

**LEARNING TO ATTEND:
MEASURING SEQUENTIAL EFFECTS OF FEEDBACK IN OVERT
VISUAL ATTENTION DURING CATEGORY LEARNING**

A Dissertation
Presented to
The Academic Faculty

By

Olga Remick

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

December 2015

Copyright © Olga Remick 2015

**LEARNING TO ATTEND:
MEASURING SEQUENTIAL EFFECTS OF FEEDBACK IN OVERT
VISUAL ATTENTION DURING CATEGORY LEARNING**

Approved by:

Dr. D. Spieler, Advisor
School of Psychology
Georgia Institute of Technology

Dr. P. Verhaeghen
School of Psychology
Georgia Institute of Technology

Dr. D. Washburn
School of Psychology
Georgia State University

Dr. A. Duarte
School of Psychology
Georgia Institute of Technology

Dr. E. Schumacher
School of Psychology
Georgia Institute of Technology

Date Approved: October 21, 2015

ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. Daniel Spieler, for his guidance and help, and my dissertation committee members, Dr. Duarte, Dr. Schumacher, Dr. Verhaeghen, and Dr. Washburn for their time, consideration, and willingness to serve on my committee.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	v
LIST OF FIGURES	vi
SUMMARY	vii
1 CHAPTER 1: INTRODUCTION	1
Defining Categorization and Related Constructs	1
Measuring Attention in Category Learners	3
Visual Attention in Category Learners	6
Effects of Error Feedback in Categorization	10
Defining Research Questions	13
2 CHAPTER 2: EXPERIMENTS	18
Experiment 1	18
Experiment 2	30
Experiment 3	38
3 CHAPTER 3: GENERAL DISCUSSION	45
Eye-movements and Error Driven Learning	47
Eye-movements and Cue Competition	48
Eye-movements and Cue Utilization	51
Evidence for Top-Down Control of Eye-Movements	52
APPENDIX A: THE TIME PROPORTION SHIFT (TIPS) MEASURE	55
REFERENCES	57

LIST OF TABLES

	Page
Table 2.3.1: Comparing Utilization of the 80% Valid Cue Across Experiments 1, 2, and 3	42
Table A1: Example of Calculating the TIPS	56

LIST OF FIGURES

	Page
Figure 2.1.1: Classification Stimuli and Feedback Schedule	22
Figure 2.1.2: Average Utilization Scores in Experiment 1	24
Figure 2.1.3: Average Proportion of Fixation Time to Cues in Experiment 1	25
Figure 2.1.4: Odds of Change in Experiment 1	27
Figure 2.1.5: Sequential Effects of Feedback Congruency in Experiment 1	28
Figure 2.2.1: Average Utilization Scores in Experiment 2	34
Figure 2.2.2: Average Proportion of Fixation Time to Cues in Experiment 2	35
Figure 2.2.3: Odds of Change in Experiment 2	37
Figure 2.3.1: Average Utilization Scores in Experiment 3	41
Figure 2.3.2: Average Proportion of Fixation Time to Cues in Experiment 3	43
Figure 2.3.3: Odds of Change in Experiment 3	44

SUMMARY

Trial-level evidence for feedback sensitivity in fixations during category learning have been previously described as weak. In this dissertation, steps were taken to overcome some methodological issues potentially obscuring the evidence for such sensitivity. Jointly, the three experiments reported here suggest that sensitivity to error in visual attention reflects cue competition, as opposed to error-driven learning of a selective visual profile. These outcomes are in agreement with previous research in human vision, which holds that fixations reflect the agent's task representation. A case is made for the top-down control of visual attention during category learning, manifested as effects of prior knowledge, long-standing expectations, decisional uncertainty, and vacillations between alternative sources of conflicting evidence. A suggestion is made that the time-based measures of visual attention may align with the continuous ratings of the perceived category membership (reflecting learner confidence).

CHAPTER 1:

INTRODUCTION

Defining Categorization and Related Constructs

To a large extent, optimal decisions are dependent on one's ability to identify the most important aspects of a problem among the multitude of other, less relevant aspects. Selective attention to relevant aspects of a situation is a hallmark of an expert in any domain of human knowledge, and is emergent over time from repeated exposures to similar experiences. Our ability to extract what is relevant from an unfamiliar environment heavily depends on our ability to think categorically, that is, to generalize across a wide range of perceptual experiences according to some functional relevance (Seger & Miller, 2010). This ability to generalize serves both cognitive economy and survival by allowing people and animals to respond quickly and efficiently in situations they have never encountered before.

The process by which people and animals extract the defining features of a problem and generalize their knowledge to new situations is the central question in research in categorization. The infinite variety of categories can be broadly classified as either formal or natural. Formal categories are rule-based, have perceptually separable dimensions, and possess discrete dimensional values. When comparing zebras to horses, for instance, a child may single out the dimension of coat color and dichotomize it in terms of 'stripes' vs. 'no stripes', thus forming a formal category. Formal categories are frequently deterministic, such that knowledge of the contingency between dimensional values and outcomes eliminates all decision error: Knowing that stripes define zebras, for instance, the child will not mistake a zebra for a horse.

Natural categories, on the other hand, frequently have integral dimensions, continuous dimensional values, and probabilistic outcomes. Subtle gradations of taste, flavor, and texture are all examples of continuous dimensions. When these dimensions characterize a culinary creation, for instance, we do not ordinarily perceive them independently of one another. Values on these dimensions are combined at a pre-decisional level, such that the decision-maker is instantly aware of whether he likes the food, but the underlying pattern of dimensional values that defines the boundary between what he likes and dislikes is not easily accessible to him (Ashby et al., 1998). When a dimensional value or a combination of dimensional values is mapped onto several categories with different degrees of probability, such categories are defined as probabilistic. The same food, for instance, may appeal on some occasions, but not other. A cloudy sky is a probabilistic cue signaling rain, although rain on any given cloudy day is uncertain. Knowledge of a probabilistic category is evident in the knowledge (whether explicit or not) of outcome probabilities, as opposed to complete elimination of error in category assignment.

Categorization is closely related to processes of memory and attention (Logan, 2002). Shepard's and Luce's approaches to quantifying similarity in terms of geometric distances (Shepard, 1957), and choice - in terms of similarity (Luce, 1959), represent a conceptual bridge between theories of memory, attention, and categorization. In its most basic expression, Shepard-Luce's choice rule holds that as similarity between objects increases, so increases their confusability. Observing probabilities of correct identifications, one can derive a measure of perceived similarity among objects. Interestingly, a measure of similarity thus derived does not predict classification

probabilities (Shepard, Hovland, & Jenkins, 1961), leading the researchers at the time to conclude that identification and categorization must differ in some important aspects – perhaps, selective attention.

The initial formulation of the Generalized Context Model (GCM, Nosofsky, 1984), demonstrated that classification probabilities can indeed be derived from the identification probabilities - if one allows a parameter that differentially weights perceived similarity along relevant and irrelevant dimensions. Conceptually, this parameter is equivalent to selective attention distributed among dimensions in the amount commensurate to dimensional diagnosticity. In spite of their conceptual differences, most theories of categorization today agree with the GCM that: (1) Sensory inputs entering the system are modulated by attention; (2) Internal category representations are located in a multi-dimensional geometric space, where similarity between units is expressed in terms of distance; (3) The process that maps inputs to internal representations and selects the output operates on the error-reduction principle (Wills & Pothos, 2012).

Measuring Attention in Category Learners

Attention to dimensions is a defining property of categorization. For instance, depending on a task at hand, a person may classify items in a refrigerator as food vs. non-food, healthy food vs. junk food, or kids' food vs. adults' food, and so on. In each instance, attention is directed to a select subset of food attributes relevant for the task. The central role of dimensional attention has been especially recognized following the publication of the seminal work by Shepard, Hovland, and Jenkins (1961). Shepard et al. introduced several classification tasks that varied in the number of dimensions (one to

four) necessary for correct classification, - everything else being equal. Compared across the tasks, error rates generated by learners showed that speed and accuracy of learning a category generally decline as the number of dimensions relevant to classification increase. The fact that different sorting rules affect speed with which the same set of six stimuli can be mapped onto two category labels suggests that category learners use task-specific dimensional attention. There are several ways of measuring dimensional attention. In this section I describe a method commonly employed with probabilistic categories.

A standard procedure for observing category learning in laboratory conditions is the feedback-based classification paradigm, where a learner classifies sequences of category exemplars and receives corrective feedback after each trial. When experienced over time, feedback carries information about objective category structure. A category structure is a set of dimensions varying in the degree of relationship they bear to each category outcome. These degrees of relationship between dimensions and outcomes are called *dimensional validities*. A highly valid dimension would signal the same outcome most of the time, if not always. Dimensional validities are based on conditional probabilities, which in turn are based on frequencies of events. In most basic terms, *a dimensional validity communicates probability of a category outcome given a specific dimensional value*. For instance, if the ‘Zebra’ outcome takes place every time the ‘Coat Color’ dimension takes on the ‘Stripes’ value, the dimension of ‘Coat Color’ is said to be 100% valid or, alternatively, $P(\text{Category}_{(\text{Zebra})} | \text{Coat Color}_{(\text{Stripes})}) = 1$. Expressed in this way, a dimensional validity could range from 0 to 100, with 0 and 100 indicating a perfectly predictive dimension (a validity of 0 suggests that a dimension always predicts absence

of a given category outcome). If both outcomes are equally likely, a validity of 50% indicates a non-diagnostic dimension. There are several ways of conveying a magnitude of dimensional validity suitable for different types of structures, and a detailed discussion of this topic can be found in Edgell and Morrissey (1987) or Kruschke and Johansen (1999).

Over the course of an experiment, a learner builds an internal category representation. Ideally, one would want a category representation align with the objective category structure. In practice, however, this is not always the case, particularly when a category is probabilistic. High error rates in the final block of trials raise a question of how the learner's internal category representation deviates from the experimental category structure. This question can be answered in terms of dimensional utilization scores. Similar to dimensional validities described earlier, dimensional utilization scores are based on conditional probabilities. These conditional probabilities, however, are generated by learners and serve as behavioral indicators of dimensional attention.

The following example illustrates the basic rationale behind dimensional utilization scores. Suppose there are 20 toys in a box, where each toy attribute is equally possible. If ten of these toys are red, then the probability of selecting a red toy, in a long run and by chance alone, is .5. This is a rate we would expect from a child who reaches in the box without paying attention to box contents. Suppose, however, that we observe a child who picks up a red toy every time he reaches in the box. One would infer that this child attends to toy colors or that a toy color reliably predicts the child's choice. In a similar way, dimensional utilization scores indicate probability of a categorical choice, given a particular category attribute. If learners' choices can be reliably predicted from

dimensional values, we may infer that the learners selectively attend to (utilize) the values for this dimension (or a set of dimensions).

Visual Attention in Category Learners

Probabilities associated with response choice data are used in development and testing of formal theories of categorization (e.g., Kruschke, 2001; Nosofsky, Stanton, & Zaki, 2005). Traditionally, such theories are mathematical, and they model performance by representing attention in terms of weights that bias models' decisions in some systematic ways. Response probabilities to old and new category exemplars serve as model outputs and are compared to empirical data. More recently, efforts have been made to augment response choice probabilities with other measures of attention, including ones afforded by the eye-tracking technique.

Application of the eye-tracking technique to categorization rests on the assumption that changes in the distribution of fixations during a task are related to changes in the distribution of attention in the decision process. Research in visual attention in natural tasks (e.g., Ballard & Hayhoe, 2009), written language comprehension (e.g., Rayner, 1998), expertise (e.g., Gegenfurtner, Lehtinen, & Säljö, 2011) and decision-making (e.g., Orquin & Loose, 2013) suggests that it is a warranted assumption. A consistent finding across a variety of experimental paradigms is that stimulus attributes which are deemed of greatest utility by an agent receive the greatest share of fixations. These subjective utility effects consistently override low-level perceptual saliency effects which frequently characterize visual scenes (Ballard & Hayhoe, 2009).

