2018

Categorical Change: Exploring the Effects of Concept Drift in Human Perceptual Category Learning

Andrew Wismer
University of Central Florida

CATEGORICAL CHANGE: EXPLORING THE EFFECTS OF CONCEPT DRIFT IN HUMAN PERCEPTUAL CATEGORY LEARNING

by

ANDREW JACOB WISMER
B.A. Messiah College, 2012
M.A. University of Central Florida, 2015

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2018

Major Professor: Corey J. Bohil

ABSTRACT

Categorization is an essential survival skill that we engage in daily. A multitude of behavioral and neuropsychological evidence support the existence of multiple learning systems involved in category learning. COmpetition between Verbal and Implicit Systems (COVIS) theory provides a neuropsychological basis for the existence of an explicit and implicit learning system involved in the learning of category rules. COVIS provides a convincing account of asymptotic performance in human category learning. However, COVIS – and virtually all current theories of category learning – focus solely on categories and decision environments that remain stationary over time. However, our environment is dynamic, and we often need to adapt our decision making to account for environmental or categorical changes. Machine learning addresses this significant challenge through what is termed concept drift. Concept drift occurs any time a data distribution changes over time. This dissertation draws from two key characteristics of concept drift in machine learning known to impact the performance of learning models, and in-so-doing provides the first systematic exploration of concept drift (i.e., categorical change) in human perceptual category learning. Four experiments, each including one key change parameter (category base-rates, payoffs, or category structure [RB/II]), investigated the effect of rate of change (abrupt, gradual) and awareness of change (foretold or not) on decision criterion adaptation. Critically, Experiments 3 and 4 evaluated differences in categorical adaptation within explicit and implicit category learning tasks to determine if rate and awareness of change moderated any learning system differences. The results of these experiments inform current category learning theory and provide information for machine learning models of decision support in non-stationary environments.

This dissertation is dedicated to my loving wife, Laura, with whom this could not have been possible, to our precious son, Elijah, and to my Lord and Savior Jesus Christ.

ACKNOWLEDGMENTS

This work is the culmination of overwhelming support from family, friends, and colleagues. In particular, I would like to thank my advisor, Dr. Corey Bohil, for his unwavering support, constructive feedback, and patience. I am also grateful for the undergraduate research assistants of the Categorization Decision Lab who helped collect the data, as well as the graduate students who helped facilitate and coordinate data collection. Finally, I would like to thank my parents and other family members for encouraging me to persevere and strive for my best.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Lucretius, a Roman poet and philosopher, once said, “The only constant in life is change”. Change is expected in the dynamic environments in which humans live, learn, and interact. Consider the field of medicine. Both the context in which diseases are diagnosed and treated, as well as the diseases themselves, are non-stationary. The introduction of new technologies changes our understanding of diagnosis and treatment. In addition, the characteristic symptoms and signs of a disease may evolve over time (e.g., symptoms of Parkinson’s Disease are varied – from symptoms affecting one’s ability to move, to pain, loss of the sense of smell, and dementia. These symptoms tend to change slowly over time, and new symptoms can suddenly surface; Ratini, 2017). Disease frequencies rise and fall through the seasons and years. The resources available to diagnose and treat disease, and the costs and benefits of diagnoses can change as the severity of the disease shifts. As a result, diagnostician training must include the training of a flexible learning process that is able to adapt to change in complex environments. To train proficient, adaptable diagnosticians, a thorough understanding of the characteristic influence of non-stationarity on the learning process is needed.

Non-stationarity has been extensively studied in machine learning via the idea of “concept drift”: a phenomenon describing change in the statistical properties of a target concept. Concept drift is an important challenge in machine learning, and there have been many efforts to characterize the ways in which concept drift can undermine the predictive accuracy of models. Concept drift characterizations have yet to be applied to human categorization, where the notion of categorical change is only beginning to be explored. The question of how humans detect and/or adapt to categorical change has important applications both for informing categorization theory and machine learning models of concept drift. This dissertation intends to systematically explore the influence of non-stationarity on human perceptual category learning with inspiration from key concept drift characterizations in machine learning.

Statement of the Problem

How do humans react to categorical change in a perceptual category learning task? Categories can change in any number of ways: from a change in decisional variables (the relative frequencies of the categories or the relative costs and benefits of classification) to a structural or sampling change in the features of categories. Additionally, this change can occur abruptly (from one trial to the next) or gradually (incrementally over a longer series of trials), and learners can be made aware of the change before it occurs or be left to discover it on their own. While there is some evidence for the fluent adaptation of humans to environmental change (Franks & Hicks, 2016; Schertz, Cho, Lotto, & Warner, 2016), there are also many cases of maladaptation (Ackermann & Landy, 2015; Norton, Fleming, Daw, & Landy, 2017). The question remains as to what types of change humans can readily adapt to, and if adaptation requires initial conscious detection of the change to succeed.

Purpose of the Research

The purpose of this research is to systemically explore and characterize the influence of categorical change in human perceptual category learning. This research will explore three characteristics of categorical drift – type, rate, and awareness of change – each drawn from primary concept drift distinctions in machine learning. This research will lead to a more comprehensive understanding of change detection and adaptation in human categorization, filling in a few important gaps in the literature regarding transfer and generalization of category learning.

Significance of the Research

Concept drift is an important challenge in machine learning (Ditzler, Roveri, Alippi, & Polikar, 2015; Gama, Medas, & Rodrigues, 2004; Tsymbal, 2004; I. Žliobaitė, 2010). However, concept drift is not restricted to machine learning contexts; natural categories and environments are also vulnerable to sources of change (Helie, Ell, Filoteo, & Maddox, 2015; Summerfield, Behrens, & Koechlin, 2011). For example, radiologists make decisions about the classification of tumors (i.e., is additional testing needed to determine whether a tumor is malignant?). These decision environments are dynamic with potential changes in the patient population, x-ray clarity, or costs and benefits over time (Brown & Steyvers, 2005). Despite the possibility for categories to change, almost all category learning research uses stationary categories and stationary statistical properties (Bohil & Maddox, 2003; Helie et al., 2015; Maddox & Bohil, 1998).

Extending our knowledge of category learning into nonstationary contexts may help further refine theories of categorization. For example, categorization theories that propose separate learning systems in the brain suggest differential performance dependent on the primary learning system engaged (Ashby, Alfonso-Reese, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994). A thorough exploration of categorical change in human perceptual category learning should provide evidence for the single vs. multiple systems debate in categorization, further informing theory regarding the process in which novel categories are learned. In addition, this research should inform machine learning models that can gain from a better understanding of human cognitive functioning amid change, providing insights for enhancing decision aids for non-stationary environments.

Theoretical Perspective

The overarching theoretical perspective for the current research is one of criterion learning in perceptual classification. In classification, observers first learn the rule that discriminates between two or more categories. This comes down to determining the appropriate dimension or dimensions of interest (e.g., categorize lines according to their orientation while length of the lines can be ignored). Once the appropriate rule is learned, the observer must learn how to group categories using the appropriate dimension(s) (e.g., acute vs. obtuse angles). With one-dimensional stimuli, the dimension of interest is obvious and the process of setting a criterion equates to the classic signal detection task (Green & Swets, 1966; Macmillan & Creelman, 2004; Wickens, 2001) with the criterion corresponding to signal detection beta (β). With two-dimensional (or higher dimension) stimuli, criterion setting can be viewed from a multidimensional extension of signal detection theory. General Recognition Theory (GRT; Ashby & Townsend, 1986) is an extension of signal detection theory (SDT) into multidimensional perceptual space. GRT assumes that observers set a decision criterion that separates the perceptual space into separate response regions; an observer responds "category A" if the percept of a stimulus falls on one side of the criterion and "category B" if it falls on the other. Learning in the context of this theoretical perspective then comes down to how well observers can shift their decision criteria to maximize performance (i.e., accuracy or reward) in light of categorical changes.

The current research is also informed by the theoretical perspective of COVIS – a neuropsychological theory of categorization called COmpetition between Verbal and Implicit Systems (Ashby et al., 1998). COVIS provides a neuropsychological interpretation to general recognition theory by postulating that there are two separate learning systems involved in category learning. The explicit learning system learns easily verbalizable rules through conscious hypothesis testing and is mediated by the prefrontal cortex. The implicit learning system, guided by procedural learning, learns rules that are difficult to verbalize and typically require an integration of stimulus dimensions at a pre-decisional (i.e., perceptual) stage. The implicit learning system is mediated by the basal ganglia and is guided by the release of dopamine through immediate feedback (Ashby, Paul, & Maddox, 2011). COVIS provides a foundation through which to assess any differences in change performance dependent on the primary learning system engaged.

Research Method

In the present research, concept drift in human categorization is studied through perceptual category learning tasks in which observers learn novel categories by completing a series of training blocks. This design enables one to evaluate how novel categories are represented by observers who have no previous experience with the artificial categories. Mid-way through the experiment, one aspect of the categorization task changes (category base-rates, payoffs, or category structure [means]). Primary performance measures include accuracy (reward) and criterion values (β from SDT or GRT modeling), with an emphasis on the immediate and long-term effects of change. The immediate effect of change on performance is analyzed by comparing performance on the first block after change to the last block before change. The long-term effects of change on performance are analyzed by comparing performance on the final block of the experiment to performance on the first block of trials following the change. These measures are compared within different types of change (base-rate, payoff, category structure), and across factorial conditions of rate of change (abrupt, gradual), and awareness of change (alerted, discovered). GRT Models are also applied to observers' data to get a better sense of the strategies people use to adapt to change and the relative timing in which criterion shifting occurs.

Definition of Key Terms

Concept drift: In machine learning, concept drift occurs when there is a change in the statistical properties of a target concept within a data stream that require a change in decision criteria and sometimes imply a change in category boundary.

Non-stationarity: Non-stationarity describes a process or state that is not stationary but rather changes in some way(s) over time.

Type of change: Changes can occur in any number of aspects of a data stream (in machine learning) or perceptual category (in human classification). Changes might occur in the base-rate presence of a category, in the costs or benefits associated with classification, or in the category structure/features.

Virtual concept drift: This type of drift occurs when a data distribution changes without changing the posterior distribution. Data distribution changes can include changes in category base-rates, payoffs, or distributions of the categories. When virtual concept drift occurs, the learning model must be retrained to adequately classify the data (Hoens, Polikar, & Chawla, 2012; Widmer & Kubat, 1993). While virtual concept drift does not change category boundaries, it does impact the optimal decision criterion.

Rate of change: Rate of change, in this dissertation, refers to the length of time it takes in full for a parameter to change from its starting value to its ending value (i.e., the parameter of the task that is non-stationary). Here, only incremental rates of change are considered – ones that change in equal increments at each time step in-between the starting and ending values.

Awareness of change: Awareness of change refers to whether the learner is made aware of the impending change to the data stream or perceptual category (i.e., alerted) or whether that learner must figure out that something has changed on his or her own (i.e., discovered). It is possible that discover conditions may never be better than alerting; but a lack of a difference between the two provides useful information about what is required for adaptation.

Expected uncertainty: Expected uncertainty is uncertainty that is known to be present in predictions. In other words, it is the type of uncertainty that one expects in a familiar environment due to unreliability of predictive cues (Yu & Dayan, 2005).

Unexpected uncertainty: Unexpected uncertainty arises when sensory observations deviate significantly from top-down expectations in an environment. In other words, a context change occurs that signals strongly unexpected observations, thereby alerting the individual that the context has changed (Yu & Dayan, 2005). However, unexpected outcomes can also be the result of persistent environmental stochasticity rather than a context change (Nassar, Wilson, Heasly, & Gold, 2010).

Base-rates: The relative frequency of categories (i.e., a priori probability).

Payoffs: The relative benefits and costs of correct and incorrect responses. Payoffs are often expressed in terms of a payoff matrix showing point values for the possible outcomes (e.g., in a two-alternative forced choice scenario, payoff matrices include point values for hits, false alarms, correct rejections, and misses).

Decision Criterion: The assumed cutoff value a person uses in deciding between one category and another. Decision criteria are dependent on category base-rates and payoffs. Decision criteria are typically modeled in signal detection theory as response bias (β).

Delimitations

The present work considers human perceptual category learning of novel categories by novice observers. Therefore, it does not consider learning of non-human animals, human concept learning (e.g., the learning of abstract concepts such as morality or justice), expert performance, or well-established categories (Ashby & Maddox, 2005). What it does is help to fill in the research gaps related to the transfer of category knowledge and the detection and adaptation of categorical change. This research provides systematic manipulation of variables such as awareness of change, rate of change, and the primary learning system (implicit, explicit) involved. This research shows people's ability to adjust their decision rule in response to change, which relies on a change in decision criterion rather than a change in perceptual representation. The manipulation of category structure may, however, provide some insight into the flexibility of category rule representation with changes in perceptual aspects of a category.

Limitations of the Research

This research is limited to use of a convenience sample of undergraduate students from the University of Central Florida and members of the surrounding community. Participants received no form of compensation beyond class credit. Because performance was not rewarded by any form of (additional) compensation, few/more conservative criterion changes could result if participant motivation was low. However, previous criterion learning research has been successful without monetary compensation (Brown & Steyvers, 2009; Helie et al., 2015; Wismer & Bohil, 2017). The stimuli used in the different experiments were low-dimensional, artificial stimuli that can be conceived of as non-interesting by participants. In addition, several aspects of concept drift and non-stationarity were left unexplored; for example, multi-variable change (e.g., base-rates and category structure together) or multiple and/or recurring changes.

Summary

Categorical change is an important topic not only in machine learning, but also in human learning. Ambiguous perceptual categories, such as those represented by tumors on x-rays or unsafe items in baggage screening, can shift over time. The frequency of high-profile items may increase during a period of time, or screeners may be warned to be less conservative in selecting bags to search individually. Changes to base-rates or payoffs shift category representations, changing the optimal decision strategy in cases of ambiguity. Even categories themselves can change, such as when the evolution of technology results in new representations of x-ray images, or when diseases take on new symptom profiles over time. Category non-stationarity is an under-researched but important area of study that can benefit from systematic application of concept drift characterizations from machine learning.

CHAPTER 2: LITERATURE REVIEW

The notion of concept drift is derived from machine learning classification. An abundance of research has investigated the best methods to handle different types of concept drift in data streams. These findings will be reviewed in the next section. Then, the current state of human category learning will be addressed, highlighting the predominant focus on a dual systems approach. Finally, key characteristics of concept drift in machine learning will be discussed for their application to human category learning, with a review of the sparse human category learning literature relevant to the selected characteristics of concept drift. This literature review will identify the primary gaps in our understanding of categorical change and accordingly end by describing the ways in which this dissertation aims to fill some of these research gaps.

Machine Learning and Concept Drift

Within the past 10-15 years, researchers have paid increased attention to the issue of drifting concepts in machine learning and predictive analytics (Ditzler et al., 2015; Indrė Žliobaitė, Pechenizkiy, & Gama, 2016). In these fields, researchers apply predictive models to datasets in an effort to reliably predict the value of some target variable. For example, one common use of learning algorithms is to understand and predict user preference for a website or service. Recommendation systems try to determine what a user will find interesting by building a user interest profile from an initial learning experiment. In typical applications, a learning algorithm is trained on a large set of past, historical data (i.e., training data) on user preferences and choices, and the model is then applied to novel information (i.e., test data) to predict if a user will find a particular item interesting (Webb, Pazzani, & Billsus, 2001). That is, the model is a classifier (it is essentially a model of the human's classification tendencies – it classifies "interesting things" and "non-interesting things"). Many different forms of learning algorithms exist that can, with a small degree of error, successfully make these predictions.

A primary challenge to the performance of these predictive models is the fact that target variables (such as user preferences) are fluid and prone to change (Webb et al., 2001). For instance, although someone may have always preferred Ford vehicles in the past, a positive experience with a Honda rental car on vacation may lead to an increased preference for Honda vehicles. As a second example, a model may be able to predict with high accuracy the brand of a car given a set of inputs describing its features (e.g., high MPG, specific safety features, etc.), but if new car models are released with a different set of standard features, the model would need to be retrained in order to maintain predictive accuracy. Thus, in order to minimize prediction error when data distributions can change, learning algorithms must incorporate the possibility of non-stationarity. In fact, it is not only concept membership change but also change in decisional factors that are included in concept drift (e.g., the availability of Honda vs. Ford rental vehicles - or the cost of selecting a rental vehicle that is not desirable – either of which may change over time).

This problem of non-stationary data distributions is termed the problem of "concept drift". Most recently, a concept has been given a probabilistic definition based on Bayes' Theorem (Webb, Hyde, Cao, Nguyen, & Petitjean, 2016). A concept involves the prior class probabilities (i.e., concept base-rate), P(Y), and class conditional probabilities (i.e., probability of feature X given concept Y), P(X|Y). The joint distribution P(X,Y) is determined by P(Y) and P(X|Y); therefore a concept = P(X,Y). Because concept drift emphasizes the idea that distributions can change over time, a concept is identified with a timestamp as $P_t(X, Y)$ to denote the concept definition at time *t* (Webb et al., 2016). Concept drift occurs when a concept $P_t(X,Y)$

does not equal $P_u(X,Y)$ - the concept at a new point in time. The probabilistic definition of concept drift motivates the investigation of different types of concept drift explored in this dissertation, which will be detailed in later sections.

Approaches

There are two primary approaches to dealing with the issue of concept drift: active and passive learning. In active learning approaches, learning models are created from past, historical data. These same (unaltered) models are applied to classify new inputs without changing any model parameters. Only if a "change detection" mechanism in the algorithm determines that concept drift is present/has occurred does the algorithm employ an adaptation mechanism. Otherwise, the same model is used to classify new inputs. During the adaptation mechanism is employed, the model either updates itself based on inspected features from the data or the classification errors, or it rebuilds itself; but this only occurs when change is detected. This method may be thought of as "detect and react" (Ditzler et al., 2015). Examples of active approaches include hypothesis tests, change-point methods, sequential hypotheses tests, and change-detection tasks. More details about these techniques can be found in Ditzler et al. (2015) or I. Žliobaitė (2010).

Alternatively, passive learning approaches do not depend on the detection of concept drift. Instead, model parameters are continually updated with each new data input, allowing concept drift to automatically revise model parameters. Passive approaches can help to avoid misses or false alarms (that are more likely in an active approach) by maintaining an up-to-date model at all times. Methods for passive approaches typically fall under single classifier models or ensemble classifier models. Again, more details regarding these techniques can be found in Ditzler et al. (2015) or I. Žliobaitė (2010).

By nature of the adaptation process (detect then react, vs. continual adaptation), active and passive approaches differ in their ability to handle changes of different time courses. Active approaches, with their requirement of change detection before adaptation, tend to perform best in environments with abrupt changes that occur quickly from one data or time point to the next. Passive approaches, given their nature of continuously adapting to each new input, perform best in environments with more gradual, slower changes. Active approaches have a harder time recognizing change that occurs more slowly over a gradual period of time (thus never triggering a change in model parameters or triggering after an extended time period), while passive approaches can give too much weight to outliers and inputs that are highly variable (Ditzler et al., 2015).

