A Tale of Two Processes: Categorization Accuracy and Attentional Learning Dissociate with Imperfect Feedback

Caitlyn M. McColeman (caitlyn_mccoleman@sfu.ca)
Department of Psychology, Simon Fraser University
8888 University Drive, Burnaby, BC V5A 1S6

Aaron J. Ancell (aaron_ancell@sfu.ca)
Department of Philosophy, Simon Fraser University
8888 University Drive, Burnaby, BC V5A 1S6

Mark R. Blair (mark_blair@sfu.ca)
Cognitive Science Program and Department of Psychology, Simon Fraser University
8888 University Drive, Burnaby, BC V5A 1S6

Abstract

The present study used eye-tracking to examine the relationship between attention and category learning in probabilistic environments. While training, participants received either perfect feedback (100% accurate), or one of three different levels of probabilistic feedback (87.5%, 75% or 62.5% accurate). It was found that participants in the 87.5% condition were more accurate than participants in the other two probabilistic feedback conditions. However, despite their greater accuracy, participants in the 87.5% condition continued to attend to irrelevant information as frequently as those in the other two probabilistic conditions. This shows that: (1) cues that are not utilized in making a categorization decision may still be frequently attended to, and (2) attentional learning is not as tightly coupled to improving accuracy as current formal models suggest.

Keywords: Category Learning; Attention; Eye-Tracking; Categorization; Probability Learning.

Introduction

Although our actual learning environment is frequently probabilistic, most of what we know about human category learning comes from experiments where participants are provided with an ideal learning environment: both perfectly predictive cues and completely accurate feedback. While there is valuable information to be gained from such studies, it is also possible that they overlook interesting learning processes found only in more natural, probabilistic environments.

In the real world, people must frequently assess the quality of the information they receive to guide their actions. Although the weather forecast might predict sunshine for the day a cloud formation over head may then further change our expectation: we may prepare for rain instead. In this probabilistic environment, the forecast is not a perfect predictor of outcome. Alternatively, the user of a damaged GPS system is unlikely to rely on the device for directions to an important meeting given that it is unreliable, while that same driver is very likely to let off the gas when the very reliable speedometer gets too far above the speed limit. Knowing what information around us is a valid predictor of the state of the environment is an important skill.

One way in which probabilistic environments are especially interesting is that they make it difficult for learners to determine which information is relevant. For example, in a classic study of learning from probabilistic cues, Castellan (1973) demonstrated that the mere presence of irrelevant cues tended to reduce the use of relevant ones. Further, he found that use of relevant cues was impacted most when they were moderately predictive. When the relevant cues were highly predictive, participants seldom used irrelevant information, and when cues were only slightly predictive, participants could not learn which which of them were predictive. Not surprisingly, the concept of attention is often invoked to account for data of this type, as in the case of Kruschke and Johanson's (1999) Rapid Attention Shifts 'N' Learning (RASHNL) model of probabilistic cue learning. In RASHNL attention to cues shifts after the agent receives an error signal. The purpose of attentional shift is solely for error reduction. The model performs its task well, in that it successfully learns to classify categories in a probabilistic environment. However the literature is beginning to show that attentional shifts in probabilistic category learning serve purposes above and beyond error reduction.

In a recent study, Little and Lewandowsky (2009) found that participants learning categories with probabilistic cues retained more information about an irrelevant cue than participants in a non-probabilistic condition. This suggests that attention to irrelevant cues increases in probabilistic environments, and that participants in such environments may allocate their attention for purposes other than reducing errors. It also suggests that people may attend to information that they do not use in their category judgments.

Unfortunately, differentiating attention to a cue from cue use has been difficult. Often methods of gauging attention to a cue have equated attention with utilization through the use of transfer stimuli (e.g. Blair and Homa, 2005). While this method does provide a great deal of insight into learning, it is unable to tell us when participants are sampling information without actually

using it. One way to examine attention independent of cue use is with eye-tracking. Eye-tracking has successfully been used to study a range of cognitive phenomena (Yarbus, 1967) such as visual search (Liversedge and Findlay, 2000) and reading (Rayner, 1998). Recently it has also been employed to study the allocation of attention in the context of category learning (e.g. Rehder and Hoffman, 2005a; Blair, Watson, Walshe, & Maj, 2009). By recording eye movements and fixations, we gain fine-grained spatial and temporal information about what participants are overtly attending that cannot be measured by via cue utilization. Support for eye-tracking as a measure of attention is found in neurological studies showing that eye movements share the same anatomical and functional networks used for covert shifts in attention (Corbetta et al., 1998). Perhaps more importantly, eye-tracking results have been shown to mirror the attentional biases and parameters found in formal models of categorization (Rehder and Hoffman, 2005b; Kruschke, Kappenman, and Hetrick, 2005).