A second point on which these literatures converge is that expertise modifies overt visual attention in a consistent way. For instance, a meta-analysis of eye-tracking research on expertise differences reports that, across several domains of human knowledge (e.g., sports, medicine, transportation) experts are reliably less likely to fixate irrelevant information, less likely to repeat fixations, more likely to fixate relevant information, need shorter time to first fixate relevant information, and have shorter fixations as compared to that of novices (Gegenfurtner, Lehtinen, & Säljö, 2011).

In a similar way, recordings of eye movements during categorization show that patterns of fixations reflect expertise: All else being equal, people enter a task sampling everything and finish the task with some established ocular-motor routine. Once a category is mastered, eye fixations become sensitive to dimensional validity (Hoffman & Rehder, 2010; Rehder & Hoffman, 2005a; Rehder & Hoffman, 2005b). A change in the number of relevant dimensions (one, two, or three) produces a similar change in the number of dimensions fixated: If a single dimension defines a category boundary, then the single dimension is fixated; if two dimensions jointly define the boundary, then the two dimensions are fixated, and so on (Rehder & Hoffman, 2005a). If dimensional validity varies on a stimulus-by-stimulus basis, eye-fixations reflect the stimulus-specific attention profile as well: The most diagnostic dimension in the context of stimulus A (which is irrelevant in the context of stimulus B) receives more fixations when stimulus A is presented, and fewer fixations when stimulus B is presented (Blair et al., 2009).

In cases where learners' accuracy reaches an asymptote while error rates are still high, eye-movements are typically less selective to differences in dimensional validity. Learners who experience high error rates during final blocks of trials are not likely to

visually favor any single dimension. The temporal order of sampling these dimensions, however, may become highly routinized at this later phase of categorization (Chen, Meier, & Blair, 2013).

Use of eye-trackers in the context of categorization is motivated, in part, by a prospect that theoretical accounts of category learning can be evaluated using measures of overt visual attention such as fixation duration, fixation sequence, or fixation probabilities. Eye-tracker variables have been applied as evidence in arbitrating among competing theoretical accounts of categorization in respect to dimensional attention (Blair, Watson, Walshe, & Maj, 2009; Rehder & Hoffman, 2005a), exemplar and prototype based learning (Rehder & Hoffman, 2005b), classification and inference types of learning (Hoffman & Rehder, 2010), and mechanisms of blocking and overshadowing in attributions of cause (Kruschke, Kappenman, & Hetrick, 2005).

Eye-movement data and response probability data share some common properties, which makes both types of data readily amenable to theoretical modeling. As has been mentioned earlier, fixations reflect category structure once the structure has been learned. Consequently, questions addressed with eye-movement data have been similar to those addressed with the choice probability data, namely: “Given a categorical structure X , what is the internal category representation formed by an average learner?” The internal representation is operationalized in terms of aggregated measures of overt behaviors (whether choice probabilities or eye-movements). Note, that the behaviors in question are measured upon completion of training. Given that aggregate measures of eye-movements reflect the central tendency in task representation, the eye-movement data readily lends

itself as an alternative measure of dimensional attention in testing mathematical predictions about category learning outcomes.

Notwithstanding the similarities between eye-movements and response choice, there are some unique properties to eye-movements. If response choice data can be compared to a dichotomous variable (where an option is either chosen or not), fixations are represented by a range of continuous variables (e.g., duration time, proportion of total trial time, fixation frequency, fixation probability, fixation probability as a function of trial time, fixation sequence, ordinal position, etc.) – each of which carries a slightly different information about the on-going behaviors. For instance, fixation times to dimensions may show no dimensional sensitivity, while fixation order may reveal some stable preferences (e.g., Chen, Meier, & Blair, 2013). Fixations sub-serve multiple functions in the execution of cognitive and motor tasks, but only a subset of these functions is directly related to the process under investigation. Fixations show sensitivity to perceptual and spatial characteristics of visual displays, which is not always consequential to learners' decisions or objectives (e.g., Glaholt, Wu, & Reingold, 2010). Fixations support memory representations and movement - both of which undergo rapid optimization with practice (e.g., Blair, Watson, Meier, 2009; Droll & Hayhoe, 2007; Haider & Frensch, 1999). Consequently, fixation duration and frequency change as a function of time on task. Proportion of total trial time spent in fixations as opposed to transitions also varies as a function of error rate (see, Kibbe & Kowler, 2011). The implications of these and other properties of the eye-movement data for the inferences being made when evaluating theories of categorization have not been systematically treated. Apart from the inferential issues, however, it is safe to conclude that the general

pattern emerging from eye-tracker data in categorization is consistent with earlier findings reported in other cognitive contexts, namely that (1) eye-movements are extremely sensitive to task representation, such that changing instructions will change the way people fixate visual displays, everything else being equal (e.g., Haider & Frensch, 1999; Yarbus, 1967); and (2) in contrast to novices, experts preferentially fixate task-relevant information in a way that optimizes speed and accuracy of response (e.g., Land & McLeod, 2000).

Effects of Error Feedback in Categorization

By which process do novices detect what is relevant in a task and ultimately arrive at a selective attention profile characteristic of experts? This question has not received as much attention in categorization literature as the question concerning the learning outcomes. Elsewhere, it has been suggested that modulation of attention during cognitive tasks is triggered by the error signal which is in turn generated by fluctuations of dopamine levels in the brain areas that map motor sequences onto perceptual inputs (e.g., Braver & Cohen, 2001; Holroyd & Coles, 2002). The idea that internal and external error signals shape behavior is central to research in attention (Ridderinkhof et al., 2004), decision-making (Glimcher, 2003), learning and motivation (Hazy, Frank, & O'Reilly, 2010; Schultz, 2006), as well as categorization (Seger, 2008; Seger & Miller, 2010).

In successful categorization, one of the most important goals is to optimize attention across stimulus dimensions in a way that favors most diagnostic dimensions. Mathematical theories accomplish this goal via the error-reduction mechanism (for a review, see Kruschke, 2001; Wills & Pothos, 2012). The mechanism is conceptually

similar to Thorndike's Law of Effect, which holds that behaviors associated with undesirable outcomes have lower probability of recurring compared to behaviors associated with desirable outcomes. Learning theories of categorization agree that a similar associative principle must operate at the level of selective attention, where learners use external feedback signals to orient themselves toward informative category attributes and away from the error-causing category attributes.

In a mathematical theory, attention is a part of a learning algorithm that differentially weights contributions from multiple dimensions toward the ultimate decision about similarity among category members. Attention thus defined, however, is not mapped *directly* onto any observable behavior – which constitutes the most frequent criticism directed at theories of categorization (e.g., McColeman et al., 2014). Can overt visual attention be mapped onto the theoretical construct invoked in models of categorization? Is it possible to observe trial-level feedback-related changes in visual attention in human learners? The aim of this dissertation is to address this question.

Adapting a categorization task for use with the eye-tracker requires separation of dimensions in space to insure that only one dimension is fixated at any point in time. If a learner considers a dimension important, he or she will allocate a greater proportion of trial time to such a dimension. A pattern of fixation times to dimensions on trial N can be compared to that on trial N-1. If eye-movements are sensitive to error-driven modulations of dimensional attention, we should be able to observe that a dimension that aided a correct response on trial N-1, draws more visual attention to itself on trial N. Conversely, we should be able to see that a dimension that led to an error feedback on trial N-1, receives less visual attention on trial N.

This rationale has been applied in a recent work (McColeman, Barnes, Chen, Meier, Walshe, et al., 2014), which examined sequential effects of error in a group of nine eye-movement datasets coming from the same lab, albeit different studies and designs. The measure of particular interest discussed in the study is the Time Proportion Shift (TIPS), which quantifies any change in the distribution of visual attention across category dimensions from trial N-1 to N: The less there is a change from trial one to trial two, the closer the TIPS is to zero (details on how the TIPS score is computed can be found in Appendix A). Aggregated TIPS scores carry information about learners' tendency to re-adjust attention distribution (or, more specifically, shift relative proportions of fixation times across dimensions) on trial-to-trial bases. Overall, ability to detect TIPS declines with block and is affected by category type. More specifically, TIPS tends to favor rule-based, non-metric categories, and less likely to emerge in categories with continuous dimensions and/or requiring information integration (as opposed to easily verbalized rules).

A second measure of interest discussed in McColeman et al. (2014) is Error Bias, which conveys the degree to which a learner is likely to evidence TIPS following an error as opposed to correct (confirmative) feedback. Error Bias can be present (1), absent (0), or be in a direction opposite from the one expected (-1), such that all change in attention occurs subsequent to correct feedback, and none subsequent to error. Error Bias is a measure comparing average TIPS following correct with average TIPS following error feedback, experiment-wise. Error Bias will not be different from 0, when the TIPS are not different from 0. Presence of TIPS, however, does not guarantee presence of Error Bias, - one reason being the fact that TIPS tends to disappear with block, which is not

taken into account in the measure of the experiment-wise Error Bias. Overall, five studies out of nine evidenced the effect of Error Bias, with effect sizes ranging from about .04 to 1. Upon reviewing the datasets, the authors concluded that the error bias is clearly present, albeit so small and illusive, it begs a question whether processes other than the error reduction mechanism dominate overt visual attention during category acquisition.

Defining Research Questions

The aim of this dissertation is two-fold: (1) to overcome some of the methodological challenges to measuring sequential effects of feedback in fixations during categorization, and (2) to gain a better understanding of processes guiding visual attention during category learning. In the following sections, I describe the specific methodological challenges addressed in this dissertation, and provide general rationale for the experiments and analyses implemented.

1. Ill Defined Problem Space

Perhaps the greatest methodological challenge to discerning central tendencies in eye-movements during categorization is the ill-defined problem space at the task outset. The feedback-based classification paradigm typically imposes no constraints on how category information is to be parsed. These constraints are implicitly present in the trial-level feedback. By the time learners have extracted these constraints from the feedback, however, internal category representations, choice patterns, and ocular-motor routines are already in place. Processes guiding the initial constraint acquisition for the most part

elude attempts at aggregation: Indeed, in a three dimensional category, learners can be testing one-, two-, or three-dimensional rules; in a probabilistic category, learners may also be testing rules involving strings of trials (e.g., see Jones & Sieck, 2003).

Aggregating these early data across multiple learners (and trials within a learner, for that matter) inevitably results in undifferentiated noise typically observed in eye-tracker data during the initial few blocks of classification learning.

Solution

One way to address this issue is to constrain the initial solution space to only one, common to all solution. This can be accomplished by embedding a familiar semantic theme in an otherwise abstract categorization task, such that learners' attention is initially drawn to the common semantic theme. To this end, Experiment 1 includes a probabilistic category introduced as a weather prediction task, where two dimensions are abstract and one dimension is represented by weather-themed pictures, suggesting either a rainy or a sunny type of weather. Aggregated measures of a shared behavior are expected to be more sensitive to any general processes that might guide the initial category acquisition.

2. Aggregation

Another challenge associated with measurement of sequential effects of feedback in eye-movements concerns the dilemma between data richness and need for aggregation. Consider the following example. Sequential effects of feedback on response probabilities have been demonstrated in metric and non-metric probabilistic categories (Jones & Sieck, 2003; Jones, Love, & Maddox, 2006). These studies show that response choice

probabilities vary as a function of perceptual similarity between two adjacent trials, such that higher perceptual overlap between two adjacent trials results in a higher probability of making a response congruent with the feedback seen on the previous trial. Multi-dimensional richness of eye movements poses a serious challenge to generating similar evidence with the eye-tracker. The underlying assumption guiding the acquisition of ‘fixations-to-dimensions’ measures is that fixations vary as a function of perceived *dimensional importance*, - not a function of perceived exemplar similarity. Strictly speaking, there is no reason for a learner to re-distribute fixations among dimensions based on perceptual overlap (or lack thereof) between two adjacent trials. There is also no direct way of verifying the perceived dimensional importance at the trial level, at the subject level (ignoring for the moment the eventual need for aggregation across trials and subjects).