In machine learning, it is possible for both active and passive approaches to be successful in application. However, characteristics of the particular application can drive the choice of approach style. Depending on dynamics of the learning scenario such as magnitude, rate of change, or computational resources available, one approach tends to be preferable. Active approaches tend to do slightly better with higher magnitude, abrupt changes (that more easily trigger a major change to model parameters) while passive approaches can better handle more slowly changing parameters through continual parameter updating (Ditzler et al., 2015). However, both active and passive approaches can handle changes of different magnitude and rates fairly well as a whole.

Characterizations

One distinction commonly made in machine learning is the division between “real” and “virtual” concept drift. In general, researchers posit real concept drift occurs when a change takes place in the actual class boundary between concepts, resulting in a change in the posterior probability [p(C|X), the probability of concept C given feature X] of a concept (Webb et al., 2016). Virtual concept drift, on the other hand, is given different definitions in the literature. Most often, virtual concept drift is said to occur when a concept appears to be changing, but it is only due to a change in the frequency (i.e., base-rate) of features or classes themselves, without a change in the likelihood of feature X given target concept C (Webb et al., 2016). To facilitate translation to human category learning, virtual concept drift in this dissertation is taken to include any change that results in a shift in the optimal decision criterion (as opposed to shift in class boundaries which would occur when actual category structure changes). It is when perceptual sensitivity of categories is low that virtual concept drift drives critical changes in response strategies. An example of virtual concept drift is when the definition of breast cancer remains unchanged (the underlying etiology and symptoms), but the frequency of people with breast cancer in a population increases or decreases driving a new for a shift in response strategy.

The difference in proposed definitions of virtual concept drift exemplifies the common critique that many researchers rely too strongly on qualitative distinctions of concept drift that vary widely between researchers. For this reason, there has been a strong push in recent years to establish quantitative measures of concept drift that can be universally applied to different domains in which concept drift might be expected. For example, Webb et al. (2016) highlight that concept drift definitions tend to be qualitative, vague, and inconsistent. To this end, they argue that quantitative measures of concept drift are necessary to fully understand the problem of concept drift in non-stationary distributions and enable comparison across concept drift research. Webb et al. (2016) provide a taxonomy of key quantitative measures of concept drift. The authors' definitions are embedded in the framework of data stream classification; however, they are intended to apply much more broadly, and they apply particularly well to human category learning.

These quantitative measures include magnitude, duration, path length, and rate of drift.

1. Magnitude = D(*t, u*); the distance between old and new concepts, where t and u are the two time points and D is the distance between them.
2. Duration = *u – t;* length of time over which drift occurs.
3. Path Length = cumulative deviation during the period of a drift (for path length formula, see Webb et al., 2016). For example, the "gradual" path in Figure 1 below has a longer path length than the "sudden" path. Path length incorporates elements of the duration and magnitude of drift, and increases with drift that "gradually" shifts between two concepts before settling on the new concept.
4. Average drift rate = $PathLen_{t,u}/(u-t)$; average rate at which the distribution is changing at time *t*.

Webb et al. (2016) choose not to specify the particular distance function for the magnitude calculation because they believe this function varies by domain. The remaining quantitative measures of concept drift follow from the choice of distance function regarding change magnitude.

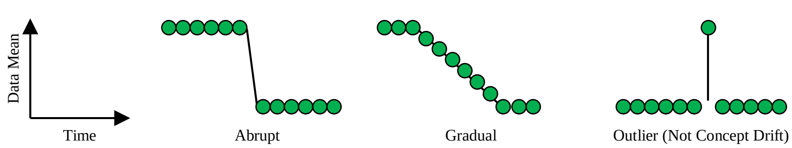

Figure 1. Different patterns of change in data distributions. Outlier is not concept drift.

From the literature on concept drift in machine learning, two common themes/variables can be seen: 1) The type of concept drift can be described by which component of the formula changes (e.g., base-rate of the concept), 2) the optimal approach to handling concept drift (active vs. passive) largely depends on characteristics of the change such as the rate of change (i.e., abrupt vs. gradual). The themes stemming from this literature review raise the question of potential similarities between machine and human learning in the presence of concept drift. Do these same concept drift distinctions matter in human learning, and if so, will they shed light on the different approaches by which humans adapt to changing categories/decision parameters?

Human Category Learning

Was that sound a gunshot or a car backfire? Is this snake poisonous or not? Is it safe or unsafe to merge lanes now? All animals – human and non-human – assign objects and events in the environment to separate categories or equivalence classes. The partitioning of objects into separate categories allows for distinct responses to be made. If the sound was a gunshot, I need to get to a safe place as quickly as possible. If the sound was car backfire, then I do not need to worry about it. The ability to successfully categorize environmental objects and events is an essential survival skill (Ashby & Maddox, 2005).

Given the crucial role categorization plays in everyday life, it is no surprise there exists a vast and wide literature on the processes underlying category learning (Ashby & Maddox, 2005). However, almost all research has focused on stationary categories in stationary environments with unchanging statistics (e.g., Bohil & Maddox, 2003; Helie et al., 2015; Maddox & Bohil, 1998). While such methods have broadened our understanding of the processes involved in category learning, they do not help explain a lot of the types of category learning that occur in the environment. Categories are not learned in a vacuum outside of all environmental influences.

Rather, category learning occurs in environments prone to change, and this change has an effect on an organism's adaptation to survive.

Take the example of classifying a vehicle merging situation as safe or unsafe. The appropriate speed one drives at and distance behind the next car change in accordance with changes in the weather. What classifies as a "safe" speed in sunny weather is different than in rainy or snowy weather, for example. Therefore, a southerner who tries to drive in a northern state for the first time will have to transfer his or her knowledge to match the shifted conditions regarding safe and unsafe speeds/distances. Situations such as these that are prone to change from any number of variables exemplify the need for us to understand the processes by which humans can adapt to changing conditions. (Note: this type of scenario also reflects the need for an individual to select the appropriately matched rule to situation. That is, a person may have learned the appropriate definitions of safe driving distances in both sunny and rainy weather, but needs to appropriately apply this knowledge in each context. As a beautiful, sunny day slowly shifts into a drizzle, rain, and then torrential down-pour, drivers need to increase their following distance and decrease driving speed over time).

Before delving any deeper into the nature of non-stationary category learning, some boundary conditions for the type of category learning that is being discussed must be delineated. There are three primary restrictions to the scope of the present work. First, this work is restricted to human category learning only and will not consider non-human animals. Second, since the interest is in the learning of categories, the current work will involve novices with no experience with the categories prior to the experiment so that these observers might be trained to asymptotic performance. As such, this type of category learning involves the learning of new – never-before-learned – categories, rather than the integration of new information into previously learned categories. Finally, categorization can encompass a variety of things, but the scope of the current work will be limited to perceptual category learning rather than abstract concepts. By this, I mean the type of categorization required to classify, for example, different varieties of skin moles (items that have observable perceptual features) rather than the ability to classify an item as “interesting” or “non-interesting” (Ashby & Maddox, 2005). These restrictions are necessary for the impact of this work.

Single System Theories

A primary distinction within category learning theories is the division between single and multiple systems of learning. Two of the most popular single system theories are Exemplar (Medin & Schaffer, 1978; Nosofsky, 1987) and Prototype Theory (Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968). Exemplar theory states that people compare an item to all category exemplars held in memory and respond with the category to which a new item is most similar. Prototype theory is similar but instead says that people only hold the “prototype” or average category member in memory and compare each new item to this prototype to determine category membership. Within human perceptual category learning, single system theorists, such as Exemplar and Prototype Theory, propose that one system can account for the host of category learning phenomena that research uncovers (Nosofsky & Johansen, 2000; Nosofsky & Zaki, 1998; Poldrack, Selco, Field, & Cohen, 1999). That is, only one system is required to learn all types of categories (e.g., based on individual memories for experienced category instances).

Multiple Systems

The debate between single and multiple-systems of category learning stems from a broader debate within cognitive science, where the majority viewpoint is that of multiple-systems of memory/learning. Indeed, many categorization researchers have posited a role of distinct learning systems, engaged to varying degrees in different circumstances (Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994), with further support from neuroscience (Ashby et al., 1998; Poldrack & Foerde, 2008; Smith & Grossman, 2008) and decision research (Kahneman, 2011). As a whole, burgeoning evidence has been given for the existence of multiple, qualitatively distinct, processes of category learning (i.e., multiple learning processes). At this point, the prominence of multiple systems far outweighs single system theories, although single system theories still provide an impetus for much of the current research in category learning (Ashby & Ell, 2002).

COVIS (COmpetition between Verbal and Implicit Systems) is one prominent theory of multiple category learning systems. COVIS is a neuropsychological theory of category learning that posits competition between separate learning systems – an active explicit learning system and a more passive implicit learning system. Ashby et al. (2011) claim COVIS is the only categorization theory rooted in neuroscience linking processes to distinct physical structures in the brain, and it is assumedly still the case today. The explicit learning system is believed to be mediated by the prefrontal cortex and is associated with verbalizable rules relying on working memory. For example, the rule "If it is tall it is a tree; if it is short it is a bush" is a verbalizable category rule learned by the explicit system. Category learning usually (but not always) begins with the explicit system attempting to find a verbalizable rule to describe the categories. When it consistently fails, the implicit system begins to dominate responding. The implicit learning system, mediated by subcortical structures (e.g., the dopamine-reward learning system in the basal ganglia) slowly learns stimulus-response associations over time. Implicit learning is believed to be based on procedural learning, meaning people learn a response generation procedure without accessing any exemplar memories during categorization (Ashby et al., 1998).

Implicit rules can be learned; however, they are typically non-verbalizable or, at least, quite difficult to verbalize (such as when one needs to integrate two or more dimensions that are typically difficult to verbalize). COVIS posits that there is competition between learning systems until one system does well and begins to determine most responses (Ashby & Maddox, 2005).

COVIS is only one multiple systems theory of category learning (see also ATRIUM; Erickson & Kruschke, 1998), but it does well to describe asymptotic categorization performance (Ashby et al., 1998). Multiple systems theories have been compared against the longstanding single system theories of Exemplar and Prototype theory. COVIS (and other decision bound theories) differ from Exemplar and Prototype in that COVIS posits people gradually learn to associate responses with select regions of perceptual space and respond deterministically (i.e., rule-based), while single system theories involve similarity comparisons of new stimuli to previous ones (Ashby et al., 2011).

The COVIS Framework

COVIS provides a valuable framework for investigating the influence of non-stationarity in human category learning. The explicit and implicit learning systems (akin to the conscious/unconscious explicit/implicit memory systems) parallel the active and passive approaches in machine learning, respectively. Explicit learning involves more active hypothesis testing and searching for a verbalizable rule. Implicit learning, based in procedural learning, is believed to involve learning to associate responses with regions of perceptual/psychological space (i.e., stimulus-response association), thus corresponding to a continually adapting "passive" approach in machine learning. Explicit and implicit category structures result in qualitatively different performance in many cases, and along with changes in decisional variables (e.g., base-rates and payoffs), provide a platform by which to explore the effects of different types of concept drift on human perceptual category learning.

Review of Related Research

Base-rates and Payoffs

Decisional components of category learning have an extensive research history in stationary conditions (Maddox & Bohil, 1998, 2000, 2001; Wismer & Bohil, 2017). Two decisional components of categorization that play a vital role in maximizing reward and/or accuracy of classification are base-rates and payoffs. Base-rates are the relative frequencies of each category and can be described as a ratio. For example, a two-category situation in which category A is 3 times as likely as category B would have a base-rate ratio of 3:1. Payoffs describe the costs and benefits of correct and incorrect classification. Benefits include correctly classifying Category A or Category B. Costs involve the negative outcomes of incorrectly saying Category A when Category B was presented, or saying Category B when Category A was presented. Payoffs are usually manipulated via points summarized in a payoff matrix. For example, the payoff matrix [3 -1; -1 3] means that correct classifications for either category result in 3 points, while incorrect classifications result in the subtraction of a point. Both base-rate and payoff changes shift the optimal decision bound between categories (and may be considered a source of virtual concept drift).

Observer performance in categorization studies is often compared to that of a hypothetical “optimal classifier” who, through perfect knowledge of the category distributions and parameters, is able to maximize long-run reward (Maddox & Bohil, 2003). The optimal classifier follows the “optimal decision function” comparing the likelihoods of two category distributions via the formula: $L_o x = F(x|B)/F(x|A)$, where $F(x|i)$ denotes the likelihood of a particular stimulus given category i. When the resulting ratio is greater than 1, the stimulus is most likely to come from category B. When the resulting ratio is less than 1, the stimulus is most likely to come from category A. However, the optimal classifier not only has perfect knowledge of category distributions but also of category base-rates and the payoffs (i.e., benefits) associated with the categories (Because the present work does not include costs, these values are excluded from the present discussion). This decisional information forms the optimal decision criterion, represented as: $\beta_o = [P(A)/P(B)] \times (V_{aA}/V_{bB})$, where category probabilities (i.e., base-rates) are represented by P(A) and P(B), and V_{aA} and V_{bB} represent payoff values associated with correct responses. Finally, the output of these two formulas form the optimal decision rule, whereby the optimal classifier responds with category B when the $l_o(x)$ is greater than β_o, otherwise it responds category A. The optimal decision rule highlights two important points related to base-rates and payoffs: 1) When $P(A)V_{aA} = P(B)V_{bB}$, the optimal decision is based solely on the category likelihoods, and 2) when base-rates and payoffs are manipulated independently (i.e., only one of the two is manipulated in a block of trials) equivalent ratios result in equivalent optimal decision criterions.

The bulk of research related to base-rates and payoffs in category learning involves stationary conditions, or at the most, category training with equal base-rates/payoffs followed by an explicit change to evaluate base-rate or payoff sensitivity in terms of criterion placement (e.g., see Bohil & Wismer, 2015). The literature reveals that criterion placement is closer to optimal for base-rates than for payoffs (Bohil & Maddox, 2001). An explanation for this is provided by the Competition between Reward and Accuracy (COBRA) hypothesis (Maddox & Bohil, 2004). This hypothesis states that observers strive to balance reward and accuracy in category learning. When only base-rates change, a criterion shift can maximize reward and accuracy at the same time. However, when payoffs are asymmetric, an observer must learn to sacrifice some amount of accuracy to maximize reward. This is typically difficult to do because of the competing goals of accuracy and reward maximization. Because of this, observers typically adjust their criterion closer to the optimal decision bound in base-rate conditions as opposed to payoff conditions. In general, criterion shifting is possible; in the case of both asymmetric base-rates and payoffs, observers typically shift the criterion in the right direction but they do not shift as far as the optimal decision bound (see Figure 9 in Chapter 3). This phenomenon has been termed conservative cutoff placement (Green & Swets, 1966; Maddox, 1995).

Purposeful evaluation of adaptation to changing base-rates or payoffs is scarce in the literature. Gifford, Cohen, and Stocker (2014) provided an evaluation of auditory categorization with varying category base-rates. The category base-rate ratio changed between blocks, randomly, from 1:1, 1:3, or 3:1 to one of the other base-rate ratios. Thus, this change occurred abruptly, and participants were left to discover the base-rate differences on their own (i.e., no explicit instructions given to alert individual to change). The authors found that participants learned the base-rates relatively quickly and used them to improve categorization. However, they analyzed performance from a different (Bayesian) perspective, making comparisons to a decision bound perspective difficult. Nonetheless, their results suggest participants might have been using a strategy that assumed non-stationary category base-rates and that participants based their base-rate estimates on only a recent number of trials. Still, Gifford et al. (2014) did not vary base-rates to look at base-rate change in particular but simply to assess general auditory categorization ability. This pattern holds true in other category learning studies that varied base-rates only abruptly between blocks of trials and the manipulation was not intended to evaluate the issue of virtual concept drift (e.g., Maddox & Bohil, 1998; Wismer & Bohil, 2017).

Much of the remaining related literature on base-rate and payoff changes comes from the decision making, recognition memory, and vigilance literatures. For example, Kantner, Vettel, and Miller (2015) tested recognition memory criterion shifts. Setting old-new discrimination at near chance level (i.e., similar to a low category discriminability), they tested criterion change to changes in base-rates (Experiment 1) and payoffs (Experiment 2). Old-new discrimination was set at a low level because it is only when memory evidence is ambiguous that criterion shifts are most helpful and necessary. Critically, the payoff experiment involved actual monetary incentives for performance to see if this would increase criterion shifting to a more optimal level. Kantner et al. (2015) found that in both experiments participants strongly relied on memory evidence rather than decision rules to match biased base-rates or payoffs, despite memory evidence being highly ambiguous. This over-reliance on memory evidence resulted in high critical error rates (i.e., misses and false alarms) despite instruction and incentives to avoid such errors.

Finally, work in vigilance (sustained attention in detection tasks over an extended time period; Warm, 1984) provides some evidence of adaptation to category base-rates via manipulations of signal probability in vigilance tasks. Parsons (2001) evaluated shifts in response bias to increasing or decreasing signal probability of a 40-minute vigilance period. While response bias tends to become more conservative over time in consistent-probability conditions, Parsons (2001) found that increasing-probability conditions led to greater conservatisms shifts while decreasing-probability conditions were associated with stable response bias. Overall, results from Parsons (2001) suggest that observers' performance did not match that of an "ideal observer" (with perfect knowledge of task parameters; Vickers & Leary, 1983) but rather a "self-doubt" hypothesis where observers interpreted signal probability changes as due to inadequacies in their own observations and over-adjusted their criterion accordingly. Critically, criterion shifts to changing signal probability were independent of whether or not forewarnings of change were provided, while forewarning increased signal discrimination over time.

Findings by Nassar et al. (2010) suggest that individual differences in criterion shifting in dynamic environments can be, at least partially, attributed to differences in underlying beliefs about the stationarity of an environment (i.e., a Bayesian prior for expected frequency of environment changes). Observers completed a numerical prediction task where they were told the numbers came from a noisy process and could change over time and were told to minimize their prediction error. A simple computational model including a parameter for belief of environmental stationarity helped explain differences in criterion shifting. Thus, beliefs about the state of the environment – manipulated by the presence or absence of explicit instruction on change – may impact criterion shifting in characteristic ways, depending on how forewarning affects an individual's belief about environmental stationarity. This finding differs from that of Parsons (2001), suggesting that the effect of forewarning may depend on task or individual characteristics.

Transfer in Category Learning

While little-to-no research has focused specifically on non-stationary category learning, researchers have studied how well individuals are able to transfer or generalize category knowledge to a new set of (related) categories. Sometimes transfer learning is simply evaluated by testing individuals with never-before-seen exemplars selected from the same distribution of average features but without response feedback given. Other times transfer learning is evaluated by maintaining feedback but shifting stimulus distributions at testing. A common type of transfer learning, known as analogical transfer, relates to the notion of virtual concept drift and is the focus of the present work related to categorical change.