The present study uses eye-tracking to further investigate the effect of probabilistic environments on attentional allocation. Participants trained on a category structure with an irrelevant dimension while receiving feedback that was either 100% accurate, or probabilistic (either 87.5%, 75%, or 62.5% accurate). Consistent with the findings of Castellan (1973) and Little and Lewandowsky (2009) it is expected that attentional allocation will be less efficient in probabilistic environments. It is also expected, based on previous research (e.g. Nosofsky & Little, 2010), that this negative impact of probabilistic feedback on attentional learning will be mirrored by a similar impact on accuracy. In light of previously documented dissociations between task accuracy and attentional allocation (e.g., Blair, Watson & Meier, 2009; Rehder & Hoffman, 2005a), we are particularly interested in evidence that participants may be strategically attending to information that does not influence their categorization decision. In the context of a probabilistic task, there may be strategic uses of attention that are not concerned with immediate response accuracy. For example, one might attend to irrelevant information just in case there is a shift in the environmental contingencies, or to avoid settling on a sub-optimal strategy. If such dissociations between attentional learning and category learning exist, then, following Castellan (1973), it is likely they will occur in conditions where the probabilities are moderate.

Some of the prior work examining probabilistic environments has looked at validity assessment. For example, Droll, Abbey & Eckstein (2009) asked participants to estimate the validity of a cue in terms of predicting the presence of a target. They found that participants slightly overestimated the validity of poor predictors (those that predicted the target on about 10% of trials) and slightly underestimated the better predictors (those that were predictive of target on 60% of trials). In the present study, we are interested to see how people assess the accuracy of the feedback they receive.

Methods

Participants
157 Simon Fraser University undergraduate students with normal or corrected-to-normal vision participated in the study for course credit. Four participants were excluded from analysis for not completing all trials in the experiment. Eleven participants were excluded for random response behavior indicated by both abnormally fast response times (1 SD below the mean) and poor performance (accuracy at chance (0.25)).

Stimuli
The stimulus resembled a three-armed microorganism, with one feature located in each arm (see Figure 1). The background image of the three-armed organism was selected randomly on each trial from nine possible images. The images differed subtly in background detail to increase the presentation of unique images in the task.

Each arm contained one of three possible feature types. The locations of the three possible features were randomized for each participant, but were presented in the same location on the micro-organism throughout the experiment. Each of the three features had two possible values. The difference between the two values was a subtle change in a feature property. This binary structure allows for eight possible combinations of specific feature values. Our design used four categories, defined by the values of two relevant features (Table 1). The third feature was always irrelevant. Relevance was assigned randomly across the possible locations and the possible feature images.

The full micro-organism subtended 16.3° of visual angle. The features contained within the micro-organism were located centrally in the arm. The features each subtended 1.3° and were separated by 10.6°.

Procedure
At the start of the experiment participants viewed a series of instructions telling them that they had been hired by a space laboratory to classify samples of alien organisms according to which chemical the organism produced: sodium, potassium, calcium, or lithium. They were told
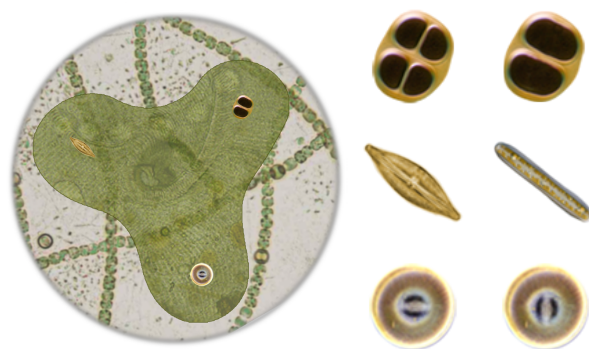

Figure 1. An example of the stimulus background (left) and the possible features values (right).

Table 1. Category Structure by Feature Value.

Relevant Feature #1	Relevant Feature #2	Irrelevant Feature	Category
0	0	0	A
0	0	1	A
0	1	0	B
0	1	1	B
1	0	0	C
1	0	1	C
1	1	0	D
1	1	1	D

that after each trial a professor would offer an opinion about which element the organism provided. Participants were informed that the professor may not be correct on all trials.