The Error Bias suggested by McColeman et al. (2014), discussed earlier, provides a coarse sense of attentional changes inherent to early stages of learning. Recall, that Error Bias is based on the TIPS (Time Proportion Shift, which is the aggregated difference score, summarizing changes in trial time to dimensions from trial N-1 to N). First, the aggregation takes place at the trial-level, then at the experiment level, and finally across subjects – with some information being lost at each step. Recall also, that reduction in fixation-time shifts across dimensions is associated with time on task, high accuracy, continuous category dimensions, and category boundaries involving information integration from multiple dimensions (Chen, Meier, & Blair, 2013; McColeman et al., 2014).

Solution

With these considerations in mind, experiments described in this dissertation employ probabilistic categories with binary (as opposed to continuous) dimensions, and easily verbalized rules based on a single dimension. In addition, eye-movement data for the first and the second half of study are analyzed separately within each experiment. Finally, in this dissertation I depart from time-based analyses of sequential effects in eye-movements, and suggest an alternative approach, which (1) is not based on a contrast between the predictive and non-predictive cues; (2) does not involve aggregated time measures, but is based on fixation probabilities. The following paragraphs describe the rationale and the mechanics of the approach.

In basic terms, the effect I am aiming to detect is this: If a learner receives error feedback - he or she attempts something different the next time; if the learner receives correct feedback - he or she repeats the action again. Two behaviors in closest temporal proximity to the event of feedback are (1) the last fixation on trial N-1, reflecting decision and (2) the first fixation on trial N, reflecting the priority in accessing cues. My working hypothesis is that *reinforcing the last fixation of trial N-1 with the correct feedback will increase the odds of executing a fixation to the same location on the following trial N. Alternatively, discouraging the last fixation of trial N-1 with the error feedback will decrease the odds of executing a fixation to the same location on the following trial N.*

In adapting this hypothesis, I make an assumption that the last fixation in a trial reflects dimensional importance. As the learners execute a key-press indicating their category choice, their eyes are typically resting on one of information sources available

on the screen. It makes rational sense to assume that the information fixated at the moment of response is the information most heavily influencing the response. Empirical evidence supporting this assumption comes from consumer and decision-making research, indicating that a chosen option has higher probability of being viewed at the end of a trial, compared to other options (Glaholt & Reingold, 2011; Orquin & Loose, 2013).

When considering information about location of the first and the last fixation from each trial, there are two scenarios of interest: (1) The last fixation on trial N-1 is repeated (i.e., is in the same location) on trial N – constituting no change; (2) The last fixation on trial N-1 is not repeated (i.e., is to a different location) on trial N – constituting a change. If the hypothesis outlined above holds, I should observe more instances of change associated with the error feedback, and fewer instances of change associated with the correct feedback. Alternative hypothesis would hold that feedback type has no effect on the odds of re-visiting previously visited location. In all three experiments reported below, the odds of change given error vs. correct feedback are computed for each learner and compared with a paired samples t-test.

CHAPTER 2:

EXPERIMENTS

Experiment 1

In its most basic expression, the learning principle holds that a behavior leading to a favorable outcome has a greater probability of re-occurring compared to a behavior leading to an unfavorable outcome (Thorndike's Law of Effect). Implicit in all formal models of categorization is an assumption that a similar learning principle operates at the level of attention, which is being directed or withdrawn from a stimulus attribute depending on the type of feedback associated with the attribute. The purpose of Experiment 1 is to demonstrate this effect in eye movements.

A critical property of Experiment 1 is that the to-be-learned category includes a semantically relevant dimension, which is contrasted with two abstract dimensions. The semantically relevant dimension is deliberately rendered unhelpful to category discrimination. Attending to the semantically relevant dimension results in sub-optimal responding characterized by frequent error. The optimal solution involves attending to one of the abstract dimensions and disregarding the other two. Theoretically, this can take place through systematic shifts of attention in response to feedback: When a learner receives feedback congruent with the semantic concept, his or her attention should be drawn to the semantic concept. When a learner receives feedback incongruent with the semantic concept, his or her attention should be drawn away from it.

The experiment is thus designed to encourage attention shifts from the semantically relevant, but empirically un-helpful dimension toward the abstract, but highly reliable dimension. If such shifts of attention occur, they should be in the same

direction for all learners, to the extent that all learners share semantic knowledge present in the weather concept.

Because the ‘dimensional importance’ in this task is strongly suggested to learners at the outset, an analysis of sequential effects following the concept-congruent and concept-incongruent feedback is possible. Concept-congruent feedback occurs when information conveyed by the weather cue is supported by feedback (*e.g.*, a picture of the cloud is followed by feedback signaling rain). Conversely, concept-incongruent feedback occurs when the information conveyed by the weather cue is contradicted by feedback (*e.g.*, a picture of the cloud followed by feedback signaling sun). It is important to keep in mind, that the independent variable of feedback type (congruent vs. incongruent) occurs during trial N-1, while the dependent variable of attention to cues (time spent fixating cue 1, 2, and 3) is measured during trial N. It is expected that attention to cues within the concept-congruent condition should show a bias toward the semantically relevant cue. This bias should not be present on trials following concept-incongruent feedback.

Finally, in order to make a statement about the trial-to-trial nature of attention shifts, only adjacent pairs of trials are considered. These pairs of trials are selected on a condition that the first member of a pair follows congruent feedback, and the second member of the pair follows the incongruent feedback.

In sum, if the task structure is effective in encouraging all learners to adopt the same initial solution of labeling the cloud-containing displays as “Rain” and sun-containing displays as “Sun”, then aggregated measures of their behaviors should be sensitive to sequential effects of feedback on overt visual attention. This paradigm,

therefore, has a potential for demonstrating the process of new learning as it unfolds through trial-to-trial modulation of overt visual attention in response to feedback.

Method

Participants

Thirty undergraduate students, with normal or corrected to normal vision, from Georgia Institute of Technology participated in the study and received course credit. Average age was 19 years (SD = 1.4); 53 percent of the sample were males; 73 percent reported English as their first language; 60 percent of the sample were White, 37 percent were Asian, and the remaining 3 percent identified themselves as “Other”.

Apparatus

Data were collected with the EyeLink System, Version 1.2 (SensoMotoric Instruments, Inc., Boston, MA), sampling at 250 Hz. Raw eye position data were parsed into fixations and saccades using the EyeLink software, where saccades were defined as movements with velocities exceeding 30 degrees/second, movements below the threshold were classified as fixations.

Analyses include time spent in fixations only. Areas of interest on the display were computed as quadrants with sides subtending approximately 14 degrees of visual angle around each symbol. Fixations falling outside the quadrants were discarded. The stimuli appeared on a 20-inch monitor, in black font against white background (see Figure 1A). Each symbol in the display subtended $\sim 1/2$ by $\sim 1/2$ degrees of visual angle, placed ~ 21 degrees apart, and at ~ 14 degrees of visual eccentricity.

Stimuli

Each display contained three dichotomous cues: Two abstract cues and one picture cue with the two values symbolizing cloudy and partially cloudy weather (from here on Weather Cue). Examples of the stimuli prototypical of Sun and Rain conditions are shown in Figure 2.1.1, panel A. Participants learned to categorize the displays as “Sunny-Day” vs. “Rainy-Day” (from here on the S vs. R categories) via a key-press on a standard computer keyboard. The hand to label mapping was counterbalanced across participants.

The logical structure of the category, component validities for each cue, and the feedback schedule associated with each stimulus are shown in Figure 2.1.1, panel B. There were eight categorization stimuli in all, half of which were representative of category S and half of category R. The column labeled P(S) shows probability of Category S feedback for each stimulus within a block (20 trials). Note that the fractions in the P(S) column represent feedback frequencies from which component validities for individual cues can be derived. Component validities for the cues were 80%, 50%, 60%. Weather cue was always 60% valid (meaning that attending to Weather Cue alone would result in 60% accuracy), and always appeared at the bottom quadrant of the display. Assignment of the abstract cues to locations and validities was counterbalanced across participants, but remained fixed for each learner. Block-wise and experiment-wise, $P(R) = P(S) = 50\%$. Each cue value occurred equal number of times during a block and the entire experiment.

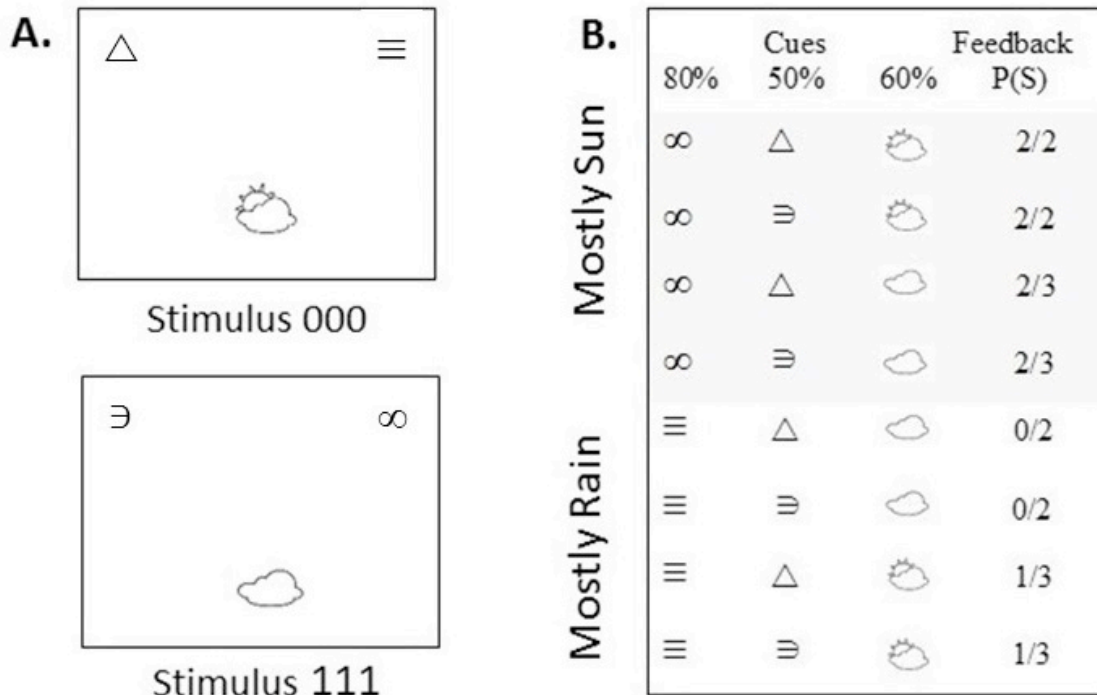


Figure 2.1.1. Classification stimuli and feedback schedule. **A:** The prototypical “Sun” (000) and “Rain” (111) instances. **B:** The feedback schedule illustrating probability calculations for feedback “Sun” for the three-cue combinations within one Block (20 trials). The denominator of each fraction is the number of times the stimulus appeared in the block. The nominator of each fraction is the number of times feedback “Sun” appeared in conjunction with the three-cue stimulus. Note that $P(S) = P(R) = 50\%$ block-wise and experiment-wise.

Procedure

Participants were told to predict weather by labeling computer displays as either “Rain” or “Sun”. They were told that perfect performance on the task is impossible, but if they try hard, they can get most of the trials correct. The fact that the task can be solved by attending to only one cue was not communicated to the participants.

On each trial, the following sequence of events took place: a) a central fixation sign (“+”) appeared for 700 ms, b) a stimulus display appeared and remained on the screen until response, c) a feedback screen showing a picture of either sun or rain appeared for 1000 ms, and d) a blank screen lasting 700 ms, was followed by a fixation

cue “+” in preparation for the next trial. The experiment included 10 blocks of 20 trials. The entire session lasted approximately 30 minutes.

Results and Discussion

Computing Cue Utilization

Recall that the category consisted of three dichotomous cues varying in predictive validity (80%, 60%, and 50%). Consider the following example. If X represents the 80% valid cue, its two values are X_0 and X_1 . If X_0 appeared 10 times during a block and was followed by feedback signaling RAIN 8 times, its validity is 80%. Because cues are symmetric, X_1 would appear 10 times during a block as well, and would be followed by feedback signaling SUN 8 times. Cue X is said to be 80% valid, and its two values are said to be symmetrical. Component validities for the other two cues can be arrived at in a similar way.