Analogical Transfer

Casale, Roeder, and Ashby (2012) define analogical transfer as “the ability to transfer knowledge despite significant changes in the surface features of the problem” (p. 434). In category learning, this means being able to transfer a learned rule for one stimulus set to a new stimulus set. Critically, Casale et al. (2012) were interested in people’s ability to do this for verbalizable rules compared to non-verbalizable, or difficult to verbalize, rules. They studied this by using rule-based (RB) and information integration (II) category structures with an analogical transfer task. Rule-based category structures typically involve the learning of a one-dimensional rule that differentiates the categories, and this rule is usually easy to verbalize. Information integration category structures involve the integration of information from two (or more) dimensions at a pre-decisional stage and are typically difficult to verbalize. Both types of category rules are possible to learn, although II rules typically take longer. Because of this, researchers sometimes use more complex conjunctive RB rules or four-category conditions as opposed to two-categories in an attempt to equate relative accuracy levels across rule types. However, the RB rule in Casale et al. (2012) was a simple one-dimensional rule which has been used in a great deal of COVIS research. Participants first learned either the RB or corresponding II categories and then transferred to a new set of stimuli within the same category structure but in a new area of the stimulus space defined by the same linear decision bound (see Figure 2). In so doing, the RB and II categories had the same perceptual similarity in both the training and transfer conditions.

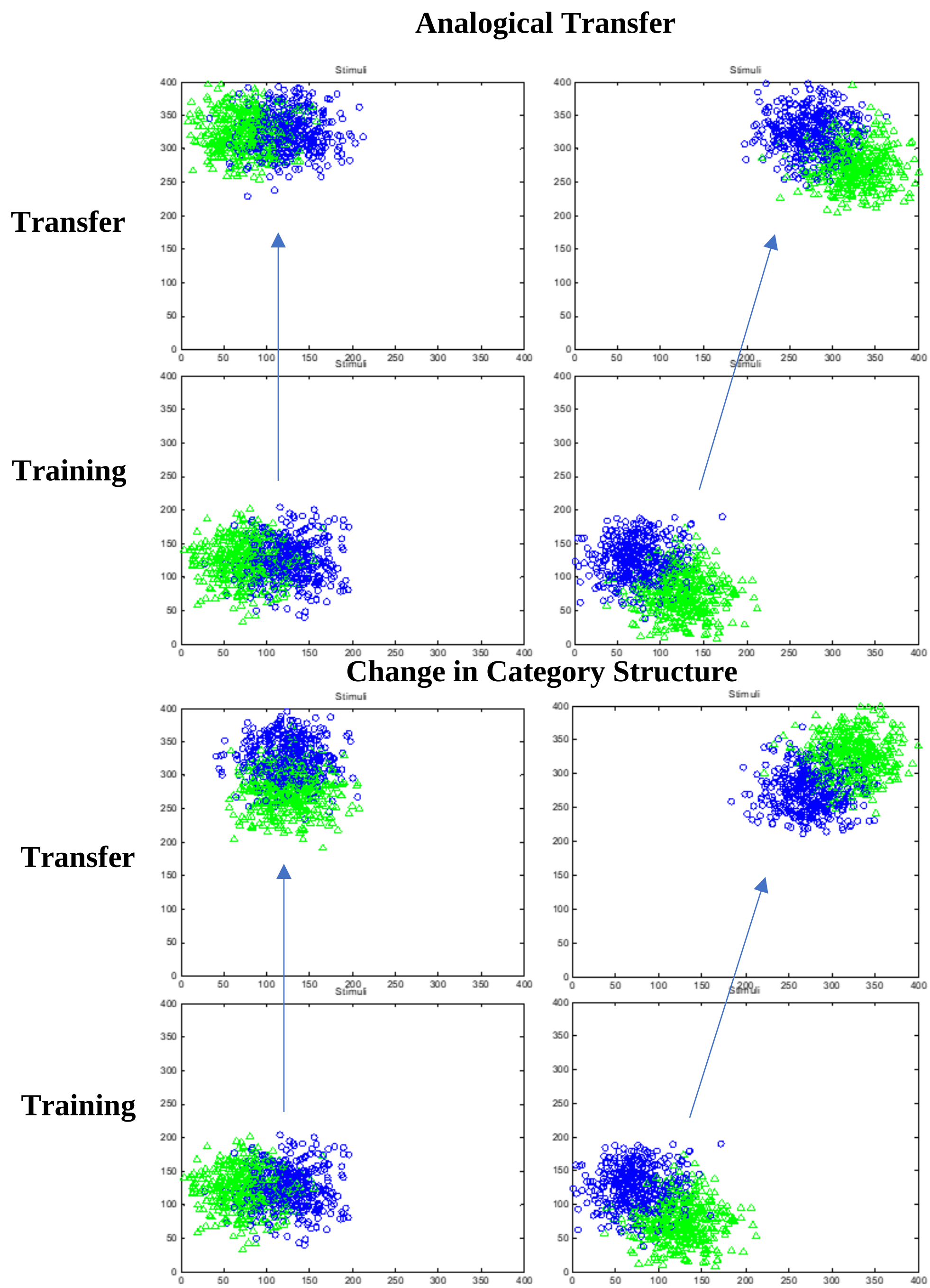

Figure 2. Example category distributions used in Casale et al. (2012, p. 436). X and Y dimensions represent perception of bar width and bar tilt, respectively. Rule-based (RB) categories are displayed in the left column; information-integration (II) categories are displayed in the right column.

This was the first study to look at analogical transfer in II categories. The authors noted that while the analogical transfer literature predicts equal transfer in RB and II (given that RB and [rotated] II categories have the same degree of perceptual similarity despite being sampled from a new space along the same decision bound), COVIS predicts transfer in RB but not II categories. COVIS makes this prediction because the explicit system learns a decision bound that is not context dependent whereas the slower implicit system learns to associate categorization responses with reasonably small regions of perceptual space rather than learning a decision bound (Ashby & Waldron, 1999; as cited in Casale et al., 2012).

Casale et al. (2012) conducted three experiments, finding in all three evidence of transfer in RB categories but no evidence for transfer in II categories. The stimuli used were sine wave gratings that varied in angle of rotation and spatial frequency. The researchers also used control conditions in which categories transferred to a new region of perceptual space along the decision bound, however the categories were then rotated 90 degrees clockwise to form new category structures within the same RB or II design (e.g., RB changed critical dimension while II changed polarity of slope and intercept of bound). The control conditions were included to test the potential effects of "meta-learning", a phenomenon in which learning on one set of category structures helps improve performance on any subsequent set of category structures due to experience with the stimuli and task (Kendler & Kendler, 1970). Control condition results revealed a significant decrease in accuracy in the first transfer block for both II analogical transfer and control conditions, but only a decrease in accuracy for the RB control condition (not RB analogical transfer condition).

A few details regarding this study are pertinent to the present work. Casale et al. (2012) had a small sample size in each condition through the three experiments (*n*'s ranging between 13 and 16). Experiment 1 used unequal training lengths for RB (300 trials) and II (700 trials), with the understanding that II categories take longer to learn than RB categories when they use the same category *d'*. Experiment 2 took an alternative approach and used unequal category discriminability values (*d'*) for RB and II categories to equate the training lengths. Both of these experiments had 200 transfer trials. Stimuli were composed of circular sine wave gratings. Experiment 3 involved the removal of feedback at transfer to see if feedback during previous transfer cases was the underlying factor rather than a difference between two learning systems. In all cases, transfer blocks introduced an abrupt change in category distributions, and participants were made aware of the presence of a change.

Smith et al. (2015) evaluated the generalization of category knowledge in humans and macaques. Critically, the authors found parallel results to Casale et al. (2012) for human participants: evidence of RB transfer along a decision bound to new regions of perceptual space but no evidence of transfer for II categories. They concluded that II learning is constrained to the original stimulus context whereas RB category knowledge has some independence from the initial stimulus context. Smith et al. (2015) used two dimensional unframed rectangles as their stimuli. These rectangles differed in size and number of illuminated green pixels. Participants received 200 different stimuli in training, and 100 new random stimuli from generalization distributions at transfer. Similar to Casale et al. (2012), the change in category distributions occurred abruptly after training, and participants were made aware that they were being asked to generalize to new category distributions.

Many other researchers have looked at category change through the lens of generalization phases but, unlike Casale et al. (2012) and Smith et al. (2015), left participants to discover the change rather than providing an alert (e.g., Cantwell, Crossley, & Ashby, 2015; Helie et al., 2015; Navarro, Perfors, & Vong, 2013; Seger, Braunlich, Wehe, & Liu, 2015; Summerfield et al., 2011). No studies to my knowledge have systematically manipulated awareness of change within a single study.

Helie et al. (2015) investigated intra- and extra-dimensional rule changes in rule-based (RB) stimuli. Following the results of Casale et al. (2012) and Smith et al. (2015), they found that there was no shift-cost (i.e., drop in accuracy) when only the irrelevant dimension changed (i.e., an intra-dimensional shift; RB analogical transfer). In addition, Helie et al. (2015) found that criterion changes were easier to make on a critical dimension if irrelevant dimensions also changed. They also found that it was harder to learn a new criterion that involved a new rule dimension, but this was facilitated by a change in the irrelevant dimension as well. The design of the two experiments involved 20-23 participants per condition, three 100-trial training blocks, three 100-trial transfer blocks, four categories instead of the traditional two, and no alerting of the change in category distributions. Thus, results suggest RB analogical transfer occurs both with alerting (Casale et al., 2012; Smith et al., 2015) and without alerting (Helie et al., 2015).

Seger et al. (2015) explored what they called representational and decision uncertainty in the generalization of category learning. Representational uncertainty relates to uncertainty about category membership while decision uncertainty is uncertainty about a decision bound. Seger et al. (2015) separated category generalization from decision bound implementation by using a particular information-integration category learning task. Participants first learned to categorize II stimuli within a limited perceptual region and then transferred to II stimuli that differed both in distance from the training region prototype as well as distance from the training decision bound. These two types of transfer tasks revealed diverse patterns of brain activation as well as differing levels of accuracy. In particular, transfer stimuli farther from the decision bound were categorized more accurately than transfer stimuli closer to the decision bound (likely due to increases in perceptual sensitivity). Flanking stimuli – stimuli that were located the same distance from the decision bound but farther from the training prototype) were categorized as well as training stimuli (see Figure 3). Boundary transfer stimuli located closer to the decision bound were categorized less accurately. Reaction times followed this identical pattern. Critically, this experiment provides the first evidence of II analogical transfer as witnessed with the flanking stimuli, which stands in contrast to the results of Smith et al. (2015) and Casale et al. (2012), who found no evidence of analogical transfer. A few key methodological detail differences may explain the different results between Casale et al. (2012) and Seger et al. (2015). Seger et al. (2015) did use abrupt change in that training and transfer blocks were separate; however, transfer blocks were conducted in an fMRI scanning session, with flanking transfer distributions, with feedback present on only half of the trials. More research is needed to understand why these methodological details led to a different pattern of II analogical transfer.

As a whole, the separate explicit and implicit learning systems posited by COVIS suggest that specific characteristics of drift may influence adaptation by these learning systems. Namely, previous research suggests that analogical transfer is relatively easy in cases of explicit learning but much more challenging in cases of implicit learning. The question remains whether drift characteristics such as rate and awareness of change influence the pattern of analogical transfer in explicit and implicit learning conditions in a way akin to the choice between active and passive learning approaches in machine learning.

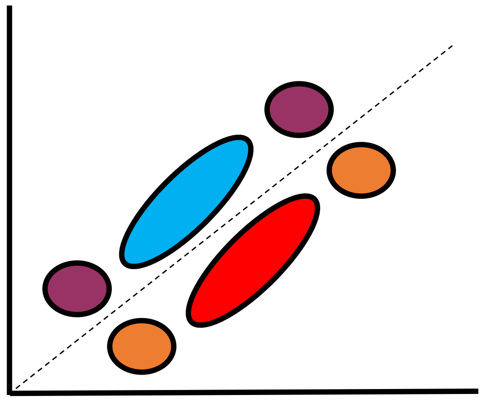

Figure 3. Example category structures used by Seger et al. (2015). Blue vs. red oval categories illustrate the starting category distributions. Purple and orange categories represent the flanking distributions (analogical transfer conditions with same means as original distributions but different areas of stimulus space).

<u>Other Changes</u>

As a side note, a few other types of category changes have been looked at in human category learning, albeit not too in-depth. The primary type of change that has received attention in both machine and human learning is periodicity effects (i.e., multiple or recurring changes). Summerfield et al. (2011) used two-dimensional rule-based stimuli defined along one dimension and introduced random mean or orientation jumps every 10-20 trials. They found that participants' responses were best fit by a working memory model with short term memory that continued to reset category means on the basis of the most recent category member encountered. This working memory model provided a significant improvement in fit compared to a Bayesian model and a reinforcement learning model.

Navarro et al. (2013) studied time varying one-dimensional categories. The stimuli were "floaters" – little objects that "floated" some height above a horizontal line. That height was the basis of category membership, and the classification boundary consistently rose on each trial.

The results reveal that participants became sensitive to the systematic pattern of change in the categories as the boundary changed over time.

Finally, some researchers have looked at the effect of category change as a rotation of the category distributions. An example of category rotation is the transformation of RB category distributions (such as in Figure 2) by a 45 degree clockwise rotation to form corresponding II category distributions (the technique used in Experiment 1 of Casale et al., 2012). For both RB and II category structures, this type of change resulted in a large decrease in accuracy (Cantwell et al., 2015) and lack of full accuracy recovery (Helie et al., 2015). The present work provides a systematic foundation on which to extend our understanding of concept drift in human category learning from cases of analogical transfer to more complex categorical changes.

Decision Criterion Learning

Studies of decision criterion learning and change in other domains are also applicable to the present work. Before describing some of what has been done in other domains, some background on decision criterion placement is in order. According to Signal Detection Theory (SDT), perceptual decisions have two components: discrimination and response bias. Discrimination, measured by *d'*, describes how well an observer can discriminate between two distributions. Smaller *d'* values indicate a lower level of discrimination. At the same time, SDT proposes that an observer sets a decision criterion, known as beta (β), when responding (see Figure 4). Beta is a measure of response bias indicating how biased an observer is for one response versus another (Green & Swets, 1966; Macmillan & Creelman, 2004; Wickens, 2001).

Decision criteria are based on likelihood ratios. SDT assumes that an observer will choose signal anytime the perceptual effect of the stimulus value results in a higher likelihood for signal than noise, and vice versa (i.e., when the signal curve in Figure 4 is higher than the adjacent point of the noise curve on the x axis). However, observers tend to be biased in their responses and may require more evidence in favor of the signal distribution before responding with that option. In this case, the observer has what is referred to as a biased criterion ($\beta \neq 1$) in favor of the noise distribution (Wickens, 2001).

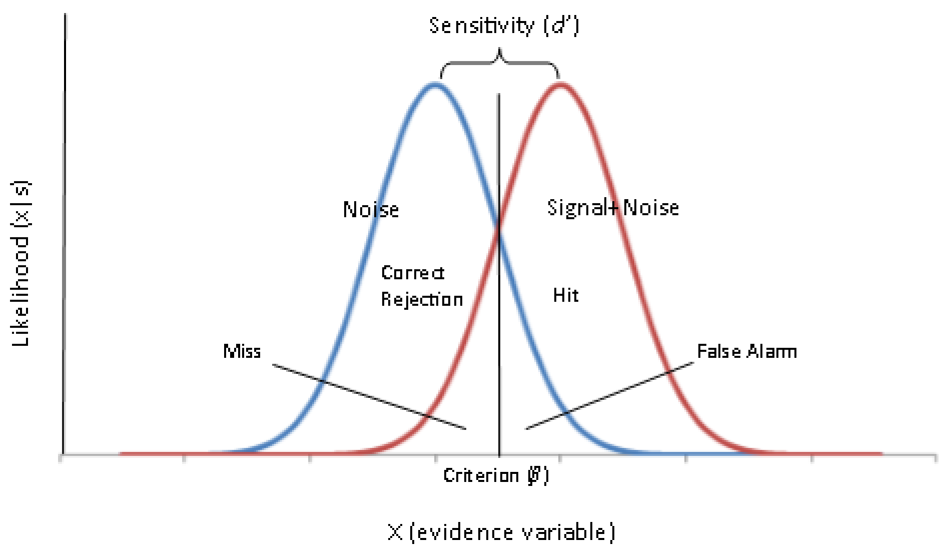

Figure 4. Signal detection curves.

The signal detection model provides a one-dimensional representation of the perceptual effect of a stimulus. General Recognition Theory (GRT), also called Decision Bound Theory, is a generalization of signal detection theory that applies to multidimensional stimuli (when muiltivariate normal perceptual distributions are assumed; Ashby & Townsend, 1986) that allows for specific representation of individual dimensions. GRT describes the separate perceptual and decisional processes involved in an action such as classification. It claims that presentation of the same stimulus does not always result in the same perceptual effect of the stimulus. GRT assumes that an observer partitions the perceptual space into separate regions, and a unique response is ascribed to each region. The decision rule separates all perceptual points into response regions. GRT also estimates both perceptual noise and decisional noise, related to imperfection in perceiving stimuli and setting a criterion from trial to trial (Alfonso-Reese, 2006). Notably, GRT provides a formal basis for COVIS (i.e., COVIS is the neuropsychological extension of GRT).

Other sources of noise include "expected" and "unexpected" uncertainty, terms coined by Yu and Dayan (2005). Expected uncertainty is uncertainty that is known to be present in predictions (or in a stochastic prediction environment). For example, we know that weather forecasts are imperfect and may have some level of reliability (say, 80%). It is therefore common for weather to sometimes be different than the forecast for that day. This type of difference is expected from time to time. On the other hand, unexpected uncertainty arises when sensory observations deviate greatly from top-down expectations. Thus, if equipment malfunction affects weather forecasting ability, weather forecasts may become no more reliable than flipping a coin or looking out the window (Hassall, 2013). When unexpected uncertainty arises in non-stationary environments, previously optimal actions become suboptimal.

These two forms of uncertainty are not mutually exclusive but rather they interact with each other in many environments. In environments with high expected uncertainty, unexpected uncertainty can be hard to detect and adapt to (Behrens, Woolrich, Walton, & Rushworth, 2007; Bland & Schaefer, 2011). In terms of categorization then, expected uncertainty arises from the probabilistic nature of the categories; in other words, the sampled category distributions are the largest source of variability. When category distributions significantly overlap, it will be expected that stimuli falling in the area under the overlapping curves will sometimes be classified as category A and other times as category B (see Figure 4). In these cases, when a categorical distribution shift occurs due to a mean and/or variance shift, such unexpected change can be difficult to detect due to the already high level of uncertainty in the environment (Hassall, 2013).