The experiment consisted of 15 blocks of 24 trials each, for a total of 360 trials. Trials began with a fixation cross which flashed between indigo and yellow in the center of a white screen. When the fixation cross flashed to indigo, participants pressed a button on the controller. If the participant did not press the button on one of the three indigo cross presentations, then the stimulus was displayed after 1500ms. This was designed to ensure that participants fixated on the central cross between trials, a necessity if accurate post-hoc drift corrections were needed. Participants then viewed the stimulus until they were ready to make a category decision. The possible responses were assigned to categories A, B, C and D (as in Table 1). Participants responded using four buttons on a Logitech game pad. After they made their category response, the stimulus remained on screen, and text with both the participant's response (e.g. "You: sodium") and the feedback response (e.g. "Professor: lithium") were presented. The location of professor feedback was counter balanced between participants (top left or top right corner of the feedback screen). Participant response was presented in the opposite corner (top left or top right).

The accuracy of the professor's feedback varied by subject condition. The possible feedback conditions, counterbalanced across participants, were 100%, 87.5%, 75% and 62.5% correct feedback. When a subject was assigned to the 100% feedback condition, the professor's feedback was correct for every trial of the experiment. In cases of imperfect feedback the professor was correct on a pre-specified number of trials. For example, in the 75% condition, the professor provided the correct answer on a randomly selected 18 of the 24 trials in each blocks. On trials where the professor was incorrect, the answer provided was randomly selected from one of the three possible incorrect answers.

At the end of the experiment, participants were asked to provide an estimate of how accurate they thought the professor was. They were reminded that chance was 25% accuracy (given that there are four possible categories). They typed in their response, prompted to respond between 0-100%.

All gaze data were collected with a Tobii X120 eyetracker (sampling at 120 Hz) accurate to within 0.5º. Raw gaze locations were transformed into fixations. We used a modified dispersion threshold algorithm (Salvucci & Goldberg, 2000) to transform raw gaze data to fixations (thresholds or of 1.9º and 75 ms were used. Circular areas of interest were defined, extending to 140 pixels from each feature's centre. Absolute values of fixation locations were corrected using information from the fixation cross to adjust for shifts in participants' posture and eye-tracker drift.

Results

Accuracy

As seen in Figure 2, participant's performance generally improved, though there were also clear differences between conditions. This was supported by a mixed model analysis of variance (ANOVA) using Condition (100%, 87.5%, 75% and 62.5% feedback accuracy) as a between subjects factor and Block (early, middle and late trials) as a within subjects factor. This showed a main effect of Condition ($F(3,142)=15.33$, $p<0.001$) and a main effect of Block ($F(1.39, 196.89)=93.71$, $p<0.001$, Huynh-Feldt correction).

The effect of training on performance was not uniform across conditions as was shown by a significant interaction between Block and Condition ($F(4.16, 196.89)=6.246$, $p<0.001$, Huynh-Feldt correction). Looking at Figure 2, it is clear that the interaction between Block and Accuracy is due to relatively rapid improvement in the 87.5% condition. As the experiment progresses, accuracy in the 87.5% condition diverges from accuracy in the 75% condition, and converges with accuracy in the 100% condition. Pairwise comparisons confirmed that accuracy in the 87.5% condition is not statistically different from accuracy in the 75% condition within Block 1 ($p=0.15$), but is different in Blocks 2 and 3 ($ps\leq0.02$).

By Block 2, accuracy in the 87.5% condition has increased to the point where it is no longer significantly different from the 100% condition ($p=0.35$), but accuracy in the 75% condition has not improved as rapidly, remaining significantly below accuracy in the 87.5% and 100% conditions ($ps\leq0.02$). These changes seen in Block 2 continue through Block 3 where there is no significant difference in accuracy between the 87.5% and 100% conditions ($p=0.69$), while accuracy in the 75% condition continues to fall behind the 87.5% condition, remaining significantly worse than in both the 87.5% and 100% conditions ($ps\leq0.005$).