The extent to which learners’ responses are consistent with the category structure is measured in terms of cue utilization. Ideally one would want cue utilization be commensurate with cue validity, but this does not have to be the case (Edgell & Morrissey, 1987; Kruschke & Johansen, 1999). Utilization scores are computed similar to cue validities. Specifically, if cue X_0 appeared 10 times during a block and a learner responded RAIN 10 times out of 10, the utilization of X_0 is said to be 100%. Ideally, utilization scores for the *two values* (e.g., X_0 vs. X_1) of *the same cue* should be similar. First analysis verifies this assumption. A 2 x 3 x 10 ANOVA with Cue Value (0 vs. 1), Cue Validity (80% vs. 60% vs. 50%), and Block (1 through 10) as repeated measures

showed no main effect of Value, $F(1,29) = 1.12, p = .30$, or interaction of Value with other two factors, confirming that learners were working with three dichotomous dimensions (as opposed to six unrelated symbols). Subsequently, information about the individual values for each cue was collapsed, and average utilization for each pair of values is reported as overall cue utilization scores.

Utilization scores are plotted in Figure 2.1.2, which shows the main effect of Cue Validity, $F(2,58) = 15.55, p < .001$, partial $\eta^2 = .35$, and Validity x Block interaction $F(18,522) = 7.62, p < .001$, partial $\eta^2 = .21$, due to changes in priorities attributed to Weather and the 80% valid cue in the first three blocks. Notice that during the initial 20 trials (Block 1) average utilization of Weather Cue surpasses that of the other two cues, confirming the effectiveness of the experimental manipulation aiming to constrain the initial solution space to a common solution.

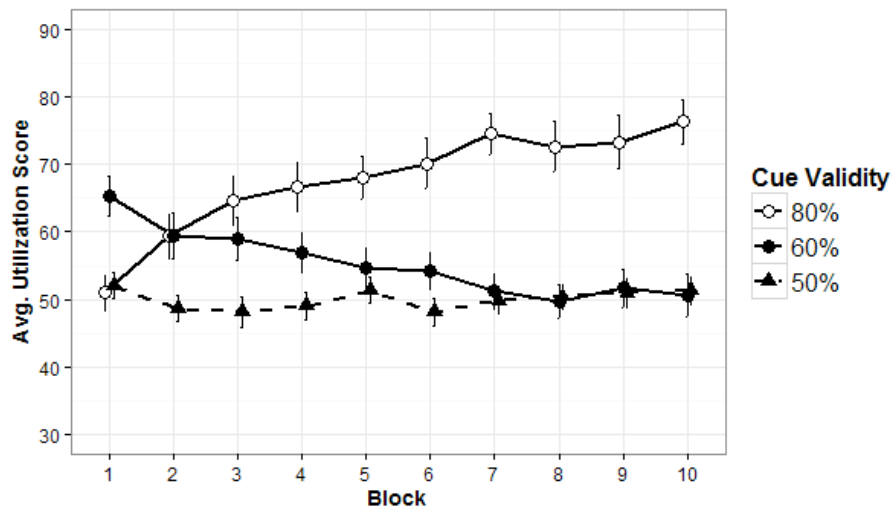


Figure 2.1.2. Average utilization scores in Experiment 1. Learners show preference for the 80% valid abstract cue, starting at Block 4. Here and throughout: (1) the error bars represent standard error of the mean; (2) the empty circles represent the most reliable cue; (3) the dotted line represents uninformative cue.

Proportion Trial Time Allocated to Cues

Ideally, one would expect that overt visual attention should follow choice behavior reflected in cue utilization. Figure 2.1.3 shows that, - somewhat similar to utilization, - fixations are re-distributed prior to Block 3, after which, attention consistently favors the 80% valid cue over the 50% valid cue. However, unlike utilization, visual attention continues favoring Weather Cue throughout the course of the experiment. These observations are confirmed by a 3 x 10 ANOVA with Validity and Block as within-subject factors, with a Validity x Block interaction, $F(18, 522) = 2.44$, $p < .05$, partial $\eta^2 = .08$, and no main effects. The interaction is due to the fact that fixations to the 50% valid cue decreased after Block 2.

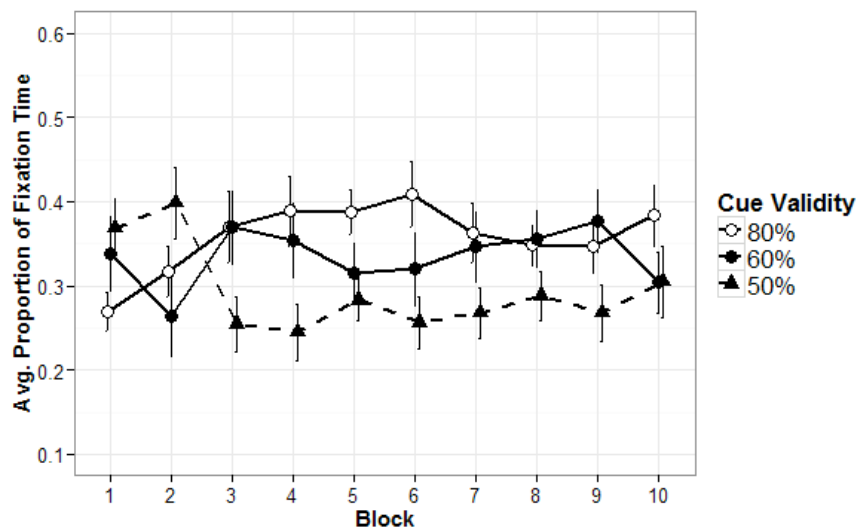


Figure 2.1.3. Average proportion of fixation time to cues in Experiment 1. Note the relatively non-discriminating visual profile persisting for the duration of study.

Utilization data suggest that the first half of training is different from the latter part, in that some re-distribution among cue priorities takes place early on, but not later in training. For the following two analyses the data from blocks 1 through 5 (first 100 trials) is examined separate from the data from blocks 6 through 10 (last 100 trials). My

main focus is on the first 100 trials, because both the utilization scores and fixation time to cues suggest that learners compute cue priorities during this initial period.

Effects of Feedback on the Odds of Revisiting Most Recently Fixated Cue

On each trial learners receive feedback which tells them whether the response they have just executed was right or wrong. Learners typically maintain their gaze on one of the three cues while executing a response, until a feedback screen interrupts their gaze. The question addressed in this section is whether the type of feedback a learner receives (error vs. correct) has any effect on the probability that the gaze to the same cue will be renewed on the subsequent trial, or whether the learners will redirect their gaze somewhere else.

The following analysis compares location of last fixation on trial N-1 with location of first fixation on trial N, counting the number of location repeats (labeled *No Change*) and location changes (labeled *Change*). Figure 2.1.4 shows the odds of revisiting the most recently fixated cue (the last cue fixated on trial N-1) as a function of feedback on trial N-1. The figure shows that sequential effects of error are evident during blocks 1 through 5, but not during blocks 6 through 10. During the first phase of learning, the odds of returning to the pre-feedback cue are lower subsequent to the error feedback and are higher subsequent to the correct feedback. Paired samples t-test confirms this observation, $t(29) = 2.43, p < .05$ (similar outcomes were found with a 2 x 2 Chi Square, $\chi^2(1) = 6.53, p < .01$). The effect was not present in the later trials.

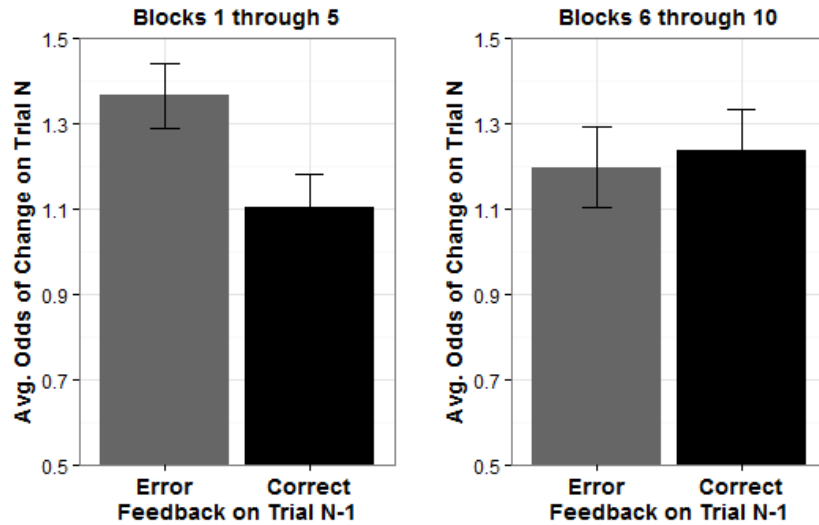


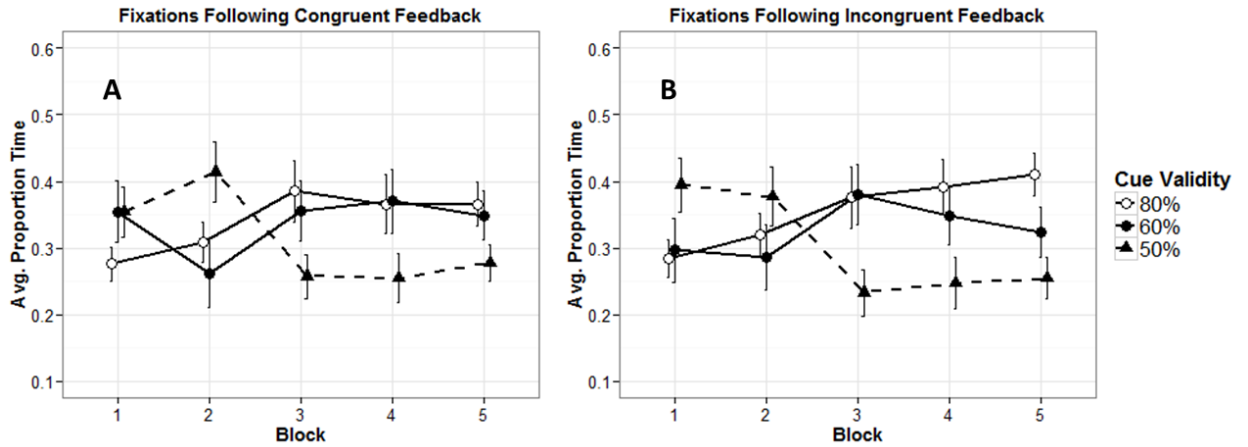
Figure 2.1.4. Odds of change in Experiment 1. The odds of change following error were computed by dividing the number of trials on which the change occurred by the number of trials on which the change did not occur – conditioned on the fact that all trials took place following error feedback. The odds of change following correct feedback were computed in the same way, with the exception that, in this case, all trials took place following correct feedback. Note that the effect of error is observed during blocks 1 through 5 only.

Sequential Effects of Feedback Congruency on Visual Attention

The final analysis compares proportion of trial time allocated to cues on trial N as a function of feedback congruency with the semantic concept suggested by the categorization stimulus on trial N-1. Concept-congruent feedback occurs when a display with a cue suggesting RAIN, for instance, is followed by feedback signaling RAIN. If the same screen is followed by feedback signaling SUN, the incongruent condition is said to take place. The expectation is that following incongruent feedback, learners should allocate less time to the generally unhelpful (60% reliable) Weather cue, - as compared to what happens during the congruent trials. The following analysis includes pairs of adjacent trials, such that the first member of a pair takes place after congruent feedback and the second member of the pair takes place after the incongruent feedback – thus, any

differences between the congruent and the incongruent conditions represent the differences between trials that happen immediately one after another.