This is the true problem posed by a non-stationary environment – how does one differentiate between noise (expected uncertainty) and actual change (unexpected uncertainty)? The present work will incorporate this aspect of expected vs unexpected uncertainty in the design of the categorization experiments. In each experiment, the categories are defined probabilistically (i.e., they are drawn randomly from overlapping, normal distributions). This means that there is considerable overlap between categories, and one stimulus value(s) may indicate category A on one trial but category B on another trial. This variability in category members (drawn from overlapping distributions) can be considered expected uncertainty. The change manipulations that occur in each experiment will provide a test of unexpected uncertainty (at least in the "discover" conditions where participants are not alerted to any change).

Criterion learning will serve as the basis of measurement to evaluate the influence of concept drift in human perceptual category learning. By incorporating distinct, characteristic elements of concept drift in human category learning, one can explore how humans learn in non-stationary environments. The combination of machine and human classification learning techniques will inform current categorization theory.

The Present Research

In summary, most category learning research has only incidentally studied non-stationarity as a side component of some larger research question (e.g., generalization, analogical reasoning, impact of uncertainty, refining models). In so doing, limited category changes have been investigated, base-rate and payoff changes have not been systematically and intentionally studied, and largely only abrupt change rates (between blocks) have been considered (see Table 1 below for an overview of some of the key non-stationarity variables that have been looked at in human category learning). To my knowledge, no one has manipulated rate of change or awareness of change within a single study.

Table 1. Summary of key variables in previous research

Citation	Rate		Awareness		Category Structure	
	Abrupt	Gradual	Alert	Discover	RB	II
Cantwell et al. (2015)[*, **]	✓			✓		✓
Casale et al. (2012)[*]	✓		✓		✓	✓
Helie et al. (2015)[*, **]	✓			✓	✓	
Navarro et al. (2013)[**]		✓		✓	✓	
Seger et al. (2015)[*, **]	✓		✓			✓
Smith et al. (2015)[*, **]	✓		✓		✓	✓
Summerfield et al. (2015)[**]	✓			✓	✓	

Note. Citations marked with one asterisk (*) denote experiments that tested analogical transfer. Citations marked with two asterisks (**) denote experiments that tested other types of category change (i.e., shifted category boundary).

Systematic investigation of these rate, awareness, and type of change variables should provide several useful outcomes. 1) It should provide further evidence for the distinction of multiple learning systems by testing differential predictions of explicit and implicit-mediated learning tasks. 2) It should provide insights for the refinement of theories for the presence of change based on different change characteristics. 3) It may provide insight back to the machine learning community regarding the effects of systematic concept drift changes on types of learners (human classifiers). In general, the aims of this dissertation are to answer the following questions. Do humans need to be aware of change to efficiently adapt to it? If so, under what circumstances is it required or most helpful? Do different learning systems respond better or worse to different rates of change? Is the impact of “virtual concept drift” in human category learning dependent on the parameter of change (i.e., category base-rates, payoffs, or features)?

Non-stationary category learning has received quite little attention in the literature. In particular, concept drift has not been the target of study in perceptual category learning; rather, it has been more of a by-product of other research questions. For example, Casale et al. (2012) were interested in analogical transfer in perceptual categorization and whether COVIS provided a platform for disambiguating explicit and implicit category learning in terms of transfer. Helie et al. (2015) were interested in testing a new model of rule selection, while Cantwell et al. (2015) had similar interests in proposing a neurobiological model of procedural category learning. For this reason, virtually every categorization study has used abrupt change rates (i.e., between blocks) rather than gradual change rates to test transfer performance. Awareness of change has not been controlled for, and changes in base-rates and payoffs have received little attention.

Overview of Experiments

This dissertation presents the results of four experiments evaluating effects of concept drift in human perceptual category learning. Each experiment consisted of 8 blocks of 50 trials for a total of 400 trials. Each experiment included a single, unique change parameter that shifted mid-experiment (either abruptly between blocks 3 and 4, or gradually over blocks 4 and 5, dependent on the rate of change). Informed by the concept drift literature in machine learning, each of the four experiments included the factorial combination of rate of change (abrupt, gradual) and awareness of change (alert, discover) as independent variables, along with a control condition in which the parameter of interest was held constant. These details are summarized below:

- **Component of virtual concept drift**
    - Base-rates (Experiment 1)
    - Payoffs (Experiment 2)
    - Rule-based (RB) analogical transfer (Experiment 3)
    - Information-integration (II) analogical transfer (Experiment 4)
- **Rate of change (Abrupt vs. Gradual)**
    - Abrupt change occurs by changing the value of the parameter of change immediately, and in full, after the third block of trials.
    - Gradual change occurs by slowly shifting the value of the parameter of change in equal increments over blocks four and five.
- **Awareness of change (Alerted vs. Discovered)**
    - Alerted = observers are forewarned of the upcoming change, and parameter of change, in the instructions before the first change/post-change block. Observers are also reminded of the change thereafter between each block of trials. Observers are told only that a specific aspect of the task (e.g., category base-rates) will change, not the specific direction or magnitude of the change.
    - Discovered = observers are told nothing of the upcoming or past change and are left to discover the change on their own.

Each of the four factorial conditions involved three stationary pre-change blocks (150 trials) and at least three stationary post-change blocks (150+ trials; blocks 6-8). Abrupt change conditions ultimately had two additional post-change blocks (blocks 4 and 5) since the change occurred in full following block 3. Gradual change occurred over blocks 4 and 5 by incrementing the parameter of change by the magnitude of the change divided by the duration of the change

(100 trials) on each change trial. Once the full magnitude of the change occurred, the change parameter remained fixed for the remainder of the experiment (see Figure 5 and Figure 6 for the design of the abrupt and gradual change rate conditions, respectively.

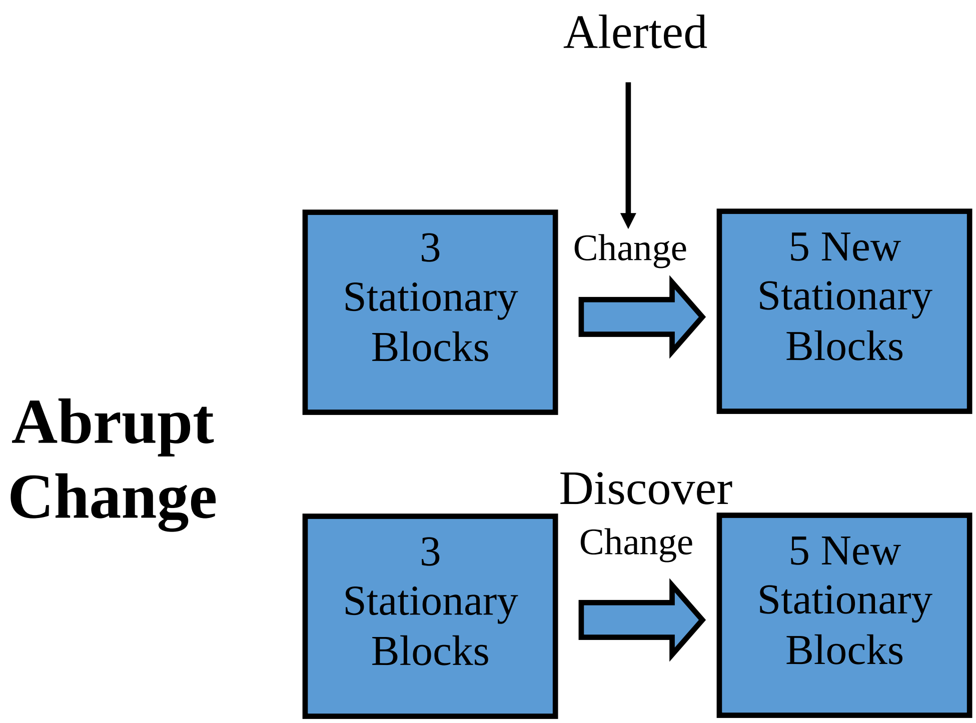

Figure 5. Experimental design for abrupt change conditions.

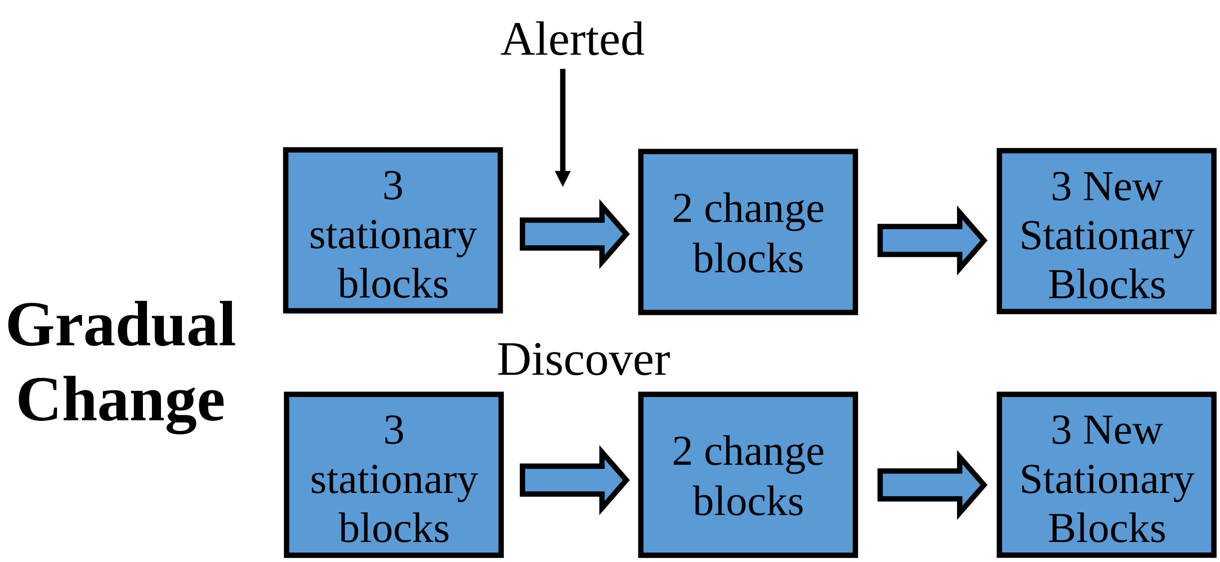

Figure 6. Experimental design for gradual change conditions.

Experimental Evaluation

COVIS provides a framework for evaluating the effect of concept drift on explicit vs. implicit learning. The application of machine learning change characteristics to perceptual category learning helps extend categorization theory from stationary contexts into new areas of non-stationarity, paralleling the dynamic and complex nature of category learning in the natural environment.

Four experiments were conducted to evaluate the effect of rate and awareness of change in different “virtual” change conditions (i.e., tasks that required a shift in decision criteria to maximize performance, while category boundaries remained the same). Experiment 1 investigated people’s ability to shift their decision criteria to changes in category base-rates. Experiment 2 investigated people’s ability to shift their decision criteria to changes in category payoffs. Experiment 3 investigated people’s ability to transfer rule-based category knowledge (i.e., explicit learning) when surface features changed (i.e., analogical transfer), while Experiment four investigated the same question within information-integration category knowledge (i.e., implicit learning). Within each experiment, the factorial combination of rate (abrupt, gradual) and awareness of change (alerted, discovered) was explored.

The experiments were designed in such a way that a direct comparison could be made between Experiment 1 and 2 (base-rate and payoff learning) and between Experiment 3 and 4 (RB and II learning). Each set of conditions provided a platform to explore the effect of virtual concept drift in tasks that tend to involve more implicit vs. explicit learning.

Decision bound models were explored to provide a qualitative comparison among conditions and statistical tests of the best fitting models. Model parameters were maximum likelihood estimators determined by a hill climbing algorithm that minimized negative log-likelihood fit measures (lnL). The most parsimonious model was determined on the basis of goodness of fit tests (χ^2 and the akaike information criterion [AIC], Ashby and Soto (1992).

CHAPTER 3: EXPERIMENT 1: BASE-RATES

Base-Rate Change

Experiment 1 explored the influence of a shift in category base-rates on categorization performance. Base-rate change is one element of concept drift in machine learning, sometimes referred to as "virtual concept drift" (Gama, Žliobaitė, Bifet, Pechenizkiy, & Bouchachia, 2014). In addition, a change from unbiased (equal) to biased (unequal) category base-rates reflects the issue of "class imbalance" in machine learning (Hoens et al., 2012). This type of change/imbalance has been studied from different perspectives in machine learning (Alaiz-Rodríguez, Guerrero-Curieses, & Cid-Sueiro, 2007; Drummond & Holte, 2006; Provost & Fawcett, 2001; Saerens, Latinne, & Decaestecker, 2002).

In categorization, base-rates provide an important source of information for criterion placement (i.e., response bias from signal detection) when categories are difficult to discriminate on features alone. A change in category base-rates requires the observer to shift their decision criterion in accordance with the changing likelihood of each category to maintain accuracy. Such a change is required in medical personnel when disease frequencies and patient populations change over time or when new technology increases discriminability. Experiment 1 provides a systematic exploration of the influence of rate and awareness of change on perceptual categorization with a shift in category base-rates.

Method

Experimental Design

Four experimental conditions resulted from the factorial combination of the between-subjects independent variables rate of change (abrupt, gradual) and awareness of change (alerted, discovered):

1. Abrupt base-rate change, alerted in instructions
2. Abrupt base-rate change, no alerting (i.e., discovered)
3. Gradual base-rate change, alerted in instructions
4. Gradual base-rate change, no alerting (i.e., discovered)

A fifth between-subjects condition was run with stationary category base-rates, serving as a control condition.

The task was described as a simulated medical diagnosis task. Participants saw one bar graph on each trial, varying in height, which simulated one person's test result. Participants were tasked with learning which disease (A or B) was more likely given the test result. Participants were told that the test was not a perfect predictor, but that they should still be able to reach a high level of performance over time.

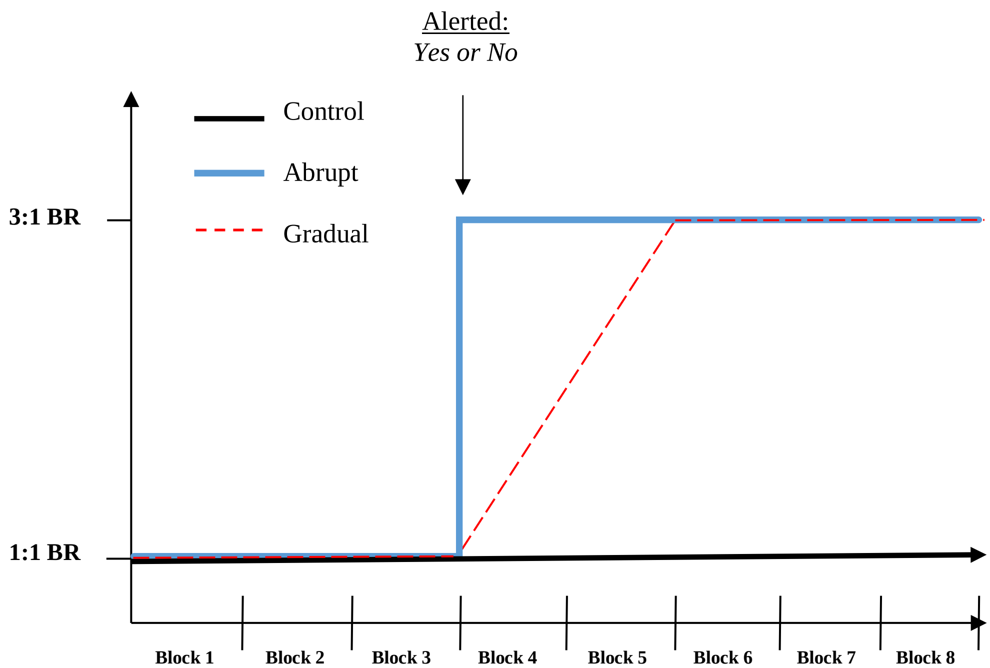

Figure 7. Experimental design for Experiment 1.

All participants started the experiment with stationary, equal category base-rates (1:1 ratio; on average, 25 exemplars from both disease A and disease B) for the first three blocks of trials – from here-on referred to as the “pre-change blocks”. Following block three, the base-rate ratio changed from 1:1 to 3:1 either abruptly (immediately between blocks 3 and 4) or gradually (in equal increments each trial across blocks four and five), resulting in, on average, 36-39 exemplars from disease A and 12-13 exemplars from disease B. Once the change was complete (i.e., the base-rate ratio reached 3:1), the new base-rate ratio remained stationary for the remaining blocks of trials. All participants completed 8 blocks of 50 trials each for 400 trials in total (see Figure 5 and Figure 6 in Chapter 2). One hundred trials of gradual change was deemed to be sufficiently gradual (i.e., a 0.25% change in base-rate on each trial is too small to notice incrementally) and allowed an equal number of blocks pre (3) and post (3) change for comparison.

The only parameter that changed during Experiment 1 was the category (disease) base-rate ratio; no other features of the categories or decision environment changed. Two of the four experimental conditions were alerted to the change on the instructions screen. In abrupt change conditions, the instructions stated “The rate of disease presentation has now changed. One category is now more frequent than the other.” In gradual change conditions, the first block of change was preceded by instructions saying “The rate of disease presentation will change. One category will become more frequent than the other.” The second change block was preceded by the instructions “The rate of disease presentation will continue to change, making one category more frequent than the other”. All post-change blocks – in alerted conditions - were preceded with instructions that stated “REMEMBER: One category is now more frequent than the other!”.

Discover conditions were not provided any of this information on instructions screens and were left to discover the change on their own.

Gradual change conditions slowly incremented the Category A base-rate parameter on each change trial of the 100 total change trials across blocks 4 and 5. This occurred by increasing Category A's base-rate by the total base-rate range (75%-50% = 25%) divided by the 100 change trials = 0.25% on each trial. Therefore, on the first change trial there was a 50.25% chance of observing a stimulus from Category A, a 50.50% chance on the second trial, a 50.75% chance on the third change trial, and so on. Once the base-rate of Category A reach its maximum value (75%) on the final change trial, the base-rates remained fixed at the new ratio levels (3:1 in favor of Category A).

Participants

Students from the University of Central Florida and volunteers from the community participated in the study. Most students signed up for and participated in the study through the SONA systems participation tool in exchange for course credit. All participants were at least 18 years of age and normal or corrected-to-normal vision. Each participant completed only one condition of the study and were assigned their condition randomly. Samples sizes were as follows: abrupt/alerted (n = 32), abrupt/discovered (n = 30), gradual/alerted (n = 28), gradual/discovered (n = 27) and control (n = 30) conditions, for a total of 147 participants. One participant in the abrupt/discover condition was removed from all analyses due to non-learning (responded with low base-rate category on the majority of trials in all post-change blocks), resulting in a sample size of 29 for this condition. One participant was removed from the gradual/alert condition due to being an overall outlier (7 out of 8 blocks their decision criterion was more than 3 *SD*'s above the mean). This resulted in 145 participants overall. These sample sizes reflect traditional conventions in the literature (Helie et al., 2015; Wismer & Bohil, 2017).