Accuracy in the 62.5% condition was significantly different from the 100% and 87.5% conditions ($ps\leq0.001$), and marginally different from the 75% condition ($p=0.07$) in Block 1. In Blocks 2 and 3, accuracy in the 62.5% condition was significantly different from accuracy in all other conditions ($ps\leq0.01$). These findings suggest that performance (measured by accuracy) is robust against some levels of probabilistic feedback.
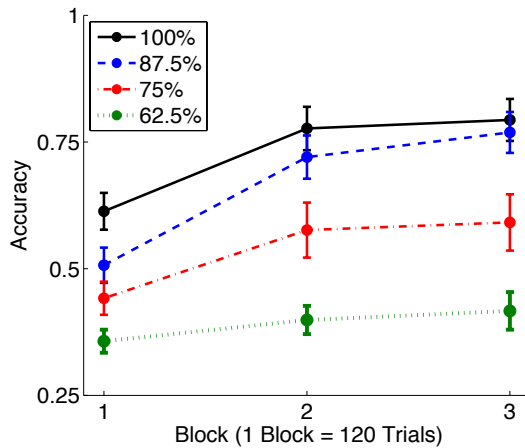
Figure 2. Classification accuracy in the early, middle and late stages of the experiment. Error bars reflect standard error of the mean.
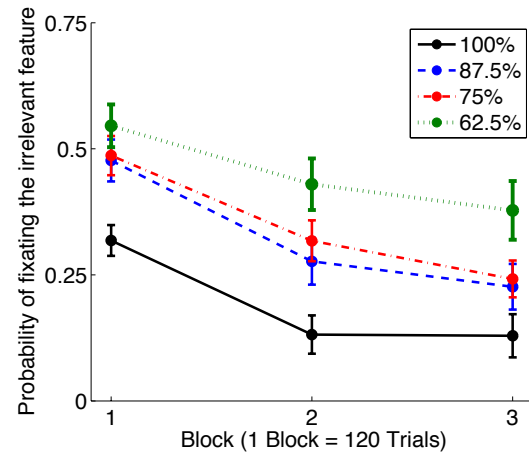

Figure 3. The probability of fixating an irrelevant dimension during stimulus presentation in the early middle, and late stages of the experiment. Error bars reflect standard error of the mean.

Attentional Learning

Attentional learning was measured by the probability of fixating the irrelevant feature. If participants are learning to more efficiently allocate their attention, then there should be a decrease in the probability of fixating this third feature.

As seen in Figure 3, participants in all conditions showed a decrease in probability of fixating the irrelevant feature over time, but there were some differences between conditions. This was supported by a repeated measures ANOVA using Condition as a between-subjects factor and Block as a within-subjects factor. There was a main effect of Block ($F_{(1.54, 219.43)}=107.7$, $p<0.001$, $\eta_p^2=0.431$ Huynh-Feldt correction), and a main effect of Condition ($F_{(3,142)}=7.134$, $p<0.001$, $\eta_p^2=0.131$). Unlike the results for accuracy, probability of fixating the irrelevant feature showed no significant interaction between Block and Condition ($F_{(4.64, 219.43)}=1.40$, $p=0.229$ Huynh-Feldt correction). This indicates that no condition changed at a rate different from the others the way that the 87.5% condition did compared to the other conditions with respect to accuracy.

The results suggest that whereas performance (measured by accuracy) is robust against some levels of probabilistic feedback, attentional learning (measured by probability of fixating the irrelevant feature) is not. There is no increase in attentional learning in the 87.5% condition relative to the 75% condition that parallels the relative increase in accuracy shown by the 87.5% condition. Although the 87.5% and 75% conditions diverge with respect to accuracy after Block 1, no similar divergence was found in the probability of fixating the irrelevant feature. Multiple comparisons showed that the 87.5% condition never significantly differs from the 75% condition with respect to probability of fixating the irrelevant feature (p=0.689). In fact, there were no significant differences in probability of fixating the irrelevant feature between any of the probabilistic feedback (87.5%, 75% and 62.5%) conditions (ps≥0.119). There were however significant differences in the probability of fixating the irrelevant feature between the 100% condition and the 62.5% and 75% conditions (ps≤0.029), and a marginal difference between the 100% condition and the 87.5% condition (p=0.071).