Early Training: Blocks 1 through 5



Late Training: Block 6 through 10

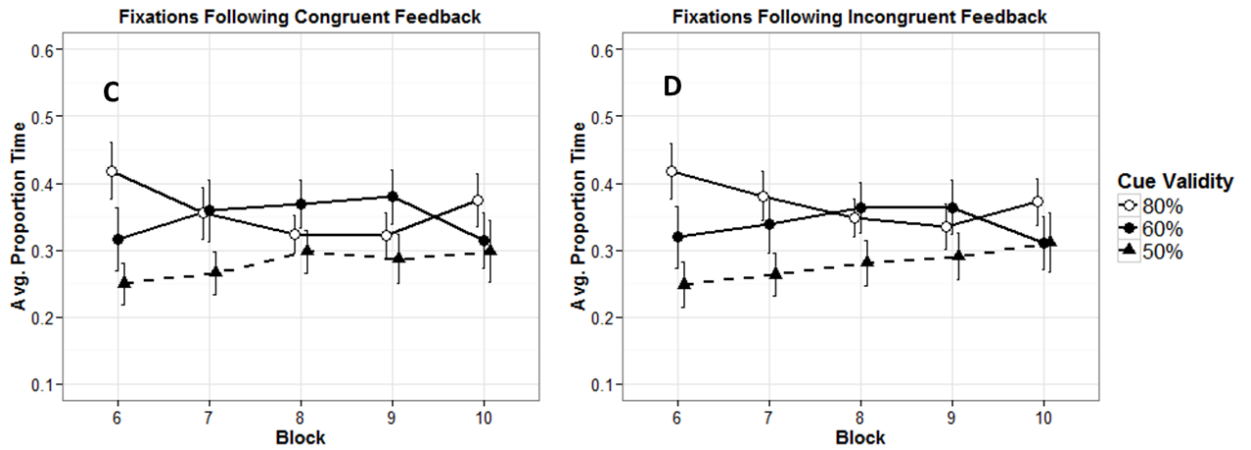


Figure 2.1.5. Sequential effects of feedback congruency in Experiment 1. Displayed are the average proportions of fixation time to cues on trial N, following concept-congruent and concept-incongruent feedback received on trial N-1. **A-B:** Three-way interaction between cue validity, feedback congruency, and block. Avg. proportion time to cues is plotted for neighboring pairs of trials, where the first member of the pair follows concept-congruent feedback, and the second member of the pair follows concept-incongruent feedback. **C-D:** There is no sequential effects of feedback congruency on time to cues during Blocks 6 through 10 (trials 101 through 200).

The results for early and late phases of the study are summarized in Figure 2.1.5. In comparing panels A and B, notice the competition between the 80% and the 60% valid cues – both represented by the solid lines – concept-congruent feedback tends to increase attention to the 60% valid cue (Blocks 1 and 5), compared to the condition characterized by the incongruent feedback. A 3 x 2 x 5 repeated measures ANOVA supports this observation, yielding a significant Validity x Condition x Block interaction, $F(8,232) = 3.24, p < .05$, partial $\eta^2 = .1$. This effect is not observed in Blocks 6 through 10, which are plotted in panels C and D.

Discussion

Although cue utilization showed that attention to Weather Cue steadily declined (disappearing completely by Block 7), fixation times to dimensions revealed that learners consistently attended to Weather Cue as much as they did to the 80% valid cue. Preference for the most reliable cue in choice data was not supported by the eye-movement data (which showed no effect of cue validity and showed persistent attention toward the semantically relevant information in spite of its low predictive value). This suggests that category learning is sensitive to factors other than experimenter feedback. Experiment 1 shows how semantic knowledge can interact with knowledge of dimensional validity in shaping learners' attention profile during categorization.

The design of Experiment 1 directed the focus of learners' attention to a specific dimension and examined how the attention responds to instances of feedback supporting or contradicting the learners' expectations about the attended dimension. There is some evidence that fixations are sensitive to feedback. There is no evidence, however, that

systematic shifts of visual attention in response to feedback shape a selective visual profile by gradually reducing fixations to dimensions associated with frequent error and increasing fixations toward the diagnostic dimension. Contrary to this hypothesis, learners utilized the most reliable dimension while employing a relatively diffuse visual attention profile and showing no sensitivity to feedback in fixations after the initial 100 trials.

A possible implication from these data is that the sequential effects detected in eye-movements reflect cue competition, as opposed to the associative type of attention learning employed in computational approaches to category learning. Cue competition takes place when two or more cues compete for learners' attention (for a detailed discussion of cue competition effects see Edgell et al., 1996). The more difficult the arbitration among candidate cues (that is, the longer time it takes a learner to settle on a solution) the lower the utilization of any individual cue. Cue competition may have consequences for visual attention in a form of sequential effects of feedback on fixation time or fixation probability observed in this experiment. Experiment 2 considers the possibility that the sequential effects of feedback in Experiment 1 reflect cue competition, where a semantically relevant but weak cue was competing with a semantically neutral but highly reliable cue.

Experiment 2

Experiment 1 revealed a degree of dissociation between attention in terms of response probabilities and overt visual attention, such that clear preference for the most diagnostic cue in terms of utilization was accompanied by a rather diffuse attention

profile in terms of times allocated to viewing the cues. Recall, that the underlying assumption guiding inference about cognitive processes in eye-movement data is that eye-movements are sensitive to subjective utility effects, so, for instance, we should expect fixation times increase as a function of perceived information importance. Results of Experiment 1, however, do not confirm this expectation. Fixations suggest that learners considered all three dimensions important, while their choices were guided primarily by the 80% valid cue. Experiment 1 suggests that eye-movement patterns do not always follow decision patterns.

The utilization data reported in Experiment 1, however, does not reflect the extent to which the semantically relevant cue was affecting utilization of the empirically relevant cue. Perhaps the non-selective fixation profile reflects the fact that learners were under-utilizing the 80% valid cue in the presence of the semantically relevant cue. This contextual effect would become apparent if utilization of the 80% valid cue increased with removal of the semantic information from the category structure. Removal of semantic component from the category structure would also clarify the nature of the sequential effects of feedback observed in Experiment 1. If sequential effects of feedback reflect the process of cue competition for visual attention, we should not observe the sequential effects in a task that does not support cue competition.

To this end, Experiment 2 replicates the design and analyses of Experiment 1, with the exception that all three cues in Experiment 2 are abstract. Experiment 2 also aims to replicate the sequential effects of feedback on eye-movements observed in Experiment 1. Because semantic information is removed from category structure in Experiment 2, the effects of feedback congruency on subjectively diagnostic cue cannot

be measured. However, the probability of re-visiting the most recently fixated cue as a function of feedback still can be assessed.

Method

Participants

Thirty undergraduate students, with normal or corrected to normal vision, from Georgia Institute of Technology participated in the study and received course credit. Average participant age was 20 years ($SD = 2.3$); 50 percent of the sample were males; 70 percent reported English as their first language; 48 percent of the sample were White, 35 percent were Asian; the remaining 18 percent identified themselves as ‘Other’.

Apparatus and Stimuli

Apparatus and stimuli were the same as these described in Experiment 1, with the exception that Weather Cue was replaced with an abstract cue, forming a three-abstract-cue probabilistic category. Component cue validities and stimulus feedback schedule can be viewed in Figure 2.1.1. As previously, component validities were 80%, 60%, and 50%, and cue-to-validity mapping remained the same as in Experiment 1.

Procedure

Participants were told to classify computer displays as either “F” or “J”, by pressing appropriate keys on a keyboard. They were told that perfect performance on the task is impossible, but if they try hard, they can get most of the trials correctly.

On each trial, the following sequence of events took place: a) a central fixation sign (“+”) appeared for 700 ms, b) a stimulus display appeared and remained on the screen until response, c) a feedback screen indicating correct category (e.g., “It was J”) appeared for 1000 ms, d) a blank screen lasting 700 ms appeared, and was followed by the fixation cue “+” in preparation for the next trial. The experiment included 10 blocks of 20 trials. The entire session lasted approximately 30 minutes.

Results and Discussion

Cue Utilization

There was no difference in utilization of two values for each cue. A 2 x 3 x 10 ANOVA – with Cue Value (coded as 0 vs. 1) by Cue Validity (80% vs. 60% vs. 50%) by Block (1 through 10) as repeated measures showed no effect of Value, $F(1,29) = 0.69$, $p = .41$, or interactions. Information about individual cue values was subsequently collapsed, and averages for each pair are reported.

Figure 2.2.1 shows average cue utilization as a function of cue validity and block. Learners showed growing preference for the 80% reliable cue which increased from 60% utilization in Block 1 to 80% in Block 5. Notice that beginning in Block 6, average utilization scores for the 80% cue approached its objective validity, seemingly reaching an asymptote. Utilization of the other two cues remained more or less stable at about 50% (suggesting that these two cues are not being used). This observation is confirmed by a 3 x 10 ANOVA with Validity and Block as within-subjects factors: The ANOVA shows a main effect of Validity, $F(2, 58) = 78.39$, $p < .001$, partial $\eta^2 = .73$; Block, $F(9, 261) =$

5.52, $p < .001$, partial $\eta^2 = .16$; and Validity x Block interaction, $F(18, 522) = 5.52, p < .001$, partial $\eta^2 = .16$.

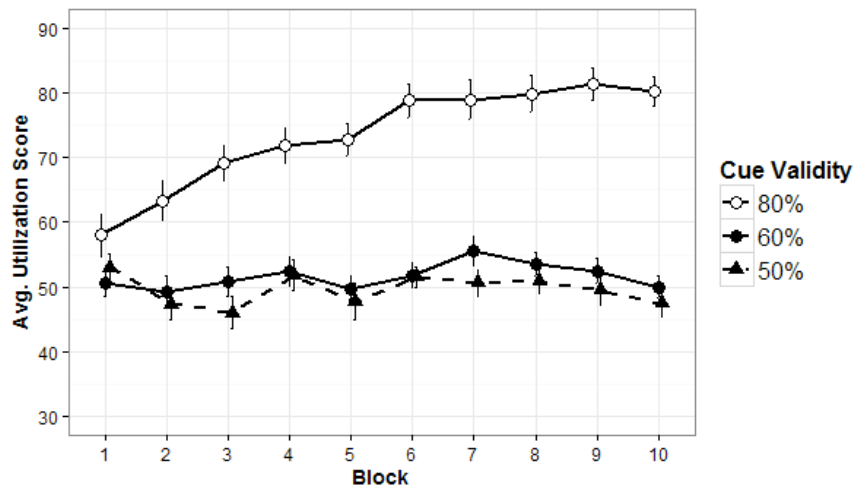


Figure 2.2.1. Average utilization scores as a function of Cue Validity and Block. Starting with Block 2, learners utilized the 80% valid cue, with little to no utilization of the 60% and 50% valid cues.

A comparison of utilization for the 80% cue between Experiments 1 and 2 showed the main effect of Block, $F(9,50) = 7.66, p < .001, \eta^2 = .57$, and no effect of Experiment or Experiment x Block interaction. Utilization of the 60% reliable cue, on the other hand was greater in Experiment 1, clearly due to its semantic relevance: a mixed effects ANOVA yielded a Block x Experiment interaction, $F(9,50) = 2.69, p < .05, \eta^2 = .33$, where utilization of the 60% cue was reliably greater in Experiment 1 during Blocks 1 through 5, but not later.