A power analysis was also conducted using G Power (ANOVA: Fixed effects, special, main effects and interactions; Effect size f = 0.3067, alpha = .05, power = .85, numerator df = 1, number of groups = 2) and suggested a total sample size of 98. A target sample size of 25 participants in each of the four experimental conditions satisfies this analysis. The chosen effect size was based on data from Wismer and Bohil (2017)'s Experiment 1 comparing participants' decision criteria (β) in the last block of baseline 1:1 base-rate to the first training block of 3:1 base-rate in response conditions ($\eta_p^2 = .089$).

Participant demographics can be seen in Table 2 by condition. Of the 145 participants across the five conditions (including the control condition), 91 participants were female and 54 were male. The average overall age was 18.84 (*SD* = 1.60). Average age, $\chi^2(24) = 23.40$, $p = .496$, and gender, $\chi^2(4) = 3.19$, $p = .527$, were equivalent across conditions.

Table 2. Participant demographics by condition in Experiment 1.

	Age		Gender Count	
Condition	*M*	*SD*	Male	Female
Abrupt, Alert (*n* = 32)	18.75	1.72	14	18
Abrupt, Discover (*n* = 29)	18.93	1.39	7	22
Gradual, Alert (*n* = 27)	18.81	1.82	10	17
Gradual, Discover	19.00	1.27	10	17
Control (*n* = 30)	18.73	1.80	13	17

Stimuli

Experiment 1 used one dimensional stimuli of bar graphs with categories defined probabilistically by the height of the bar graph (see Figure 8). Two categories were sampled from overlapping normal distributions with different mean heights and identical standard deviations (see Figure 9). The category separation was small (*d'* = 1; discriminability from signal detection) such that there was significant overlap in the two categories resulting in an optimal accuracy level of approximately 69% in an equal base-rate condition and 78% in a 3:1 base-rate condition. As mentioned previously, it is only in cases of significant category overlap when criterion shifting is required to continue responding accurately. Categories with a *d'* of 1 have been used in previous studies of unidimensional criterion learning (Bohil & Wismer, 2015; Maddox & Bohil, 2001; Wismer & Bohil, 2017).

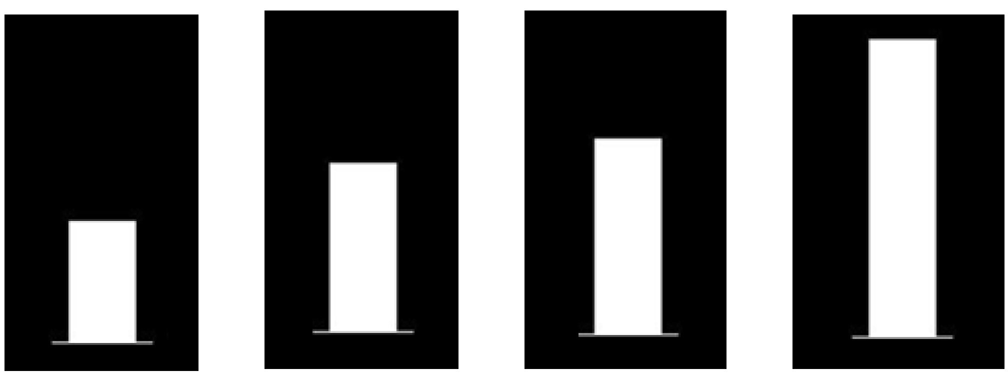

Figure 8. Example bar height stimuli used in Experiment 1 (and Experiment 2). One bar stimulus presented on screen per trial.

On each trial, one random stimulus from either Category A or B was presented on the center of the screen. The category was determined by a random selection based on the base-rate parameter. For example, since the experiment began with a 1:1 base-rate, there was a 50% chance of observing either a Category A or a Category B stimulus on a given trial. When the base-rate ratio changed to 3:1 (i.e., 75% in favor of Category A), the same random sampling method was used, resulting in many more stimuli from the category with a 75% base-rate. Actual bar heights were randomly selected from a normal distribution with a mean value equal to the mean of the selected category distribution, and the standard deviation (which is equal for the two categories; see Table 3). Because of the probabilistic nature of stimuli and stimulus value selection, individuals encountered a slightly different ratio of A:B stimuli on any given block, and base-rate ratios also varied slightly among individuals. Nonetheless, base-rate ratios were highly consistent across participants and across equivalent ratio blocks (pre-change blocks 1-3: $M = 0.49$, $SD = 0.06$; post-change blocks 6-8: $M = 0.76$, $SD = 0.05$).

Table 3. Category statistics in Experiment 1.

Time	Category	μ	σ	Base-rate ratio
Start	A	99	21	1:1
	B	120	21	1:1
End	A	99	21	3:1
	B	120	21	1:3

Note. The ratio of category A:B stimuli started at 1:1 and ended at 3:1.

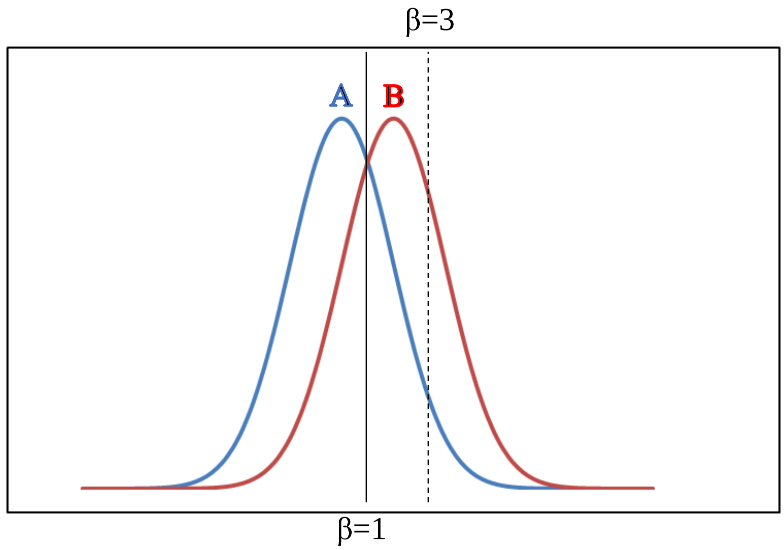

Figure 9. Distribution of categories for Experiment 1 and 2. The solid line indicates the optimal decision criterion (β = 1) in pre-change blocks, while the dotted line indicates new optimal decision criterion following the base-rate change (β = 3). When only base-rates are manipulated (i.e., no payoffs differences), the accuracy and reward maximizing criteria are equivalent (i.e., both are defined by the optimal decision criterion β = 3).

Procedure

Participants provided verbal informed consent prior to study participation. They were then directed to the research lab and seated at a computer with a high-resolution monitor. The categorization task was employed through MATLAB via the Psychophysics Toolbox extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Participants were provided instructions on how to complete the simulated medical diagnosis task using the ‘A’ and ‘B’ labeled keys (the ‘z’ and ‘m’ keys, respectively) on the keyboard on each trial. They were also instructed that on some trials they would be asked to provide a rating of confidence on a scale from 1 (very unsure) to 5 (very confidence) using these respective keys. Participants were instructed to read through the instructions on the screen before beginning the task.

On each trial, one randomly selected stimulus was presented on the center of the screen (the category was randomly selected based on the base-rate parameter, while the category value was randomly sampled from the normally distributed category; see Table 3). The stimulus was on screen for 250ms and then replaced with the words “Disease A or B” until the participant responded. If the participant did not respond within 10 seconds, they were notified that they took too long and the trial repeated. On 17 random trials per 50-trial block (i.e., 1/3 of all trials; randomized across participants), the categorization response was followed by a prompt asking for a confidence rating prior to receiving correct or incorrect feedback on their categorization response[1]. The participant again had 10 seconds to respond with a confidence rating, at which point categorization feedback was displayed for 1000ms. The inter-trial interval was 2000ms.

[1] Confidence ratings were chosen to be collected only on 1/3 of trials rather than on every trial for the following reasons: 1) confidence is secondary to criterion shifting and accuracy, 2) confidence is analyzed by block averages rather than trial-by-trial responses, 3) to reduce the risk of responses becoming too “rote” or repetitive, and 4) to not greatly increase the length of the experiment.

Participants completed eight 50-trial blocks. Participants could take short breaks between blocks as needed when an instruction screen was presented as a reminder of the task. The category base-rates changed upon completion of the third block of trials, either abruptly or gradually depending on the experimental condition. At the end of the experiment, the participant was thanked, awarded credit for participation, and dismissed.

Predictions

Base-rates are known to be poorly understood and applied when presented in descriptive, summary formats (i.e., “75% of stimuli are from category A”; Kahneman & Tversky, 1973). However, recent research has revealed that base-rates are better understood when learned through direct experience (Koehler, 1996; Spellman, 1996). Indeed, new evidence suggests base-rate sensitivity is driven largely by implicit learning through stimulus-response association (Bohil & Wismer, 2015; Wismer & Bohil, 2017).

For these reasons, it was anticipated that base-rate change would be difficult to adapt to in the sense that observers would largely differ from an ideal observer who has perfect knowledge of base-rates, base-rate change, and who can apply a consistent decision criterion. Considering the implicit nature of base-rate sensitivity and results of Parsons (2001), it was predicted that awareness of changing base-rates would not significantly impact performance. In addition, given the implicit nature of base-rate sensitivity, categorization may follow a gradual Bayesian updating process in which gradually changing base-rates may be better adapted to than abruptly changing base-rates. No prior work has investigated the effect of gradually changing category base-rates on human perceptual category learning.

Hypothesis 1: Based on previous research, participants will shift their decision criterion in the optimal direction, but they will show conservatism (i.e., sub-optimal criterion shifting)

Hypothesis 2: Given evidence of the implicit nature of base-rate sensitivity, performance (accuracy, criterion-shifting) in discovery conditions will be no worse than in alerted (forewarned) conditions.

Hypothesis 3: Given evidence of the implicit nature of base-rate sensitivity, there will be greater performance (i.e., accuracy and criterion values closer to optimal following the change) in gradual change rate conditions compared to abrupt change rate conditions.

Results

Results were analyzed using Matlab, Excel, and JASP (JASP Team, 2018). Primary dependent measures included signal detection criterion values (SDT-β), decision bound theory/general recognition theory criterion values (GRT-β), and accuracy (all directly linked to specified hypotheses). Secondary measures that were of interest but were not linked any a priori hypotheses included average confidence and the frequency of responding with the high base-rate category (i.e., category A post-change). Signal detection analyses were conducted on responses, and the resulting hit and false alarm rates were used to estimate each participant's decision criterion (β). Perceptual sensitivity (*d'*) is not presented here as there are no theoretically motivated predictions regarding *d'*, and historically *d'* has not been influenced in base-rate and payoff studies (indeed, *d'* was found to not differ by condition [p = .223] or condition by block [p = .704] in Experiment 1). Cases of hit or false alarm rates equaling 0 or 1 (preventing calculation of interpretable β values) were corrected following the method of Wixted and Lee (n.d.). For N = 25 items from Category A, values of 0 were replaced with $1/2N$ = .02 and values of 1 were replaced by $1 – (1/2N)$ = 0.98. Any data points that were classified as outliers (more than three standard deviations above or below the group mean) were replaced with the mean of the group when outlier values were removed. The number of outlier data points replaced in each condition can be seen in Table 4 below.

Table 4. Number of outliers in Experiment 1 by condition and measure.

	SDT-β	Accuracy	Number data points per measure (# blocks x # participants)
Abrupt, alert	10	1	256
Abrupt, discover	2	0	232
Gradual, alert	3	0	216
Gradual, discover	4	0	216
control	4	0	240

Note. No outliers were identified for average confidence or proportion of Category A responses.

The general outline of analyses is as follows. First, a series of results are reported for accuracy, β-values (SDT and GRT), confidence, and frequency of responding with high base-rate category comparing experimental (change) conditions to the control condition. This set of analyses serve as a manipulation check to test for pre-change differences and then assess the impact of the base-rate manipulation relative to the control condition. Then, a series of analyses are reported for each dependent measure regarding the effect of base-rate change across the four experimental conditions. These analyses include the immediate and long-term effects associated with the base-rate change, as well as base-rate sensitivity across post-change blocks. Whenever the assumption of sphericity (in ANOVA) was violated, Greenhouse-Geisser corrections were applied. All post-hoc comparisons were completed using Bonferonni's method (these points are true of all analyses reported in this dissertation).

Comparison to Control

Signal Detection Decision Criterion (β)

Average decision criterion (β) values can be seen by block and condition in Figure 10 and Table 5. To ensure there were no differences in average criterion placement prior to base-rate change, a 3 (pre-change blocks 1-3) x 5 (conditions; including control) mixed factor ANOVA was conducted on average SDT β-values. Indeed, there were no differences among conditions with respect to average β-values, $F(4, 140) = 0.55$, $p = .700$, $\eta_p^2 = .02$. A similar ANOVA was then conducted on change/post-change blocks (4-8) as a manipulation check to assess response bias to base-rate change in the four experimental conditions compared to the control. Average β-values increased across change/post-change blocks, $F(1.78, 249.22) = 9.09$, $p < .001$, $\eta_p^2 = .06$, and differed by condition, $F(4, 140) = 5.58$, $p < .001$, $\eta_p^2 = .14$. Post-hoc comparisons revealed significantly larger β values in abrupt/alert ($M = 1.52$) and abrupt/discover ($M = 1.52$) conditions compared to both the control condition ($M = 1.02$; p = 004, p = .006) and the gradual/discover condition ($M = 1.10$; $p = .036$, $p = .045$). These results suggest that the base-rate manipulation did influence response bias in the experimental conditions. Further differences between rate and awareness of change conditions in post-change blocks are evaluated in a later section.

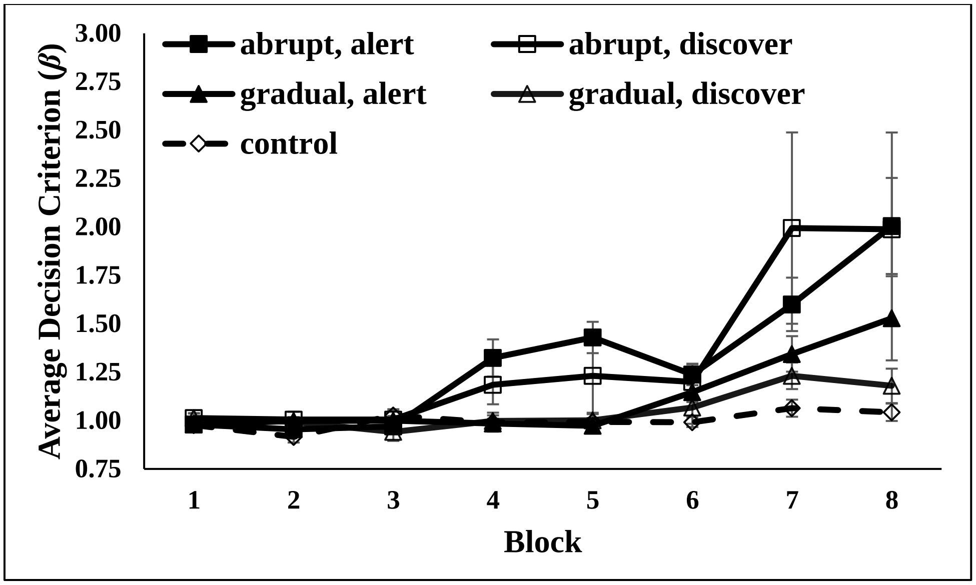

Figure 10. Average SDR-β by condition and block in Experiment 1. Error bars represent standard error.

Table 5. Average SDT-β (and *d'*) by condition and block in Experiment 1.

	Block 1 β (*d'*)	Block 2 β (*d'*)	Block 3 β (*d'*)	Block 4 β (*d'*)	Block 5 β (*d'*)	Block 6 β (*d'*)	Block 7 β (*d'*)	Block 8 β (*d'*)
Abrupt, Alert (*n* = 32)	0.98 (0.58)	0.95 (0.53)	0.97 (0.53)	1.32 (0.80)	1.43 (0.62)	1.24 (0.53)	1.60 (0.72)	2.01 (0.60)
Abrupt, Discover (*n* = 29)	1.01 (0.41)	1.01 (0.45)	1.01 (0.56)	1.18 (0.67)	1.23 (0.42)	1.20 (0.50)	1.99 (0.62)	1.99 (0.47)
Gradual, Alert (*n* = 27)	0.98 (0.46)	1.00 (0.51)	1.00 (0.46)	0.98 (0.59)	0.97 (0.48)	1.15 (0.55)	1.34 (0.60)	1.53 (0.51)
Gradual, Discover (*n* = 27)	1.01 (0.44)	0.99 (0.37)	0.94 (0.43)	1.00 (0.62)	1.00 (0.43)	1.07 (0.39)	1.23 (0.53)	1.18 (0.33)
Control (*n* = 30)	0.98 (0.23)	0.92 (0.73)	1.03 (0.55)	0.99 (0.56)	0.99 (0.58)	0.99 (0.45)	1.06 (0.72)	1.04 (0.51)

Decision Bound Models

Decision bound models were fit to the data to provide another estimate of the criterion each participant used, and they allow one to test hypotheses at the individual participant level. Three models were examined: 1) an unbiased model that assumed the participant used an unbiased ($\beta = 1$) criterion, 2) a free-boundary model that freely estimated the criterion the participant used, and 3) an optimal model that assumed the participant used the optimal criterion pertaining to each block ($\beta_o = 1$ before change, $\beta_o = 3$ after change completed, $\beta_o = 1.5$ during first block of gradual change, and $\beta_o = 2.5$ during second block of gradual change). The number of free parameters in each model are as follows: unbiased (1: noise), free-boundary (2: noise, criterion), and optimal (1: noise). Analyses report the proportion of participants in each condition best fit by each of the three models (see Table 6). The β's estimated in the free-boundary model were subjected to the same series of analyses as the β's from traditional signal detection theory. Data points considered outliers in the signal detection β-values were also removed as outliers in the GRT-β estimates to maintain consistency.

A 3 (pre-change blocks 1-3) x 5 (condition; including control) mixed factor ANOVA was evaluated to determine if any differences existed in GRT-β estimates prior to the base-rate change. As with signal detection β's, there was no effect of condition on average GRT-β estimates (all *p*'s > .151). A subsequent 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA revealed a main effect of block, $F(3.81, 529.82) = 11.31$, $p < .001$, $\eta_p^2 = .08$, such that average GRT-β values increased across blocks (in experimental conditions). There was also a main effect of condition, $F(4, 139) = 8.90$, $p < .001$, $\eta_p^2 = .20$, such that average GRT-β estimates were higher in all four experimental conditions compared to the control condition (all p's < .003). Results from GRT-β estimates corroborate what was found from the signal detection β-values (see Figure 11).