Estimates of Feedback Accuracy

Participants in all conditions tended to underestimate feedback quality. The mean estimates were 72% (SD=26.9) for the 100% condition, 65% (SD=18.5) for the 87.5% condition, 60% (SD=18.0%) for the 75% condition, and 48% (SD=17.1) for the 62.5% condition. However, participants also tended to provide higher
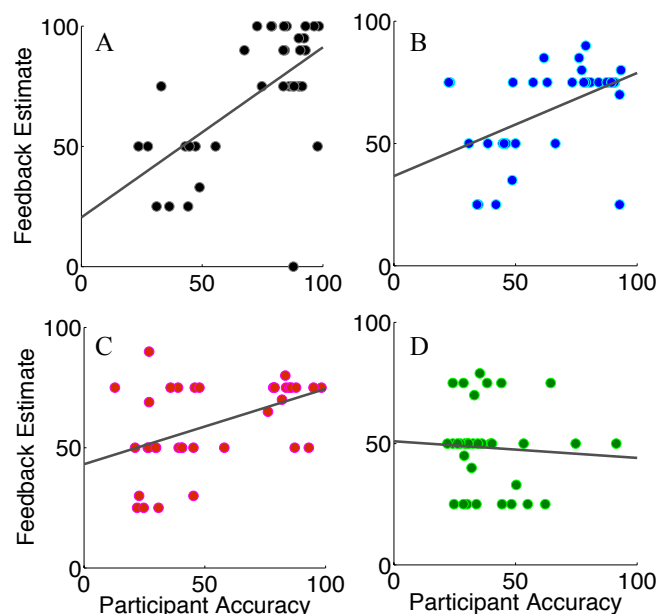

Figure 4. Estimate of feedback accuracy versus participant accuracy for (A) the 100% condition, (B) the 87.5% condition, (C) the 75% condition and (D) the 62.5% condition.

feedback estimates if their own accuracy was higher, except in the 62.5% condition (Figure 4D).

Analyses show significant strong correlations between participant accuracy and estimate of feedback accuracy in the 100% condition ($r = 0.615$, $p < 0.001$), 87.5% condition ($r = 0.517$, $p < 0.001$), and the 75% condition ($r = 0.474$, $p = 0.003$). There is no significant correlation in the 62.5% condition ($r = -0.063$, $p = 0.724$). This suggests that participants' perceptions of the accuracy of feedback were influenced by their own overall performance, though it is also possible that performance was improved by increased confidence in the feedback.

Discussion

The present study reveals an interesting difference between category learning (measured by accuracy) and attentional learning (measured by probability of fixating the irrelevant feature). Of particular interest to us was the effect of moderate probabilistic feedback. We see that average accuracy in the 87.5% feedback condition diverges from the other two probabilistic conditions, and rapidly matches accuracy in the 100% correct feedback condition. However, the probability of fixating the irrelevant feature displays no parallel improvement. The probability of fixating the irrelevant feature in the 87.5% condition never differs from that of the other two conditions (62.5% and 75%), and remains higher than in the 100% condition. This suggests that attentional learning is more sensitive to probabilistic feedback than performance. In other words, when faced with some levels of probabilistic feedback, people are able to improve their accuracy even though they do not show a corresponding improvement in attentional learning. Although attending to all of the information available on the stimulus is inefficient, participants may have had strategic reasons for continuing to attend to it, for example, to reduce the risk that meaningful information might be skipped over.

In the categorization literature, "attention" is often operationalized as the influence of a cue on a behavioral response. However, the dissociation of attentional learning and accuracy improvement supports the idea that there is more to attention than using a cue in making a category decision. Participants can attend an irrelevant feature without the irrelevant cue influencing their response accuracy (Figure 2). The idea that category learning and attentional learning are distinct is supported by recent research. For example, Rehder and Hoffman find that participants will optimize their attention only after their responses are mostly accurate (Rehder & Hoffman, 2005a). Further understanding the complex relationship between category learning and attentional learning is important for developing a comprehensive model of categorization. In future work, we intend to examine the relationship between the duration, ordering and frequency of fixations in addition to the probability of fixating an irrelevant feature as it was reported above. These measures would provide a richer description of attentional learning to inform the development of a model of categorization.

This paper takes a step in answering a broader research question: what happens in category learning beyond classification? Attention seems to function in a more conservative and less error driven way, whereas in models, changes in attention are strictly for error reduction (e.g. Krushke & Johansen, 1999). Our findings also suggest that participants are more active than these models suggest. Humans actively sample the information in the environment even when not using it to make judgments, which is inconsistent with these models' predictions that learning agents passively wait for feedback and correct weights to reduce error.