Proportion of Trial Time Allocated to Cues

Figure 2.2.2 shows average proportion time spent in fixations during a trial as a function of cue validity and block. To allow for a direct comparison between Experiment

1 and 2, the 60% cue occurred on the bottom of the screen for all subjects - as it did in the weather prediction task - while the 50% and the 80% cues occurred at the top portion of the screen, their assignment counterbalanced across learners

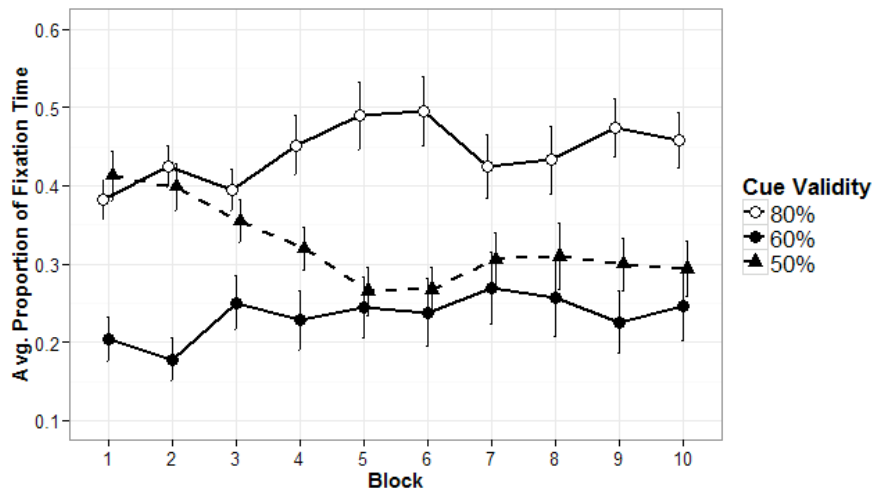


Figure 2.2.2. Average proportion of fixation time to cues in Experiment 2. Displayed is a two-way interaction between Block and Cue Validity, due to the fact that visual bias toward the upper portion of the screen containing the 50% valid cue declines from Block 1 to 5.

Apparent in the measure of proportion time toward the 50% cue (plotted with the dotted line) is the visual preference for the top portion of the screen during the first three blocks of trials. Beginning in Block 4, visual attention clearly favors the most reliable cue. These observations are supported by a 3 x 10 repeated measures ANOVA, which yields a main effect of Validity, $F(2,58) = 11.35, p < .001$, partial $\eta^2 = .28$, and Validity x Block interaction (with Huynh-Feldt correction), $F(18, 522) = 2.22, p < .01$, partial $\eta^2 = .07$.

Recall that there was no effect of cue validity on the average proportion of fixation time to cues during Experiment 1. Fixations to cues show two divergent patterns of visual attention employed in Experiments 1 and 2 (compare Figures 2.1.3 and 2.2.2), where the former is diffuse and the latter is selective. In both experiments learners show equivalent knowledge of the most reliable dimension (equivalent use of the 80% cue). However, learners in Experiment 1 make additional use of the semantically relevant, empirically unreliable information (greater use of the 60% cue).

Similar to the semantic properties of Weather cue in Experiment 1, the location properties of the 50% cue in Experiment 2 had a perceptible influence on the allocation of visual attention during learning. This visual attention bias was not translated into the decision bias: notice the dissimilarities between the choice behavior (Figure 2.2.1) and the eye-movement behavior (Figure 2.2.2) in Experiment 2 during the initial four blocks. Utilization scores suggest that category choices were virtually independent of the 50% cue (as they should be, given that the cue is not diagnostic). Time spent fixating cues, however, suggests a degree of attentional bias toward the unreliable cue, due to its location properties, as opposed to its diagnostic properties. As in Experiment 1, we observe some evidence that visual attention is sensitive to category properties other than dimensional importance. Visual preference for the upper portion of the display is a consistent finding in vision research (Glaholt, Wu, & Reingold, 2010), representing long-standing expectations people have about information order.

Sequential Effects of Error on Odds of Revisiting Most Recently Fixated Cue

As in Experiment 1, the odds of re-fixating a cue immediately after error vs. correct feedback were computed for each learner. The odds are plotted in Figure 2.2.3,

with Blocks 1 through 5 and Blocks 6 through 10 being placed on separate panels. As is apparent from the figure, no sequential effects of feedback on the odds of re-visiting the most recently fixated cue emerged in the early or late phases of learning.

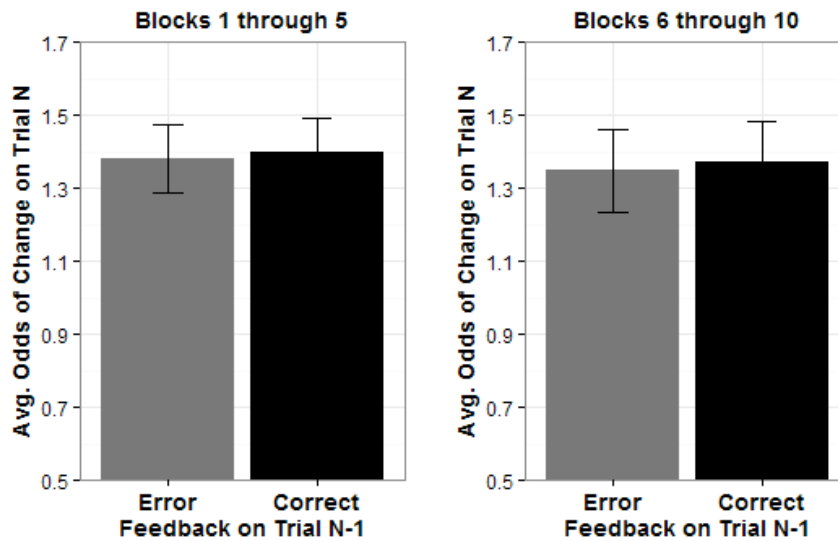


Figure 2.2.3. Comparing the last fixation on trial N-1 with the first fixation on trial N (Change vs. No Change), as a function of feedback (Error vs. Correct) on trial N-1. There is no visible effect of feedback on the odds of fixating the cues in the early or late blocks of trials.

Discussion

Removing the semantic component from the category structure eliminated the sequential effects of feedback on fixations to cues. This outcome needs to be considered in the context of cue utilization (Figure 2.2.1), which showed unambiguous preference for the most reliable cue starting as early as Block 2. Similarly, fixations to dimensions in Figure 2.2.2 show that the 80% valid cue consistently dominated visual attention. Potentially, the high contrast in validity between the 80% cue vs. the other two cues

resulted in early arbitration among the cues, which precluded emergence of the sequential effects of feedback in eye-movements. This outcome seems to support the possibility that sequential effects of feedback in eye-movements reflect cue competition, as suggested in Experiment 1.

Experiment 3 is designed to further ascertain this possibility. If it is the case that sequential effects of feedback reflect cue competition, manipulations of validity contrasts among cues should be associated with the ability to detect sequential effects of feedback in visual attention, to the extent that these manipulations introduce or remove cue competition.

Experiment 3

Effects of cue competition on dimensional utilization have been studied in the contexts of multiple cue probability learning and probabilistic category learning (Castellan, 1973; Edgell et al., 1996; Kruschke & Johansen, 1999). Research suggests that inclusion of additional *relevant* dimension has a greater degrading effect on the utilization of a given *relevant* dimension than an addition of an irrelevant dimension (Edgell et al., 1996). Recall that in the Experiment 2 learners interacted with a category characterized by the 80%, 60%, and 50% reliable cues, where only the 80% valid cue stood out as diagnostic. Both, decision patterns and fixation patterns during Experiment 2 were characterized by a degree of attentional selectivity favoring the most reliable, 80% cue. No sequential effects of feedback were detected in Experiment 2 during early or late trials.

Experiment 3 further addresses the possibility that sequential effects of feedback reflect cue competition by reducing the validity contrast between the most diagnostic and the weakly diagnostic cues. In Experiment 3, the weakly diagnostic 60% cue is replaced with a moderately diagnostic 70% cue, so in contrast to the earlier experiments, Experiment 3 uses the 80%, 70%, and 50% reliable cues. Previous research suggests that this manipulation should reduce utilization of the 80% and the 70% reliable cues. If sequential effects of feedback in fixations reflect cue competition, we should also expect that introducing two moderately predictive cues will bring about the sequential effects of feedback in eye-movements.

Method

Participants

Thirty undergraduate students, with normal or corrected to normal vision, from Georgia Institute of Technology participated in the study and received course credit. Average participant age was 19.8 years ($SD = 1.4$); 66 percent of the sample were males; 70 percent reported English as their first language; 40 percent of the sample were White, 46 percent were Asian; the remaining 14 percent identified themselves as ‘Other’.

Apparatus and Stimuli

Apparatus and stimuli were the same as these described in Experiment 1, with the exception that the feedback schedule was adjusted such that the 60% valid cue was now 70% valid. Cue-to-validity mapping was completely counterbalanced across learners,

forming a total of six versions of a three-cue non-metric category with binary dimensions characterized by 80%, 70%, and 50% valid cues.

Procedure

Procedure and instructions to learners were the same as described in Experiment 2, with the exception that the learners completed six blocks of trials, as opposed to 10 blocks. The rationale for shortening the session was two-fold: (1) no sequential effects were expected after the initial 100 trials; (2) utilization scores tend to reach an asymptote after the initial 100 trials.

Results and Discussion

Cue Utilization

As in previous experiments, examination of learner responses to individual cue values with repeated measures 2(Values) x 3(Validities) x 6(Blocks) ANOVA showed no effect of Value, $F(1, 28) = 1.04$, n.s., and no interactions. Subsequently, utilization scores for the two values of each cue were collapsed, and are reported as a single score for each cue. Utilization scores for the 80%, 70%, and 50% cues are plotted in Figure 2.3.1.

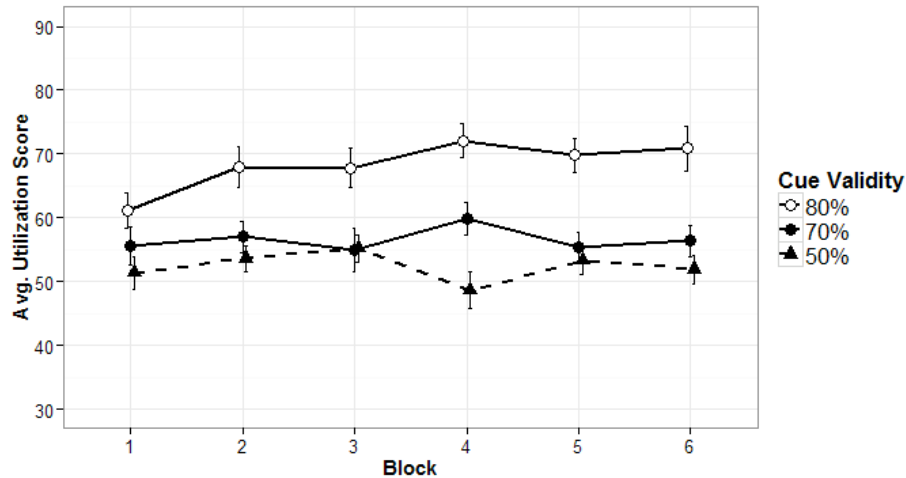


Figure 2.3.1. Average utilization scores in Experiment 3. Cue utilization as a function of cue validity and block shows that learners consistently preferred the 80% valid cue. Note that neither the 80% nor the 70% cue were utilized to the extent of their objective predictive validity.

Average utilization scores suggest that a pattern of cue utilization is more or less established by Block 2 and is maintained for the remainder of the study. Learners consistently underutilized both the 80% and the 70% valid cues, with the more reliable cue showing the largest impact on category selection. This preference for the 80% cue is confirmed by a 3(Validity) x 6(Block) repeated measures ANOVA yielding a main effect of Validity, $F(2, 27) = 15.76, p < .01, \text{partial } \eta^2 = .54$, and no effects of Block or interaction.

Cue Competition

Presence of cue competition in Experiments 1 and 3, but not Experiment 2 would mean that utilization of the 80% cue was commensurate with its objective validity in Experiment 2, but was less than 80% in Experiments 1 and 3. Application of ANOVA to group means did not detect group differences because of large within-group error terms and insufficient power (.31). For this reason, one-sample t-test was chosen to compare

group means at Block 6 against the score of 80 in Experiments 1, 2, and 3. Block 6 was chosen because it represents learning achieved in the final 20 trials of Experiment 3, against which the learning in the other two experiments is compared.

Table 2.3.1 displays Experiment means and t-values, confirming that by Block 6 utilization of the 80% valid cue was less than 80% for learners in Experiments 1 and 3. Utilization for learners in Experiment 2, however, was not different from 80. Taken together, these data represents evidence for cue competition in Experiments 1 and 3, but not Experiment 2.