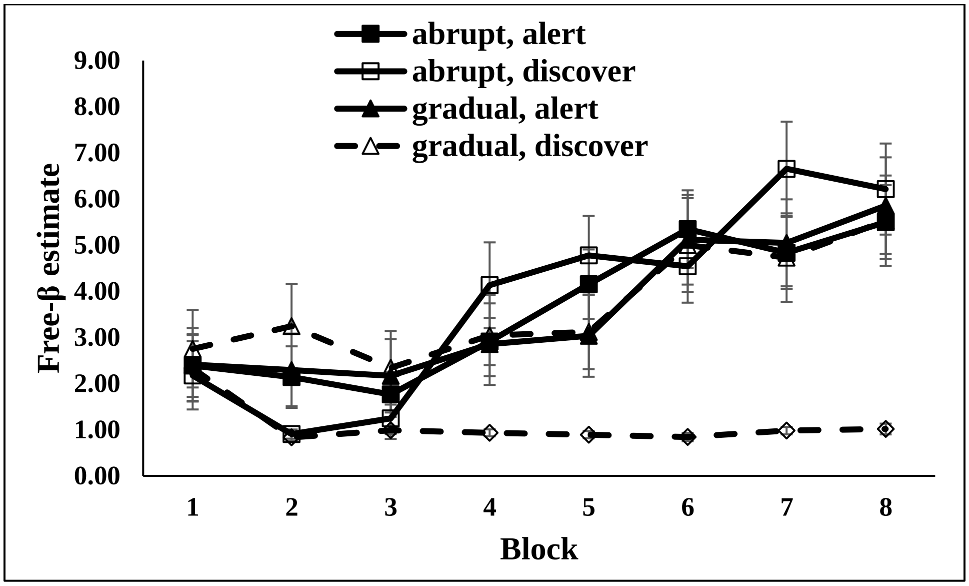

Figure 11. GRT-β estimates from free-boundary model in Experiment 1. Error bars represent standard error.

Table 6 displays the proportion of participants best fit by each model in each block. The optimal and unbiased models were equivalent for the first three blocks since the optimal criterion was unbiased ($\beta_o = 1$), as well as for the entire control condition. It is obvious that most participants began by using an unbiased (i.e., the optimal) criterion, which is expected given no information otherwise regarding base-rate bias. Once the change started to occur or occurred in full at block 4, the proportion of participants' best fit by the unbiased model decreased across blocks, suggesting sensitivity to the base-rate manipulation. Indeed, z tests (1-tailed) comparing each experimental condition to the control condition within each of the 8 blocks with respect to the proportion of participants best fit by the unbiased model found that the control condition was best fit by the unbiased model more often than the experimental conditions in every block (all p's

< .05). When looking at the individual effects of rate and awareness of change, the only significant difference found was between gradual and abrupt rates of change in block 5 – the final change block for the gradual conditions – with a higher proportion of participants best fit by the free-boundary or optimal model in abrupt conditions, $z = 2.71$ (1-tail), $p = .003$. This suggests that participants in the last block of gradually changing base-rates were more conservative in their response bias than participants in their second block of post-change base-rate values.

Table 6. Proportion of participants in each condition of Experiment 1 best fit by unbiased, free-boundary, and optimal boundary models.

Condition		Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
Individual									
Abrupt, Alert	UNB	0.78	0.90	0.74	0.29	0.23	0.17	0.23	0.16
	FRB	0.22	0.10	0.26	0.10	0.20	0.20	0.27	0.26
	OPT	0.78	0.90	0.74	0.61	0.57	0.63	0.50	0.58
Abrupt, Discover	UNB	0.79	0.86	0.86	0.41	0.28	0.30	0.21	0.14
	FRB	0.21	0.14	0.14	0.14	0.14	0.19	0.21	0.38
	OPT	0.79	0.86	0.86	0.45	0.59	0.52	0.59	0.48
Gradual, Alert	UNB	0.96	0.93	0.81	0.52	0.67	0.35	0.35	0.22
	FRB	0.04	0.07	0.19	0.11	0.11	0.27	0.23	0.26
	OPT	0.96	0.93	0.81	0.37	0.22	0.38	0.42	0.52
Gradual, Discover	UNB	0.78	0.77	0.74	0.48	0.42	0.30	0.23	0.19
	FRB	0.22	0.23	0.26	0.07	0.15	0.15	0.15	0.15
	OPT	0.78	0.77	0.74	0.41	0.42	0.56	0.62	0.65
Control	UNB	.97	.90	.83	.80	.90	.77	.80	.60
	FRB	.03	.10	.17	.20	.10	.23	.20	.40
	OPT	.97	.90	.83	.80	.90	.77	.80	.60
Average (proportion best fit by either frb or opt model)									
Abrupt					.65	**.75**	.77	.78	.85
Gradual					.50	**.45**	.68	.71	.79
					#	**	#	#	#
Alert					.60	.56	.75	.71	.81
Discover					.55	.65	.70	.78	.84
					#	#	#	#	#

Notes. UNB = unbiased boundary model; FRB = free boundary model; OPT = optimal boundary model. UNB and OPT proportions are identical for blocks 1, 2, and 3 in experimental conditions, and all blocks in the control condition, given unbiased optimal criterion ($\beta_o = 1$). Significance of comparisons across rate and awareness of change conditions is reported below proportions for average conditions (# p >= .05; * $p < .05$; ** $p < .01$).

Accuracy

Average percent accuracy can be seen by block and condition in Figure 12 and Table 7. A 3 (pre-change blocks 1-3) x 5 (condition; including control) mixed factor ANOVA revealed no effect of condition on average accuracy, $F(4, 140) = 0.62$, $p = .649$, $\eta_p^2 = .02$. Next, a 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA was conducted on average accuracy to evaluate the effect of the base-rate manipulation on average accuracy in experimental conditions relative to the control condition. Average accuracy increased across blocks 4 through 8, $F(3.78, 529.32) = 9.17$, $p < .001$, $\eta_p^2 = .06$, and differed by condition, $F(4, 140) = 11.61$, $p < .001$, $\eta_p^2 = .25$. Critically, there was an interaction of block and condition, $F(15.12, 529.32) = 1.83$, $p = .028$, $\eta_p^2 = .05$, suggesting that accuracy in all experimental conditions did not increase relative to control until blocks 6-8, when the base-rate change was complete.

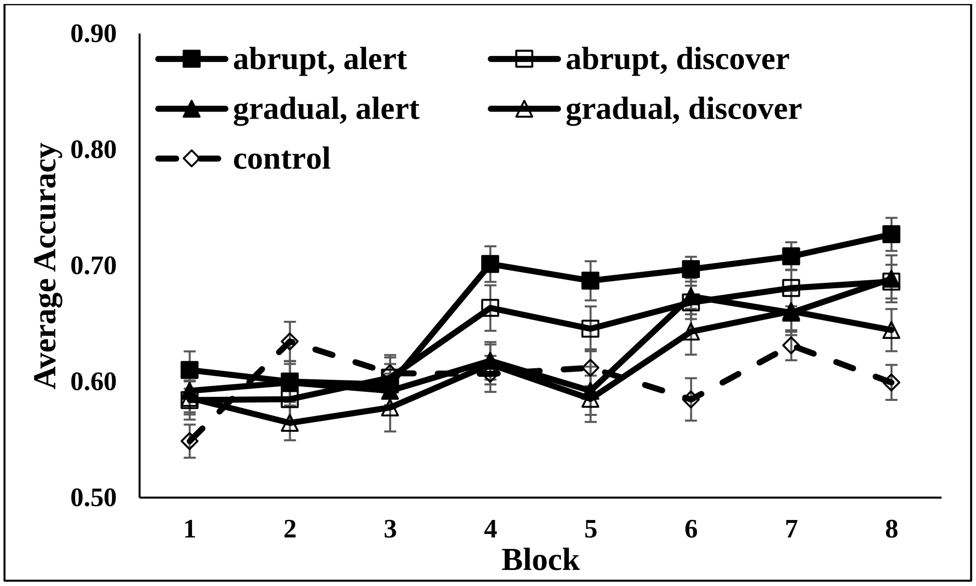

Figure 12. Average accuracy by condition and block in Experiment 1. Error bars represent standard error.

Table 7. Average accuracy (and frequency of responding category A) by condition and block in Experiment 1.

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
Abrupt, Alert (*n* = 32)	0.61 (0.50)	0.60 (0.48)	0.60 (0.50)	0.70 (0.70)	0.69 (0.75)	0.70 (0.76)	0.71 (0.78)	0.73 (0.81)
Abrupt, Discover (*n* = 29)	0.58 (0.49)	0.58 (0.48)	0.60 (0.47)	0.66 (0.64)	0.65 (0.68)	0.67 (0.72)	0.68 (0.74)	0.69 (0.77)
Gradual, Alert (*n* = 27)	0.59 (0.50)	0.60 (0.49)	0.59 (0.48)	0.62 (0.54)	0.59 (0.56)	0.67 (0.72)	0.66 (0.72)	0.69 (0.75)
Gradual, Discover (*n* = 27)	0.59 (0.49)	0.56 (0.49)	0.58 (0.48)	0.61 (0.52)	0.59 (0.54)	0.64 (0.69)	0.66 (0.68)	0.64 (0.69)
Control (*n* = 30)	0.55 (0.49)	0.63 (0.45)	0.61 (0.50)	0.61 (0.49)	0.61 (0.49)	0.58 (0.47)	0.63 (0.49)	0.60 (0.51)

Confidence

Average confidence ratings can be seen by block and condition in Figure 13. A 3 (pre-change blocks 1-3) x 5 (condition; including control) mixed factor ANOVA found no significant differences in average confidence by condition, $F(4, 140) = 0.24$, $p = .914$, $\eta_p^2 = .01$. A subsequent 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA found that average confidence increased across blocks 4-8, $F(3.11, 434.71) = 3.40$, $p = 0.017$, $\eta_p^2 = .02$, but did not differ among conditions, $F(4, 140) = 1.21$, $p = .311$, $\eta_p^2 = .03$, or by block and condition, $F(12.42, 434.71) = 0.87$, $p = .583$, $\eta_p^2 = .02$. Qualitatively, average confidence during and after change was lowest in the gradual/discover condition (i.e., gradual/discover showed the least increase in confidence).

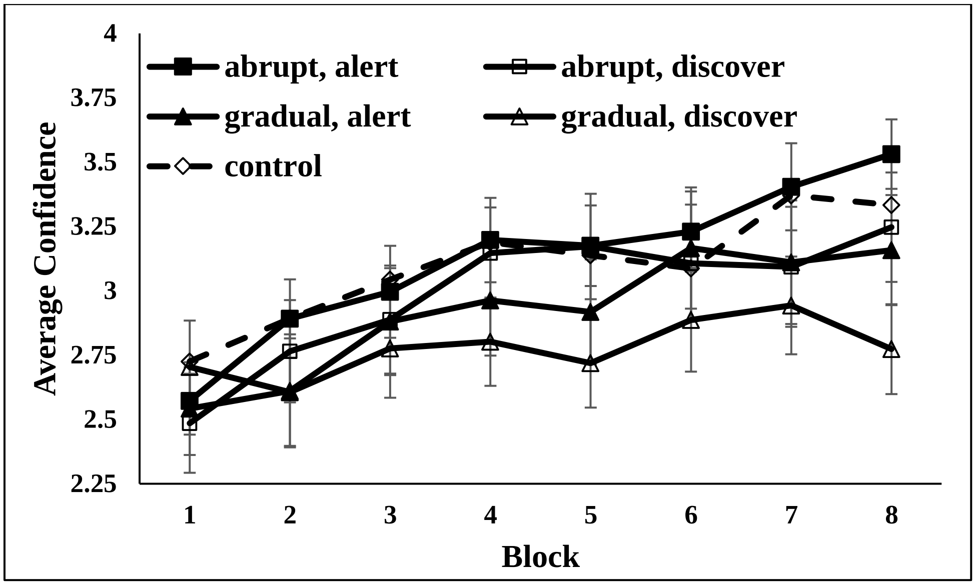

Figure 13. Average confidence by condition and block in Experiment 1. Error bars represent standard error.

Proportion of Category A Responses

The average proportion of Category A responses by block and condition can be seen in Figure 14. A 3 (pre-change blocks 1-3) x 5 (condition; including control) mixed factor ANOVA found no significant differences in the proportion of Category A responses among conditions, $F(4, 140) = 0.21$, $p = .934$, $\eta_p^2 = .01$. A subsequent 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA revealed a main effect of block, $F(3.37, 472.03) = 50.68$, $p < .001$, $\eta_p^2 = .27$, a main effect of condition, $F(4, 140) = 35.62$, $p < .001$, $\eta_p^2 = .50$, and an interaction between block and condition, $F(13.49, 472.03) = 5.67$, $p < .001$, $\eta_p^2 = .14$. All experimental conditions responded category A more frequently, on average, than the control condition – suggesting they were sensitive to the base-rate manipulation (even in gradual change conditions where SDT β-values did not reflect this as much). The frequency of category A responses was greater in blocks 4 and 5 in abrupt than gradual change conditions, as was expected given that the category A base-rate was lower in gradual change conditions over these two change blocks. By the final three blocks, however, there were no significant differences in the proportion of Category A responses. This is particularly interesting given that SDT criterion estimates suggested lower sensitivity to base-rates in gradual conditions – gradual/discover in particular. When proportion of high-base-rate category responses is considered, it appears that all change conditions are sensitive to the change in base-rates.

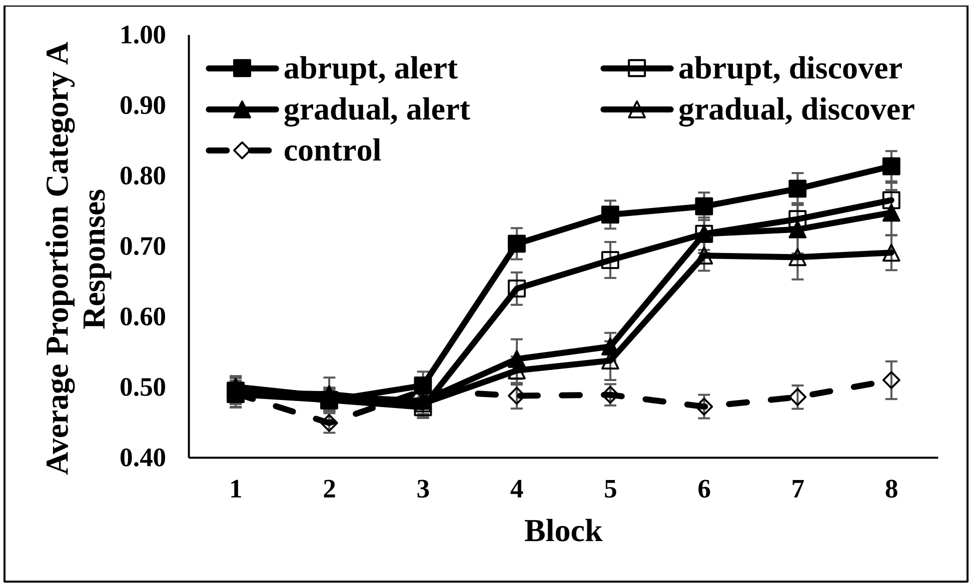

Figure 14. Average proportion of Category A (the high base-rate category following the change) responses by block and condition in Experiment 1. Error bars represent standard error.

Change Effects

The most critical evaluations in Experiment 1 center around the effect of the base-rate change on the dependent measures of interest within rate and awareness of change conditions. As such, the remaining analyses focused only on the four experimental conditions. Three approaches are taken to answer this question for each dependent measure of interest. First, the immediate effect of base-rate change is assessed. Second, the long-term effects of change are investigated (e.g., the ability to recover to pre-change levels of performance when there is a performance cost). The calculation of immediate and long-term effects are based on similar measures – termed cost and recovery when related to drops in performance associated with category transfer – that have been used in previous category learning studies (Cantwell et al., 2015; Maddox, Glass, O'Brien, Filoteo, & Ashby, 2010). A third set of analyses evaluates performance trends at the end of the experiment (blocks 6-8) by rate and awareness of change to determine how change characteristics impact behavior across several post-change blocks.

The immediate effect of change is calculated as a difference score between the final pre-change block (block 3) and the first post-change block (block 4 in abrupt conditions; block 6 in gradual conditions). The immediate effect is calculated as post-pre, such that negative values indicate a decrease in a dependent measure following change. Note that when change is associated with a decrease in performance, the immediate effect can be considered a cost (Maddox et al., 2010). The long-term effect of change is calculated as a difference score between the first and last post-change blocks (Block 8 – block 4 in abrupt change conditions and block 8 – block 6 in gradual change conditions). Since optimal values of dependent measures (e.g., accuracy) change from pre- to post-change, both immediate and long-term change effect analyses compare rate and awareness of change conditions using "difference-from-optimal" measures at each time point. Thus, analyses are a difference of difference scores (e.g., acc-acc$_{opt}$ post – acc-acc$_{opt}$ pre).

Finally, performance trends in post-change blocks are evaluated through 3 (post-change blocks 6-8) x 2 (rate: abrupt, gradual) x 2 (awareness: alert, discover) mixed factor ANOVAs on the different dependent measures. Such an analysis provides information on continued base-rate learning following change and any differences in base-rate sensitivity across change conditions.

Decision Criterion Placement (β)

A pair of 2 (rate) x 2 (awareness) between-subjects ANOVAs were conducted to evaluate the immediate and long-term effects of change with respect to decision criterion placement (β). Since the optimal decision criterion changed from pre-change blocks ($\beta_o = 1$) to post-change blocks ($\beta_o = 3$), the natural log of β-values were analyzed in terms of difference from optimal ($\ln\beta - \ln\beta_o$). The natural log of β-values was used to linearize the scale and make comparisons easier to evaluate. The results revealed no main or interactive effects of rate and awareness of change on the immediate effect of change (all *p*'s > .186) or long-term effect of change (all *p*'s > .313) on $\ln\beta - \beta_o$. See Table 8 for the average immediate effect of change on $\ln\beta - \beta_o$ by condition and Table 9 for the average long-term effect of change on $\ln\beta - \beta_o$ by condition (across all dependent measures).

A 3 (post-change blocks 6-8) x 2 (rate) x 2 (awareness) mixed factor ANOVA revealed a main effect of block, $F(1.69, 187.49) = 8.07$, $p < .001$, $\eta_p^2 = .07$, with significantly larger β-values in blocks 7 and 8 compared to block 6, suggesting asymptotic performance occurred by block 7. There was also a main effect of rate of change on average β-values in post-change blocks, $F(1, 111) = 5.18$, $p = .025$, $\eta_p^2 = .05$, $d = .21$ with larger β-values in abrupt than gradual conditions. There was no main effect of awareness of change on average β values in the post-change blocks, $F(1, 111) = 0.03$, $p = .855$, $\eta_p^2 < .01$.