Our findings also show that there are qualitative differences between different levels of probabilistic feedback. Even though the feedback quality was spaced evenly between conditions, there were qualitative differences isolated to certain groups. Only the participants in the 87.5% feedback condition showed dissociation between accuracy and attentional learning. It appears that below a certain threshold, the limited information available in a probabilistic learning environment more severely hinders both accuracy and attentional learning. This is exemplified by poor performance and reduced attentional learning the 62.5% condition. Further support for a nonlinear difference between the four feedback conditions comes from participants' estimates of feedback accuracy. Participants with moderate probabilistic (75% or 87.5%) or perfect (100%) feedback accuracies tend to estimate higher feedback accuracy as their own performance improves (see Figure 4), but this finding does not extend to the low feedback quality condition (62.5%). This suggests that there is an important difference between 62.5% and 75% conditions. In our experiment, 12.5% feedback accuracy evenly separates the adjacent conditions. These qualities found in some conditions but not others (the accuracy versus attentional learning dissociation in 87.5% condition and the non-predictiveness of performance in estimating feedback accuracy in the 62.5% condition) indicate that the linear manipulations of the probabilistic learning environment - equally separating conditions - do not impact learning in a linear way.

Participants assigned to conditions that provide moderate feedback quality (75% and 87.5%) estimate that their feedback quality is similar (60% and 65%, respectively). These groups also showed a similar probability of fixating an irrelevant feature (see Figure 3). It may be that confidence in feedback (as measured by participant estimate of feedback accuracy) influences participants' strategy. Those who are less confident in their feedback quality, but who still have sufficient information to gauge their performance (the 87.5% and 75%) are more likely to adopt a conservative attentional allocation strategy and fixate irrelevant features.

We suggest that the observed dissociation between attentional learning and accuracy results from a delay in attentional learning (Rehder & Hoffman, 2005a; Blair, Watson & Meier, 2009). A next step would be to increase the number of trials in the experiment to see if accuracy in lower probabilistic conditions (e.g. 75% and 62.5%

feedback) would improve to match participants in a perfect feedback baseline condition for poor feedback conditions (as seen with the the the 87.5% condition).

Another direction for future research is to investigate whether merely being told that feedback might be probabilistic, even when it is not actually probabilistic - as occurred with participants in the 100% condition - is enough to influence accuracy, attentional learning, and estimates of feedback accuracy. We are currently conducting an experiment wherein all participants receive perfect feedback, but half receive instructions that indicate that feedback is perfect, while the other half receive instructions that indicate that feedback might be probabilistic.

We should note that probabilistic environments are created in different ways. Past research has investigated exception patterns (Homa, Proulx, & Blair, 2008), probabilistic cuing (Little & Lewandowsky, 2009) and probabilistic feedback (Krushke & Johansen, 1999). While some of these cases are structurally identical, it is not clear whether participants would actually treat them identically. For example, fixating irrelevant stimulus features may persist even longer in situations where the participants believe the cues are probabilistic, than in situations where they believe it is the feedback is unreliable. This would not be the first finding of an asymmetry in categorization. For example Ross and Markman (2003) provide evidence that people do not treat the category label as just another cue. Probabilistic environments, including ones created by imperfect feedback, provide an interesting platform for experimentation. They act as improvements over ideal learning environments with respect to ecological validity and they elicit interesting learning strategies. The results of this paper show an example of such a learning strategy: conservative attentional allocation in the face of an uncertain environment.

Acknowledgments

References

Blair, M. R., & 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.

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.

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.

Castellan, N. J. (1973). Multiple-Cue Probability Learning with Irrelevant Cues. *Organizational Behavior and Human Performance, 9*(1). 44-64.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, 21(4), 761–773.

Droll, J. A., Craig, C. K. & Eckstein, M. P. (2009). Learning cue validity through performance feedback. *Journal of Vision*, 9(2), 1-22.

Homa, D., Proulx, M. J., & Blair, M. R. (2008). The modulating influence of category size on the classification of exception patterns. *The Quarterly Journal of Experimental Psychology, 61*(3), 425-443.

Kruschke, J. K. & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*(9), 1083-1119.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associated blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 830–845.

Little, D. R. & Lewandowsky, S. (2009). Better Learning With More Error: Probabilistic Feedback Increases Sensitivity to Correlated Cues in Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1041-1061.

Liversedge, S. P. & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science, 4*(1), 6-14.

Markman, A. B. & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*(4), 592–613.

Nosofsky, R. & Little, D. R. (2010). Classification response times in probabilistic rule-based category structures: Contrasting exemplar-retrieval and decision-bound models. *Memory & Cognition, 38*(8), 916-927.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rehder, B. & Hoffman, A.B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*(1), 1-41.

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

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium, 1*, 71-78.

Yarbus, A. L. (1967). *Eye Movements and Vision.* New York: Plenum Press.