Table 2.3.1. Comparing Utilization of the 80% Valid Cue across Experiments 1, 2, and 3.

Experiment	Mean	SD	t-score	p-value
Experiment 1	70.17	22.11	- 2.44	.02
Experiment 2	78.83	17.45	- .37	.72
Experiment 3	70.86	19.91	- 2.47	.02

Proportion Trial Time Allocated to Cues

Effects of cue validity and training block on fixation times are plotted in Figure 2.3.2. Average proportions of trial time spent fixating each of the three cues did not change with block or with cue, suggesting a non-discriminative attention profile. In this case, eye-movements did not reflect learner preference for the 80% reliable cue apparent in the cue utilization scores discussed earlier.

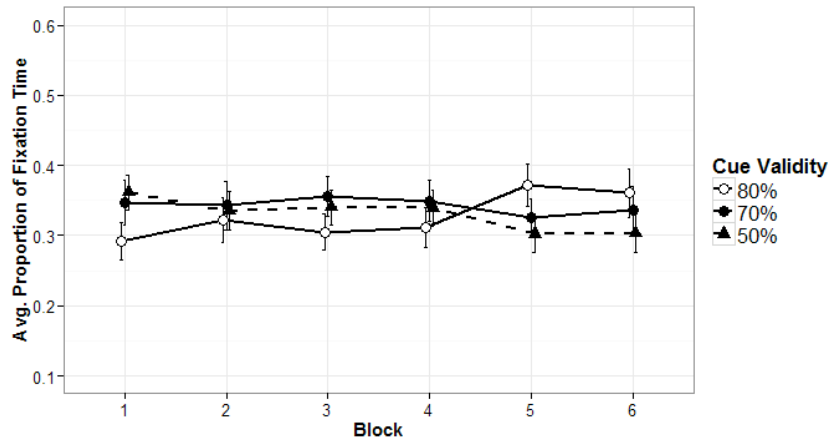


Figure 2.3.2. Average proportion of fixation time to cues in Experiment 3. The time measures suggest that there was no sensitivity to cue validities in eye-movements during Experiment 3.

A comparison of proportion times with a repeated measures 3(Validity) x 6(Block) ANOVA showed no effects of Validity, Block, or Validity x Block interaction.

Effects of Feedback on Odds of Revisiting Most Recently Fixated Cue

The odds of change in location between the last fixation on trial N-1 and the first fixation on trial N are visualized in Figure 2.3.3. The figure shows that (similar to Experiment 1) the odds of change are higher following error feedback as compared to the odds of change following a correct feedback. A paired-samples t-test confirms the difference between feedback conditions, $t(28) = 2.17, p < .05$.

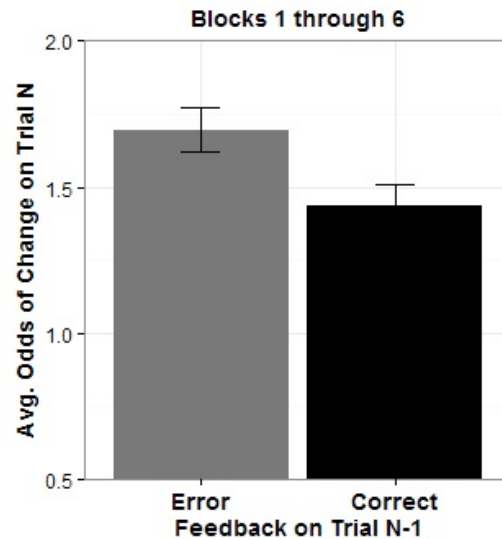


Figure 2.3.3. Odds of change in Experiment 3. Odds of fixating a different cue (change) after error feedback are higher compared to that following correct feedback.

Discussion

Introducing two moderately predictive cues into the category re-introduced the diffuse fixation profile and the sequential effects of feedback previously observed in Experiment 1. Experiment 3 also showed a dissociation between utilization (which consistently favored the 80% valid cue) and fixations to cues (which did not evidence dimensional sensitivity). When combined with the previous two experiments, outcomes of Experiment 3 suggest that: (1) attention during category learning is sensitive to information beyond dimensional validity (e.g., semantic and perceptual category properties); and (2) sequential effects of feedback in eye-movements are associated with presence of cue competition.

CHAPTER 3:

GENERAL DISCUSSION

Because foveated stimuli undergo the most detailed perceptual analysis, fixations represent a limited attention resource allocated serially based on internally generated hierarchy of importance (Hayhoe & Ballard, 2005). Knowing gaze positions, sequences, and routines during a cognitive task, therefore, can reveal something about the way people attend during the task. Eye-tracker variables have been extensively used in research on reading (Rayner, 1998), language comprehension (Barr, 2008), decision making (Orquin & Loose, 2013), attention (Chelazzi, Perlato, Santandrea, & Liber, 2013), and vision in daily tasks (Ballard & Hayhoe, 2009). More recently, the eye-tracker methodology has been adopted by researchers in the area of categorization, as well (Rehder & Hoffman, 2005a). Because eye-tracker data have been used as evidence in theoretical discussions of categorization (e.g., Kim & Rehder, 2011; Kruschke, Kappenman, & Hetrick, 2005; McColeman et al., 2014; Rehder, Colner, & Hoffman, 2009), it is important to examine potential and limitations associated with this behavioral measure of attention. One of the aims in this dissertation is to gain a better understanding of the processes guiding eye-movements during category learning.

Category learning in a feedback-based classification paradigm represents a complex problem-solving behavior, which involves (among other things) learning of probabilities. Effects of statistical learning and reinforcement learning on control of visual attention have been extensively investigated in the context of simple perceptual tasks involving stimulus detection and stimulus discrimination (Chelazzi, Perlato, Santandrea, & Libera, 2013). Effects of probability learning and reinforcement learning

on visual attention in the context of complex reasoning tasks, such as categorization, have not received similar attention.

In perceptual tasks involving selective visual attention (*e.g.*, visual search, flanker, contextual cueing), gaze positions and response times confirm that people are sensitive to statistical regularities imbedded in the task components. This sensitivity facilitates performance and can be measured in eye-movements at both local (trial-to-trial) and global (experiment-wise) levels (*e.g.*, Chelazzi et al., 2013; Ludwig, Farrell, Ellis, Hardwicke, & Gilchrist, 2012).

Neurophysiological studies of visual saccadic decisions in primates suggest that visual systems work closely with the reward systems of the brain (Glimcher, 2003). Recent developments in neurophysiology have revealed that neurons involved in selection of behavioral alternatives have patterns of firing that follow a function similar to that used in mathematical models of reinforcement-learning (Schultz, 2006). Specifically, dopaminergic neurons in the basal ganglia increase firing above the tonic levels following unexpected rewards and decrease firing below the tonic levels when an expected reward is not received. Neuro-computational models of learning concur that these dopamine reward signals generated in the basal ganglia regulate flow of information to and from the cortical areas responsible for behavior representations ranging from simple saccades to complex cognitive operations (Cohen & Frank, 2009).

The basal ganglia represent a group of sub-cortical structures central to both cognitive and motor learning. The basal ganglia are richly interconnected with cortical (frontal eye fields, dorsolateral prefrontal, and lateral intra-parietal) and sub-cortical (superior colliculus) sites responsible for planning and execution of saccades. Given the

extent to which the basal ganglia are involved in category learning and control of movements (for a review, see Seger, 2008), one would expect that eye-movements *should* be sensitive to trial-level feedback during categorization. Encouraging results observed in the context of simple perceptual tasks are commonly generalized to complex cognitive behaviors and are cited as tentative explanations for how these behaviors are formed (e.g., Chelazzi, 2013; Glimcher, 2003; Hayhoe & Ballard, 2005). The general expectation is that fixations in complex cognitive tasks should become progressively more selective in response to feedback, because selective visual attention is ultimately shaped by the error signal. This dissertation represents an effort at generating empirical evidence to this effect.

Eye-Movements and Error Driven Learning

Trial-level evidence for reinforcement-type learning in fixations have been previously described as weak (McClelland et al., 2014), because aggregate measures of attention change showed that, in many cases, fixations were not sensitive to error. In this dissertation, steps were taken to overcome some methodological issues potentially obscuring the evidence for such sensitivity.

In Experiment 1 efforts were undertaken to maximize sensitivity of aggregate measures of fixation times to the effects of feedback. This was accomplished by including a semantic component into an otherwise abstract category, so that the semantic component could serve as a common starting point for the entire group and a reference dimension for subsequent analyses. The expectation was that fixations (shaped by feedback) would gravitate away from the semantically relevant unreliable dimension

toward the abstract reliable dimension. This expectation was not confirmed: The sequential effects of feedback disappeared after the initial 100 trials, while fixations toward the two unreliable dimensions persisted for the remainder of the experiment (additional 100 trials). This suggested that, although present, sensitivity to error in visual attention serves a function other than that of shaping a selective visual profile.

Two subsequent experiments investigated a possibility that the sequential effects of error in fixations reflect a process of cue competition. The conditions for cue competition were removed in Experiment 2 and re-introduced again in Experiment 3. Across all three experiments, the sequential effects of feedback were present when the conditions for cue competition were present. The sequential effects of error were absent, when conditions for cue completion were removed. Collectively, the eye-movement data suggest that the sequential effects of feedback observed in this dissertation are attributable to cue competition, as opposed to the reinforcement-type learning initially hypothesized.

Eye-Movements and Cue Competition

Cue competition frequently characterizes environments with multiple non-metric separable dimensions, where two or more dimensions compete for learners' attention with the result that neither dimension is optimally utilized (Edgell et al., 1996; Kruschke & Johansen, 1999). Cue competition may result in over-utilization of unreliable dimensions and underutilization of reliable dimensions. Utilization data in such experiments typically show that the learning peaks and reaches an asymptote within initial 100 trials or so. This pattern of responding favors speed of learning over precision,

is observed in many species, and is evolutionary adaptive (Kruschke & Johansen, 1999). Note, that this pattern of utilization is especially apparent in Experiment 3, which included 80%, 70%, and 50% valid cues.

A theory best describing the cue competition effects is the RASHNIL - Rapid Attention Shifts in Learning (Kruschke & Johansen, 1999). Among the most important assumptions of the theory is that early learning is characterized by large and rapid shifts of attention away from the dimensions associated with error on a preceding trial. Computationally, these rapid shifts of attention quickly reduce error by minimizing interference from the old learning, such that new learning is rapidly accumulated while the old learning is being preserved.

The RASHNIL also makes an assumption about annealing of learning rates. According to the theory, large adjustments of attention in response to each instance of error do not persist beyond the initial blocks of trials. Subjectively, the learners 'resign' to the inevitability of probabilistic error and maintain their pattern of responding, discounting the error. This assumption allows the model to approximate empirical data optimally, where cue utilization reaches an asymptote relatively early in training before reaching the objective cue validities. There is no unambiguous empirical evidence, however, showing that probabilistic learners discount error after having familiarized themselves with the task (see Craig, Lewandowsky, & Little, 2011). This kind of evidence is not directly accessible via the measures of response probabilities. The indirect way involves changing the dimensional validities in the mid-study, unbeknown to learners, and measuring time it takes the learners to notice the change. Whether this type

of evidence bears direct relation to the processes operating in stationary probabilistic environments is not clear.

Note, that unlike the choice probability data, the eye-movement data allow a direct observation of attention-related behaviors predicted by the RASHNIL, including the error discounting in a stationary probabilistic environment. Rapid trial-level shifts of attention away from the error-causing dimension were observed in eye-movements during Experiment 1, and are clearly visible in Figure 2.1.4, illustrating time to cues allocated during the neighboring congruent and incongruent trials. Consistent with the RASHNIL's predictions, these shifts were not present during the second half of the experiment. Presently, eye-movement data reported in this dissertation represent the most direct empirical evidence for the theoretical assumptions adopted in the RASHNIL.