Table 8. Mean (SD) immediate effect of change by condition in Experiment 1.

Condition	$\ln(\text{SDT-}\beta) - \ln(\beta_o)$	$\ln(\text{GRT-}\beta) - \ln(\beta_o)$	Accuracy ($acc - acc_{opt}$)	Confidence	Proportion Category A Responses
Abrupt, Alert (*n* = 32)	-0.83 (0.36)	-0.44 (1.11)	0.01 (0.12)	0.20 (0.64)	0.20 (0.16)
Abrupt, Discover (*n* = 29)	-0.98 (0.34)	0.00 (1.36)	-0.03 (0.13)	0.26 (0.57)	0.17 (0.12)
Gradual, Alert (*n* = 27)	-0.98 (0.27)	0.17 (1.86)	-0.01 (0.09)	0.30 (0.97)	0.24 (0.16)
Gradual, Discover (*n* = 27)	-1.01 (0.52)	0.08 (1.50)	-0.02 (0.13)	0.11 (0.83)	0.21 (0.14)

Table 9. Mean (SD) long-term effect of change by condition in Experiment 1.

Condition	$\ln(sdt\beta)-\ln(\beta_o)$	$\ln(grt\beta)-\ln(\beta_o)$	Accuracy ($acc - acc_{opt}$)	Confidence	Proportion Category A Responses
Abrupt, Alert (*n* = 32)	0.31 (0.77)	0.61 (1.04)	0.03 (0.12)	0.33 (0.77)	0.11 (0.16)
Abrupt, Discover (*n* = 29)	0.23 (0.68)	0.65 (1.38)	0.02 (0.12)	0.10 (0.77)	0.13 (0.16)
Gradual, Alert (*n* = 27)	0.18 (0.48)	0.04 (1.45)	0.01 (0.10)	-0.01 (0.78)	0.03 (0.14)
Gradual, Discover (*n* = 27)	0.13 (0.40)	0.08 (1.21)	0.00 (0.12)	-0.11 (1.00)	0.00 (0.11)

Decision Bound Modeling Criterion (GRT-β estimate)

A pair of 2 (rate) x 2 (awareness) between-subjects ANOVAs were conducted to evaluate the immediate effect of change (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term effect of change (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) with respect to decision criterion (GRT-β') estimates by the free-boundary model. As with the signal detection β's, GRT-β values were analyzed in terms the difference from optimal, after taking the natural log of each β-value [$\ln(\text{GRT-}\beta)-\ln(\beta_o)$]. There were no main or interactive effects of rate and awareness of change on the difference in the immediate effect of change on $\ln(\text{GRT-}\beta)\text{-}\ln(\beta_o)$, all p's > .211. In terms of the long-term effect of change, there was a main effect of rate of change, $F(1, 111) = 5.72$, $p = .018$, $\eta_p^2 = .05$, with greater increases in GRT-β across post-change blocks in abrupt than gradual change rate conditions. There was no main effect of awareness of change, $F(1, 111) = 0.03$, $p = .865$, η_p^2 .01, or interaction of rate and awareness of change, $F(1, 111) < 0.01$, $p = .978$, $\eta_p^2 < .01$, on the average long-term effect of change on GRT-β estimates. An analysis of GRT β-values within post-change blocks (blocks 6-8) found no significant main or interactive effects of block, rate of change, and awareness of change (all p's > .230).

Accuracy

A pair of 2 (rate) x 2 (awareness) between-subjects ANOVAs were conducted to evaluate the immediate (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) effects of change on accuracy. Since optimal accuracy changed from pre-change 1:1 base-rate blocks ($acc_{opt} = 0.69$) to post-change 3:1 base-rate blocks ($acc_{opt} = 0.78$), accuracy values were analyzed in terms of difference from optimal ($acc\text{-}acc_{opt}$) for each respective block of trials. Positive difference scores indicated an increase in accuracy relative to optimal accuracy, while negative difference scores indicated a decrease in accuracy relative to optimal accuracy.

There were no main or interactive effects of rate and awareness of change on the immediate effect of change (all p's > .195) or long-term effect of change (all p's > .416) on average accuracy in terms of deviation from optimal. Table 8 shows that for all conditions other than abrupt/alert, there is a general cost to deviation from optimal accuracy that is associated with base-rate change. This is likely due to optimal accuracy increasing by 9% over a short period of time. Table 9 reveals that in most cases (except for the gradual/discover condition) accuracy does tend to move closer to optimal over time. However, it seems that neither rate nor awareness of base-rate change impact the immediate or long-term effects of base-rate change on accuracy.

A 3 (post-change blocks 1-3) x 2 (rate) x 2 (awareness) mixed factor ANOVA revealed main effects of both rate of change, $F(1, 111) = 8.55$, $p = .004$, $\eta_p^2 = .07$, and awareness of change, $F(1, 111) = 4.65$, $p = .033$, $\eta_p^2 = .04$, with larger post-change average accuracy in abrupt ($M = 0.70$) compared to gradual ($M = 0.66$) change conditions as well as in alerted ($M = 0.69$) compared to discovery ($M = 0.66$) conditions. No other main or interactive effects were significant (all p's > .252).

Confidence

The immediate (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) effects of change on average confidence were analyzed using a pair of 2 (rate of change) x 2 (awareness of change) between-subjects ANOVAs. There were no differences in the immediate effect of change on average confidence (all p's > .394) or long-term effect of change on confidence (all p's > .075) due to base-rate change, although confidence tended to increase more in abrupt than gradual change rate conditions. From Table 8, it is clear confidence continued to increase on the first post-change block, on average. However, confidence only then continued to increase in abrupt change conditions (see Table 9). An analysis of average confidence within post-change blocks (blocks 6-8) by rate and awareness of change revealed no main or interactive effects (all p's > .083).

Proportion of Category A Responses

The immediate (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) effects of change on the frequency of Category A responses were calculated. From Table 8 it is clear category A response frequency increased immediately following base-rate change, and from Table 9, continued to increase (particularly in abrupt change conditions). The pair of 2 (rate) x 2 (awareness) between-subjects ANOVAs revealed no main effects of rate or awareness of change on the immediate effect of change on category A response frequency (all p's > .160). There was a greater continued increase in the frequency of responding category A post-base-rate change in abrupt than gradual change rate conditions, $F(1, 111) = 13.61$, $p < .001$, $\eta_p^2 = .11$ (all other p's > .435).

A 3 (post-change blocks 6-8) x 2 (rate) x 2 (awareness) mixed factor ANOVA found a main effect of block on average proportion of category A responses, $F(2, 222) = 5.35$, $p = .005$, $\eta_p^2 = .05$, with the average proportion of category A responses increasing across post-change blocks. There was also a main effect of rate of change, $F(1, 111) = 3.02$, $p = .009$, $\eta_p^2 = .06$, such that the average frequency of Category A responses was higher during post-change blocks in abrupt ($M = 0.76$) compared to gradual ($M = 0.70$) conditions.

Discussion

Participants learned to categorize bar graphs simulating the results of a medical test into one of two disease categories. One group served as a control condition and learned the categories with equal base-rates in each block. The four experimental conditions, crossing rate (abrupt, gradual) and awareness (alert, discover) of change experienced a shift in category base-rates from 1:1 to 3:1 in favor of disease category A. Three hypotheses were tested in Experiment 1: 1)

Participants would shift their decision criteria in the optimal direction but display conservative cutoff placement, 2) discovery conditions would be no worse than alert conditions, and 3) performance would be better in gradual than abrupt change conditions.

The signal detection β-values provide support for Hypothesis 1. Participants in the four experimental conditions shifted their decision criteria in the optimal direction, as opposed to the control condition which displayed a stable, unbiased criterion, as expected. Despite shifting in the optimal direction in the experimental conditions, the maximum average SDT-β by the end of the experiment was still conservative ($\beta_{avg} = 2$ vs. $B_{opt} = 3$). Overall, the pattern of SDT-β values across blocks resembles the pattern seen in previous studies, particularly in the abrupt change conditions which most closely mirror previous base-rate conditions. It should be noted that criterion estimates from the GRT models likely overestimated β-values due to an artifact in the modeling procedure (i.e., insufficient constraints on β values), leading to an over-estimation of the β parameter; for this reason, GRT model analysis focuses on qualitative relationships between conditions and statistical tests between models.

If base-rate learning is largely implicit (Bohil & Wismer, 2015; Wismer & Bohil, 2017), then alerting one to base-rate changes should not help. For this reason, base-rate adaptation was predicted to be no different in alerted and discovery conditions. Indeed, there were no differences between alerted and discovery conditions with respect to either the immediate or long term effect of the base-rate change across all dependent measures (SDT and GRT criterion placement, accuracy, proportion of high base-rate category responses, and confidence), providing support for Hypothesis 2. Thus, the present results pertaining to awareness of change in base-rate adaptation provide additional support for claims of the role of implicit learning in base-rate sensitivity (Bohil & Wismer, 2015; Wismer & Bohil, 2017). However, it is important to point out that this conclusion is based on a null-result. Additional research, along with Bayesian data analysis, may be required to corroborate this conclusion.

The rate of change variable (abrupt, gradual) was selected from the machine learning literature, where active approaches (that require change detection prior to adaptation) perform better with abrupt concept drift, and passive approaches (that do not require change detection, but instead continually take new inputs into account for future predictions) perform better with more gradual concept drift. Given evidence that base-rate sensitivity develops largely through the slower, implicit learning system that learns stimulus-response associations (akin to a passive machine learning approach), it was predicted that base-rate adaptation would be greater in response to gradual rather than abrupt base-rate changes. Gradual change blocks provide an opportunity for observers to develop a response bias earlier and thus exhibit a smaller change cost in terms of performance in the new (3:1) base-rate compared to the previous (1:1) base-rate. One might then expect long-term adaptation to base-rate change to be greater in abrupt change conditions where higher initial change costs are expected. However, it is also possible that experience with gradually changing base-rates could heighten sensitivity to base-rates and lead to greater improvements in base-rate sensitivity (i.e., criterion placement) over time, suggesting improved adaptation in gradual change conditions. Therefore, it is possible to observe no difference in long-term adaptation between abrupt and gradual change conditions despite seeing a greater change cost in abrupt conditions.

Actually, Experiment 1 revealed no differences in the immediate effect of base-rate change between abrupt and gradual conditions. Visual inspection of Figure 10, as well as GRT model comparisons within the gradual change blocks, reveal that gradual change conditions did not develop a response bias during gradual change blocks (blocks 4 and 5), contrary to expectations. This provides a possible explanation for why abrupt change conditions did not exhibit greater costs to base-rate change than gradual conditions. Despite a lack of differences with respect to the immediate effect of base-rate change, abrupt conditions exhibited greater long-term adaptation with respect to GRT β estimates, average confidence, and proportion of high base-rate category responses. Together, this disconfirms Hypothesis 3, suggesting that abrupt change may improve base-rate sensitivity compared to gradual change.

However, it is possible that the difference between abrupt and gradual change adaptation is dependent on operational definitions of the immediate and long-term effects of change. The long-term adaptation metric used in the present work was based on three blocks in gradual conditions (block 8 – block 6) but five blocks in abrupt conditions (block 8 – block 4). The final block, used in the calculation of long-term change effects in both abrupt and gradual conditions, was the fifth block of 3:1 base-rates for abrupt change conditions but only the third block of 3:1 base-rates for gradual change conditions. Equating blocks in terms of experience in 3:1 base-rate conditions (i.e., comparing SDT-β's in block 6 of abrupt to block 8 of gradual) reveals no difference in average criterion values among change rate conditions (all p's > .128). Yet, comparing long-term SDT-β adaptation defining adaptation as the difference in β's between the first block post-change in the new base-rate ratio (block 4 abrupt, block 6 gradual) to three blocks later (block 6 abrupt, block 8 gradual) again reveals a main effect of rate of change, $F(1, 111) = 4.13$, $p = .045$, $\eta_p^2 = .04$, with greater increases in average β-values in abrupt ($M = 0.25$) compared to gradual ($M = -0.04$) change rate conditions. These results highlight some of the difficulties in comparing abrupt vs. gradual change in terms of operationally defining the immediate and long-term change effect metrics. Future work that provides lengthened exposure to new base-rates may provide insights into this issue.

An interesting observation in the confidence data reveals that confidence was lowest in the gradual/discover condition, despite evidence that performance did increase in this condition with respect to accuracy, frequency of high base-rate category responses, and criterion shifting following the base-rate change. Implicit learning is associated with knowledge that is hard to verbalize, and it can occur without conscious awareness. The present results then suggest that implicit learning of the new category base-rates did occur.

Despite these interesting findings, of central importance to the question of concept drift in perceptual category learning is the difference in adaptation to changing base-rates vs. payoffs, and the extent to which change characteristics may or may not mediate differences in base-rate/payoff adaptation. Experiment 2 reports the results of an analogous experiment conducted with payoff change replacing the base-rate change used here.

CHAPTER 4: EXPERIMENT 2: PAYOFFS

Payoff Change

The costs and benefits of classification/decision making are an important consideration. Option alternatives must be weighed against their potential outcomes, and after-the-fact outcomes should be valuated to inform future decisions (O'Doherty, Cockburn, & Pauli, 2017). This process is particularly difficult because valuations are prone to change, both from internal motivational state changes as well as external stimulus changes (O'Doherty et al., 2017). For instance, our decisions about what dinner food would satisfy our hunger changes the hungrier or thirstier we get, and smells and sounds can influence the valuation we attach to objects (O'Doherty et al., 2017). As things change both internally and externally, we must be able to shift our decision criteria.

Experiment 2 serves as an initial investigation into the effect of payoff change on classification performance. When one option is more rewarding than another (and there are no costs associated with incorrect decisions), our responses should be biased toward the option with a higher reward, to maximize long-run reward. This type of "no-loss" payoff matrix was selected as a simple way to provide an initial test of adaptation to payoff change, and it provides a comparison to previous work on cost-benefit learning in categorization (Maddox & Bohil, 2000, 2005). As noted in the literature review, previous work has shown sensitivity to category payoffs, but observers have typically 1) displayed conservatism in cutoff placement (i.e., do not shift criterion as far as optimal classifier) and 2) are more conservative than equivalent base-rate ratio conditions. Experiment 2 explores whether payoff adaptation is influenced by rate and or awareness of payoff changes.

Method

Experimental Design

As in Experiment 1, Experiment 2 included four conditions resulting from the factorial combination of rate of change (abrupt, gradual) and awareness of change (alerted, discovered) as well as one control condition:

1. Abrupt, alerted
2. Abrupt, discovered
3. Gradual, alerted
4. Gradual discovered
5. Control condition with no change across the eight blocks

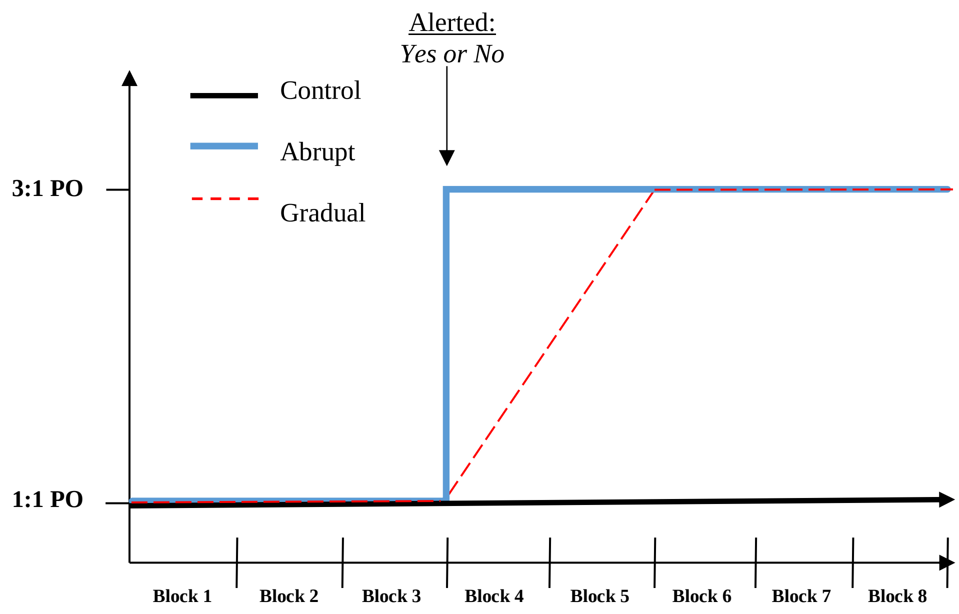

Figure 15. Experimental design for Experiment 2.

Participants

Students from the University of Central Florida and volunteers from the community participated in the experiment. Most students signed up for and participated in the study through the SONA systems participation tool in exchange for course credit. All participants were at least 18 years old and reported normal or corrected-to-normal vision. Each participant completed only one condition of the experiment. There were 25 participants randomly assigned to each of the four experimental conditions, as well as to the control condition, for a total of 125 participants (see Table 10 for participant demographics overall and by condition). This sample size is based on the same power analysis that was run for Experiment 1 since it is a direct comparison to base-rate change in Experiment 1. Indeed, previous work has used equal sample sizes for base-rate and payoff conditions (Maddox & Bohil, 1998, 2003, 2005).

Table 10. Participant demographics by condition in Experiment 2.

	Age		Gender Count	
Condition (*n*'s = 25)	*M*	*SD*	Male	Female
Abrupt, Alert	20.40	3.06	12	13
Abrupt, Discover	21.20	5.24	9	16
Gradual, Alert	20.48	3.66	9	16
Gradual, Discover	20.52	3.08	9	16
Control	18.60	0.96	12	13
Overall (*n* = 125)	20.24	3.53	51	74

Stimuli

The same stimuli and stimuli selection process as in Experiment 1 were used in Experiment 2. The only difference was that category base-rate parameters remained fixed at 1:1 (equal A:B ratio) through the entirety of Experiment 2. Payoff values were displayed – and changed mid-experiment – during Experiment 2 instead.

Procedure

Experiment 2 investigated the influence of category payoff (benefits) change on categorization performance. The procedure for Experiment 2 was identical to Experiment 1 with the exception of change to the payoff parameters rather than base-rate parameter. This manipulation was done by awarding points for correct responses (while receiving 0 points for incorrect responses). The four possible payoff outcomes produced in a two-alternative forced choice classification task can be described in a two-by-two payoff matrix (Respond A/B x Category A/B). Experiment 2 used what is called a no-loss payoff matrix. Points were awarded only for correct responses; incorrect responses did not impact point totals (i.e., 0 points awarded or taken away). Experiment 2 began with a balanced payoff matrix (1:1) where correct responses to both Category A and B stimuli resulted in 100 points for each trial. After the abrupt or gradual change, the new payoff matrix awarded 150 points for correct Category A responses but only 50 points for correct Category B responses (again with 0 points for incorrect responses). No other features of the categories or decision environment changed (See Table 11). Again, half of the participants were alerted to this change in the same manner as in Experiment 1; the other half were left to discover the change. These point values were selected to match the 3:1 ratio change in base-rates in Experiment 1.