It is not clear whether the sequential effects of error described in McColeman et al. (2014) are due to cue competition as well. The facts that the effects described in their review tended to decline with block and were sensitive to experimental design variables, suggest that there is a possibility that cue competition for learners' attention may underlie the effects reported in McColeman et al. (2014) as well. Future research should further examine the effects of experimental design variables on the ability to detect the sequential effects of feedback in fixations. If the cue-competition explanation holds, then competing evidence for two or more non-metric dimensions should bring about the sequential effects of feedback in fixations.

Eye-Movements and Cue Utilization

A normatively correct behavior in a probabilistic category is to deterministically choose the most likely alternative (to *maximize*). Consistent with prior research (e.g., Friedman & Massaro, 1998), however, aggregate measures of utilization in this dissertation did not reflect maximization behavior; instead they suggested probability matching. Reasons for probability matching in probabilistic learning are still debated. Probability matching, however, can be substantially reduced (although not entirely eliminated) with reduction of uncertainty. Although uncertainty is not apparent in a binary choice, it becomes transparent in a continuous choice. Comparisons between the binary ('*select A vs. B*') and the continuous ('*how confident are you it is A?*') modes of classification suggest that the binary classification consistently overshoots the continuous in estimating outcome likelihoods. Reductions of uncertainty reduce the differences between the two modes of responding and foster maximization. For instance, probabilistic learners who receive running summaries of category outcomes relative to their choices are more likely to both maximize and to align continuous ratings with the binary choices (Friedman & Massaro, 1998).

Recall that the utilization data in this dissertation reflected preference for the 80% diagnostic cue, while the eye-movement data (particularly in Experiments 1 and 3) reflected diffuse attention. Information conveyed by the time-based measures of eye-movements seem to resemble the information contained in the continuous choice, such that both types of behavior 'under-shoot' objective probabilities, relative to binary choice. If this is the case, then the information about attention conveyed by fixations should be more in line with that contained in the continuous mode of classification, may

vary as a function of continuous classification, and, similar to continuous classification, should be sensitive to manipulations of uncertainty. Future research should certainly address this possibility.

Evidence for Top-Down Control of Eye-Movements

Visual attention profiles that emerged in Experiments 1 and 2 reveal influences of semantic knowledge and, more generally, top-down control of visual attention. Unlike utilization scores, fixations in Experiment 1 were consistently influenced by the weather-prediction theme. Contrary to what was expected, frequent error associated with the semantically relevant, unreliable dimension did not reduce visual attention toward this semantically relevant dimension. Experiment 1 convincingly demonstrates that visual attention in categorization is being guided primarily by the top-down, conceptual representations, not the bottom-up statistical learning.

The influence of long-standing conceptual and perceptual knowledge learners bring (and apply) to a new task is difficult to express in a theoretical model and therefore is not frequently acknowledged. These influences, however, systematically modify new learning. Prior research shows that, when relevant, both perceptual (Kruschke & Johansen, 1999) and conceptual attributes (Love & Markman; Murphy & Kaplan, 2000) associated with predictive dimensions greatly facilitate learning. When such information is associated with unreliable dimensions, it significantly reduces learning of empirically more reliable dimensions. The visual bias toward the unreliable information documented in Experiments 1 and 2 strongly suggests that prior knowledge may influence visual attention more so than it does the binary choice.

In a classification task, it can be shown that optimization of visual attention toward relevant dimensions continues well after the point at which both the error and the feedback have been eliminated (Blair, Watson, & Meier, 2009; Kim & Rehder, 2011). This dissertation illustrates an opposite scenario, where visual attention to unreliable information persists in the face of very frequent error. Taken together, these two observations strongly suggest that eye-movements in complex cognitive tasks are not sensitive to incremental associative learning from feedback. Instead, they reflect a subjective, goal-driven state of a learner, as evidenced in the effects of prior knowledge, long established (heuristic) behaviors, decisional uncertainty, or vacillations between alternative sources of conflicting evidence. Placed in the context of broader research on human vision, data in this dissertation reflect the notion that the best way to understand fixations is to understand what the agent is *doing* (Ballard & Hayhow, 2009).

I suggest that the gradual accumulation of trial-level evidence about statistical properties of the category is not manifest in fixations, until the cumulative weight of small changes shifts the learners' representation of the task in some perceptible to the learner way, so the learner subjectively 'does' the task differently. I suggest that formation of selective visual attention during categorization unfolds at a rate different from that observed in a binary choice. There is some preliminary evidence that visual attention tends to lag a bit behind the binary choice (Rehder & Hoffman, 2005a). Future research should further investigate the nature of this relationship. Specifically, can the two types of behavior be related by a common function? Perhaps other indexes of categorization behavior (such as response time, continuous choice, and subjective report) may bear a more direct relation to the eye-movement behavior. Eye-movements provide

unique and rich information about the way we accomplish cognitive and motor tasks, and how we change our understanding and performance of the task with the developing expertise. The question of how experts arrive at a selective visual profile is far from being answered.

APPENDIX A:
THE TIME PROPORTION SHIFT (TIPS) MEASURE

The Time Proportion Shift (TIPS) score was suggested by McColeman et al. (2014) as a measure of between-trial attention change during category learning. The TIPS score is computed in two steps.

(1) The relative proportion of trial time spent in fixations is computed for each cue as follows:

$$\eta_{i,t} = \frac{\varphi_{i,t}}{\sum_i \varphi_{i,t}}, \text{ where}$$

$\eta_{i,t}$ - proportion time spent on feature i during trial t ;

$\varphi_{i,t}$ - total time spent on feature i during trial t ;

(2) The change in proportion of trial time allocated to cues from trial N-1 to N is computed as follows:

$$\delta_i = \sum_i |\eta_{i,t} - \eta_{i,t-1}|, \text{ where}$$

δ_i - is the TIPS, or the change in proportion of trial time allocated to cues from trial N-1 to trial N.

For an example, consider Table A1.

Table A1. Example of calculating the TIPS. The example shows two trials from a three-cue category, where the change in proportion of trial time (TIPS) from Trial 1 to Trial 2 is 1.4.

Trial	Proportion to Cue 1	Proportion to Cue 2	Proportion to Cue 3	Row Total
(N-1)	.3	.7	0	1
N	.7	0	.3	1
 N – (N-1) 	.4	.7	.3	1.4

REFERENCES

- Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., & Waldron, E.M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442-481.
- Ashby, F.G., & Shawn, W.E. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5).
- Ballard, D.H., & Hayhoe, M.M. (2009). Modeling the role of task on the control of gaze. *Visual Cognition*, *17*, 1185-1204.
- Barr, D.J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457-474.
- Blair, M.R., Watson, M.R., & Meier, K.M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, *112*, 330-336.
- Blair, M.R., Watson, M.R., Walshe, R.C., & F. Maj (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1196-1206.
- Botvinick, M.M., Braver, T.S., Barch., D.M., Carter, C.s., Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624-652.
- Braver, T.S., & Cohen, J.D. (2001). Working memory, cognitive control, and the prefrontal cortex: Computational and empirical studies. *Cognitive Processing*, *2*, 25-55.
- Castellan, N.J., Jr. (1973). Multiple-cue probability learning with irrelevant cues. *Organizational Behavior and Human Performance*, *9*, 116-29.
- Chelazzi, L., Perlato, A., Santandrea, E., & Libera, Ch.D. (2013). Rewards teach visual selective attention. *Vision Research*, *85*, 58-72.
- Chen, L., Meier, K. M., & Blair, M. R. (2013). Temporal characteristics of overt attentional behavior during category learning. *Attention, Perception, and Psychophysics*, *75*, 244-256.
- Cohen, M.X., & Frank, M.J. (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. *Behavioral Brain Research*, *199*, 141-156.

- Craig, S., Lewandowsky, S., & Little, D. (2011). Error discounting in probabilistic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 673-687.
- Droll, J.A., & Hayhoe, M.M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1352-1365.
- Edgell, S.E., & Morrissey, J.M. (1987). Delayed exposure to additional relevant information in nonmetric multiple cue probability learning. *Organization Behavior and Human Decision Processes*, 40, 22-38.
- Edgell, S.E., Castellan, N.J., Roe, R.M., Barnes, J.M., Bright, R.D., & Ford, L.A. (1996). Irrelevant information in probabilistic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1463-1481.
- Friedman, D., & Massaro, D.W. (1998). Understanding variability in binary and continuous choice. *Psychonomic Bulletin & Review*, 5(3), 370-389.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23, 523-552.
- Glaholt, M.G. & Reingold, E.M. (2011). Eye movement monitoring as a process-tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, 4(2), 125-146.
- Glaholt, M.G., Wu, M., Reingold, E.M. (2010). Evidence for top-down control of eye movements during visual decision making. *Journal of Vision*, 10(5):15, 1-10.
- Glimcher, P.W. (2003). The neurobiology of visual saccadic decision making. *Annual Reviews of Neuroscience*, 26, 133-179.
- Haider, H. & Frensch, P.H. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *JEP: Language, Memory, and Cognition*, 23(1), 172-190.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188-194.
- Hazy, Th.E., Frank, M.J., & O'Reilly (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience and Biobehavioral Reviews*, 34, 701-720.
- Hoffman, A.B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *JEP: General*, 139(2), 319-340.

- Holroyd, C.G., & Coles, M.G.H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Reviews*, *109*, 679-709.
- Jones, M., Love, B.C., & Maddox, W.T. (2006). Recency effect as a window on generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 316-332.
- Jones, M., & Sieck, W.R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 626-640.
- Kibbe, M.M., & Kowler, E. (2011). Visual search for category sets: Tradeoffs between exploration and memory. *Journal of Vision*, *11*(3), 1-21.
- Kim, Sh.W., & Rehder, B. (2011). How prior knowledge affects selective attention during category learning: An eyetracking study. *Memory & Cognition*, *39*(4), 649 – 665.
- Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.
- Kruschke, J.K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812-863.
- 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.
- Kruschke, J.K., Kappenman, E.S., & Hetrick, W.P. (2005). Eye-gaze and individual differences consistent with learned attention in associative blocking and highlighting. *JEP: Learning, Memory, and Cognition*, *31*(5), 830-345.
- Lamberts, N. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 695-711.
- Land, M.F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience*, *3*, 1340-1345.
- Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*(2), 376-400.
- Love, B.C., & Markman, A.B. (2003). The non-independence of stimulus properties in human category learning. *Memory & Cognition*, *31*(5), 790-799.

- Luce, R. D. (1959). *Individual Choice Behavior*. New York: Wiley.
- Ludwig, C.J.H., Farrell, S., Ellis, L.A., Hardwicke, T.E., & Gilchrist, J.D. (2012). Context-Gated statistical learning and its role in visual-saccadic decisions. *Journal of Experimental Psychology: General*, *141*(1), 150-169.
- McColeman, C.M., Barnes, J.I., Chen, L., Meier, K.M., Walshe, R.C., & Blair, M.R. (2014). Learning-induced changes in attentional allocation during categorization: A sizable catalog of attention change as measured by eye movements. *PLoS ONE*, *9*(1).
- Murphy, G.L., & Kaplan, A.S. (2000). Feature distribution and background knowledge in category learning. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *53A*(4), 962-982.
- Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104-114.
- Nosofsky, R.M., Stanton, R.D., & Zaki, S.R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory and Cognition*, *33*(7), 1256-1271.
- Orquin, J.L., & Loose, S.M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*, 190-206.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.
- Rehder, B., Colner, R.M., & Hoffman, A.B. (2009). Feature inference learning and eyetracking. *Journal of Memory and Language*, *60*, 393-419.
- Rehder, B., & Hoffman, A.B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 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), 311-829.
- Ridderinkhof, K.R., van den Wildenberg, W.P.M., Segalowitz, S.J., Carter, C.S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*, 129-140.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Reviews of Psychology*, *57*, 87-115.

- Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their role in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, 32, 265-278.
- Seger, C.A., & Miller, E.K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, 33, 203-219.
- Shepard, R. N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22, 325-345.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75(13), 1-42.
- Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin*, 138(1), 102-125.
- Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.