Gradual payoff change was implemented in much the same manner as the gradual base-rate change in Experiment 1. On each change trial, beginning with the first trial of block four, the point value for Category A incremented by the point value range (150-100=50) divided by 100 change trials = .50, and the point value for Category B decreased the same amount starting at 100 and ending at 50 by the end of block five.

Table 11. Category statistics for Experiment 2.

Time	Category	μ	σ	Base-rate ratio	Payoff ratio Resp A	Resp B
Start	A	99	21	1:1	100	0
	B	120	21	1:1	0	100
End	A	99	21	1:1	150	0
	B	120	21	1:1	0	50

Point feedback replaced the correct/incorrect category feedback presented in Experiment 1. Participants were told which disease category had been presented, how many points they earned for their response, and how many points they had accumulated for the current block of trials (see Table 12). Block point totals were shown on screen again at the end of the block for 5000ms to inform participants how they did. The maximum point totals for 1:1 and 3:1 payoff conditions are 5000 and 6250, respectively. However, given the low category discriminability (*d'* = 1), optimal point totals are 3450 and 3881 for the 1:1 and 3:1 payoff conditions, respectively (since optimal accuracy is 69% in 1:1 base-rate condition with category *d'* = 1).

Table 12. Example category point (payoff) feedback in Experiment 2.

That was:	Disease A
Points Earned:	100
Your Total Points:	1000

Predictions

Experiment 2 provides an investigation of the influence of change rate and awareness of change on payoffs (i.e., benefits) in category learning in a one-dimensional, rule-based categorization task. Payoffs are assumed to be learned with conscious awareness – at least when payoffs are provided in numerical form on each trial. In categorization, payoffs are typically presented as points for correct or incorrect responses and are summarized in a payoff matrix, similar to Signal Detection Theory (SDT) outcomes. Assuming the explicit nature of payoff sensitivity, the following hypotheses were made:

> Hypothesis 1: Criterion change will be more conservative to payoffs in Experiment 2 compared to base-rates in Experiment 1, in line with the COBRA hypothesis.
>
> Hypothesis 2: Performance (point totals, criterion-shifting) following payoff change will be better in alerted (forewarned) compared to discovery conditions.
>
> Hypothesis 3: Criterion change will be better in response to abrupt rather than gradual payoff change.

Results

The analysis approach followed that of Experiment 1, with the addition of reward (i.e., point total) as a primary dependent measure to assess adaptation to changing payoffs. The same outlier replacement approach was taken as in Experiment 1. The number of outlier data points identified and replaced in each condition can be seen in Table 13 below.

Table 13. Number of outliers in Experiment 2 by condition and measure.

	Beta (β)	Accuracy	Total Data Points (# blocks x # participants)
Abrupt, alert	7	0	200
Abrupt, discover	5	0	200
Gradual, alert	4	1	200
Gradual, discover	5	0	200
Control	3	0	200

Note. No outliers were identified for average confidence, point total, or proportion of Category A responses.

Comparison to Control

Signal Detection Decision Criterion Placement (*β*)

A 3 (pre-change blocks) x 5 (condition; including control) mixed factor ANOVA was conducted on average β-values during pre-change blocks. There was an interaction of block and condition, $F(7.54, 226.06) = 2.58$, $p = .012$, $\eta_p^2 = .08$, such that β-values in block l were larger in the abrupt/discover condition than other conditions, but were similar in blocks 2 and 3 by the end of the pre-change blocks (see Figure 16). Next, a 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA was conducted to evaluate the sensitivity to the payoff change in the experimental conditions compared to the control condition. There was no main effect of block on average β-values in blocks 4-8, $F(1.16, 139.05) = 2.60$, $p = .104$, $\eta_p^2 = .02$, but there was a main effect of condition, $F(4, 120) = 3.85$, $p = .006$, $\eta_p^2 = .11$. Bonferroni post-hoc comparisons revealed that β-values were significantly larger in the abrupt/alert condition compared to both the gradual/discover ($p = .032$) and control ($p = .004$) conditions. There was no interaction of block and condition, $F(4.64, 139.05) = 2.00$, $p = .088$, $\eta_p^2 = .06$; however, it is obvious from Figure 16 that the main effect of condition is driven only by the abrupt/alert condition in block 8. This increase in β in block 8 in the abrupt/alert condition is likely the results of outliers. Three of the 25 participants had extremely large β-values, and only one of these values was removed in the +/- 3SD outlier analysis. Removal of the remaining two large β-values results in a more modest β-estimate of 1.19. (As a side note, once again there was no effect of condition on estimates of *d'* from SDT; all *p*'s > .264).

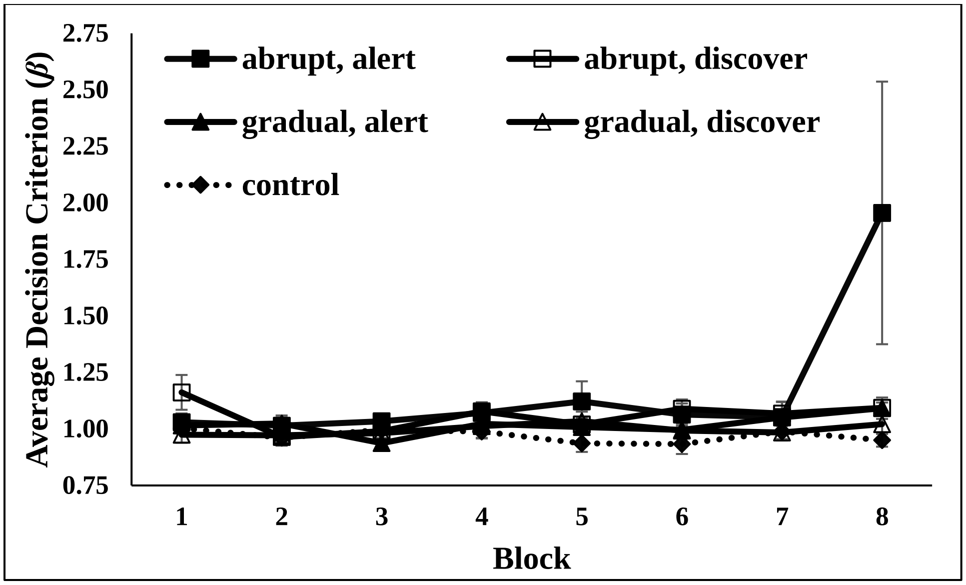

Figure 16. Average SDT-β values by block and condition in Experiment 2.

Decision Bound Models (GRT-β Estimate)

The decision bound modeling analysis for Experiment 2 was the same as for Experiment 1 given that the optimal decision criterion in each block was identical. The same three decision bound models as in Experiment 1 were applied to the data (unbiased, free-boundary, and optimal models). The β's estimated by the free-boundary model were subjected to the same analyses as the SDT-β's (see Figure 17). A 3 (pre-change blocks) x 5 (condition; including control) mixed factor ANOVA revealed a main effect of block on average GRT- β, $F(1.79, 215.18) = 4.89$, p = .011, $\eta_p^2 = .04$, with larger GRT-β estimates in block 1 than block 3 (p = .011). There was also a main effect of condition, $F(4, 120) = 2.95$, p = .023, $\eta_p^2 = .09$, with larger GRT-β estimates in the abrupt/discover condition compared to the abrupt/alert condition (p = .026), but no differences between any of the experimental conditions and the control condition (all p's > .957).

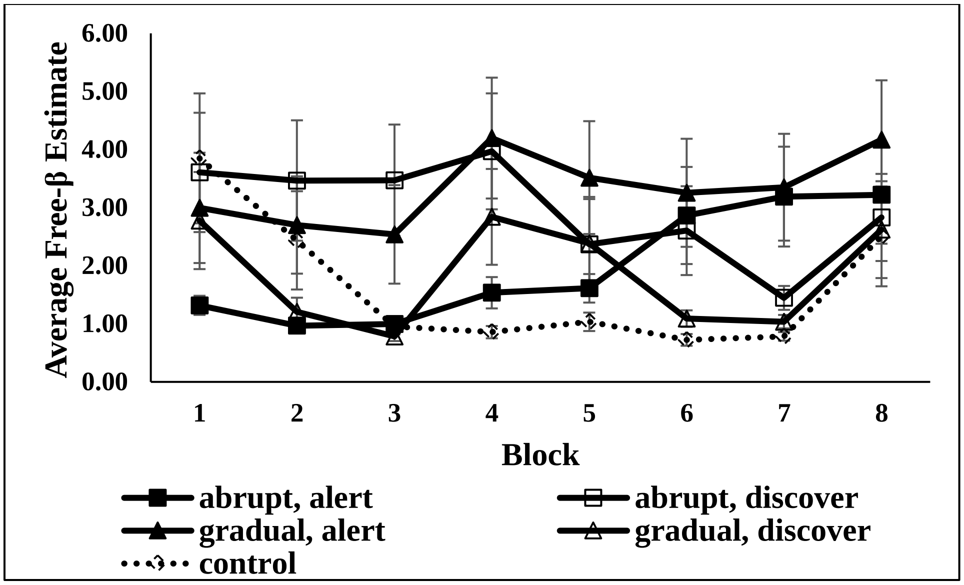

Figure 17. Average GRT-β estimates from free-boundary model by block and condition in Experiment 2.

An analogous ANOVA was conducted on average GRT-β estimates in change/post-change blocks (4-8) to evaluate the effect of base-rate change on GRT-β estimates in the experimental conditions compared to the control condition. There was a main effect of block, $F(3.33, 399.64) = 3.40$, $p = .014$, $\eta_p^2 = .03$, with significantly larger GRT-β estimates in block 8 compared to block 7 ($p = .015$). There was also a main effect of condition, $F(4, 120) = 3.31$, $p = .013$, $\eta_p^2 = .10$, such that GRT-β estimates were significantly higher in the gradual/alert condition compared to the control condition ($p = .006$). No other conditions were significantly different than the control condition (all p's > .433).

Table 14 displays the proportion of participants best fit by each model in each block. The optimal and unbiased models were equivalent for the first three blocks given an unbiased optimal criterion ($\beta_o = 1$). In block 4, more participants were best fit by the unbiased model in abrupt change than gradual change conditions, $z = 2.01$, $p = .044$. In block 5, more participants were best fit by the unbiased model in discover than alert conditions, $z = 2.24$, $p = .025$. This suggests that participants used closer to optimal decision criteria when payoffs changed gradually and when they were alerted to the change. Additional *z*-tests were conducted for each change/post-change block comparing the proportion of participants best fit by the unbiased model in each experimental condition to the control condition. In block 4, control participants were best fit by the unbiased model significantly more often than abrupt/discover, gradual/alert, and gradual/discover conditions (all p's $< .05$), more often than the gradual/alert condition in block five ($p < .01$), and more often than all four experimental conditions in block 7 (all p's $< .05$). These results therefore provide additional support for payoff sensitivity in the experimental conditions, with the highest sensitivity in the gradual/alert condition.

Table 14. Proportion of participants in each condition of Experiment 2 best fit by unbiased, free-boundary, and optimal boundary models.

Condition		Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
Individual									
Abrupt, Alert	UNB	.84	.80	.72	.60	.56	.56	.40	.52
	FRB	.16	.20	.28	.12	.04	.20	.28	.12
	OPT	.84	.80	.72	.28	.40	.24	.32	.36
Abrupt, Discover	UNB	.84	.80	.84	.52	.68	.72	.60	.48
	FRB	.16	.20	.16	.04	.12	.04	.04	.16
	OPT	.84	.80	.84	.44	.20	.24	.36	.36
Gradual, Alert	UNB	.76	.72	.76	.44	.40	.56	.56	.52
	FRB	.24	.28	.24	.16	.24	.12	.08	.12
	OPT	.76	.72	.76	.40	.36	.32	.36	.36
Gradual, Discover	UNB	.96	.76	.72	.28	.72	.76	.64	.52
	FRB	.04	.24	.28	.16	.08	.04	.24	.24
	OPT	.96	.76	.72	.56	.20	.20	.12	.24
Control	UNB	.84	.92	.84	.84	.76	.72	.88	.64
	FRB	.16	.08	.16	.16	.24	.28	.12	.36
	OPT	.84	.92	.84	.84	.76	.72	.88	.64
Average (proportion best fit by unbiased model)									
Abrupt					**.56**	.62	.64	.50	.50
Gradual					**.36**	.56	.64	.60	.50
					*	#	#	#	#
Alert					.52	**.48**	.56	.48	.52
Discover					.48	**.70**	.74	.62	.50
					#	*	#	#	#

Notes. UNB = unbiased boundary model; FRB = free boundary model; OPT = optimal boundary model. UNB and OPT proportions are identical for blocks 1, 2, and 3 in experimental conditions and all blocks for control condition [optimal criterion is unbiased (β_o = 1)]. The bottom portion of the table describes the proportion of participants best fit by a biased model in the post-change blocks. Significance of comparisons across rate and awareness of change conditions is reported below proportions for average conditions (# p >= .05; * $p < .05$; ** $p < .01$).

Reward (Point Total)

A 3 (pre-change blocks 1-3) x 5 (condition; including control) mixed factor ANOVA found no differences in average point totals among conditions, $F(4, 120) = 0.49$, $p = .743$, $\eta_p^2 = .02$. A subsequent 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA was run to evaluate the effect of payoff change on experimental conditions relative to the control condition in change/post-change blocks. There was no main effect of block, $F(3.85, 462.44) = 0.88$, $p = .475$, $\eta_p^2 = .01$, or condition, $F(4, 120) = 0.46$, $p = .769$, $\eta_p^2 = .02$. There was, however, an interaction between block and condition, $F(15.42, 462.44) = 1.72$, $p = .043$, $\eta_p^2 = .05$. Average point total by block and condition can be seen in Figure 18 below.

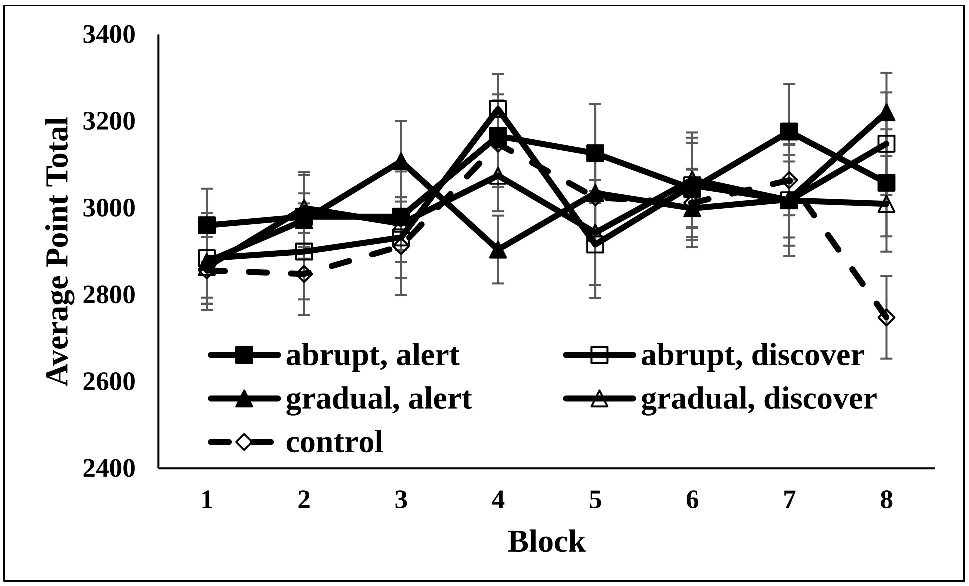

Figure 18. Average point totals by block and condition in Experiment 2. Error bars represent standard error.

Accuracy

There was no main effect of condition on average accuracy in the pre-change blocks, $F(1.92, 7.67) = 0.41$, $p = .804$, $\eta_p^2 = .01$. A subsequent ANOVA on average accuracy in the change/post-change blocks (4-8) found no main or interactive effects of block and condition on average accuracy (all p's > .080). Average accuracy by block and condition can be seen in Figure 19.

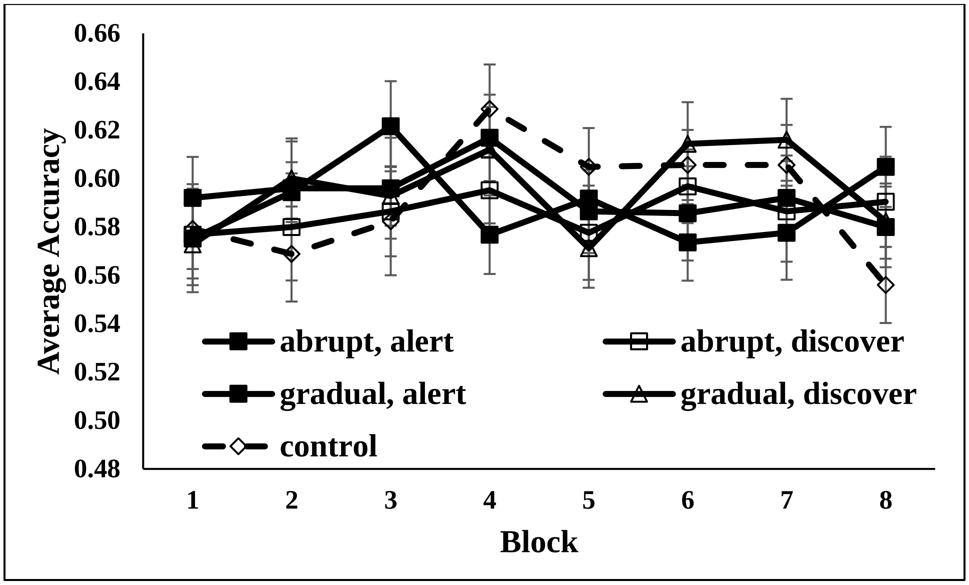

Figure 19. Average accuracy by block and condition in Experiment 2. Error bars represent standard error.

Confidence

A 3 (pre-change blocks) x 5 (condition; including control) mixed factor ANOVA found that average confidence increased across pre-change blocks, $F(1.66, 198.82) = 6.62$, $p = .003$, $\eta_p^2 = .05$, but did not differ by condition, $F(4, 120) = 0.21$, $p = .935$, $\eta_p^2 = .01$, or by block and condition, $F(6.63, 198.82) = 0.46$, $p = .853$, $\eta_p^2 = .02$. A subsequent ANOVA run on change/post-change blocks (4-8) found no main or interactive effects of block and condition (all *p*'s > .485). Average confidence by block and condition can be seen in Figure 20.

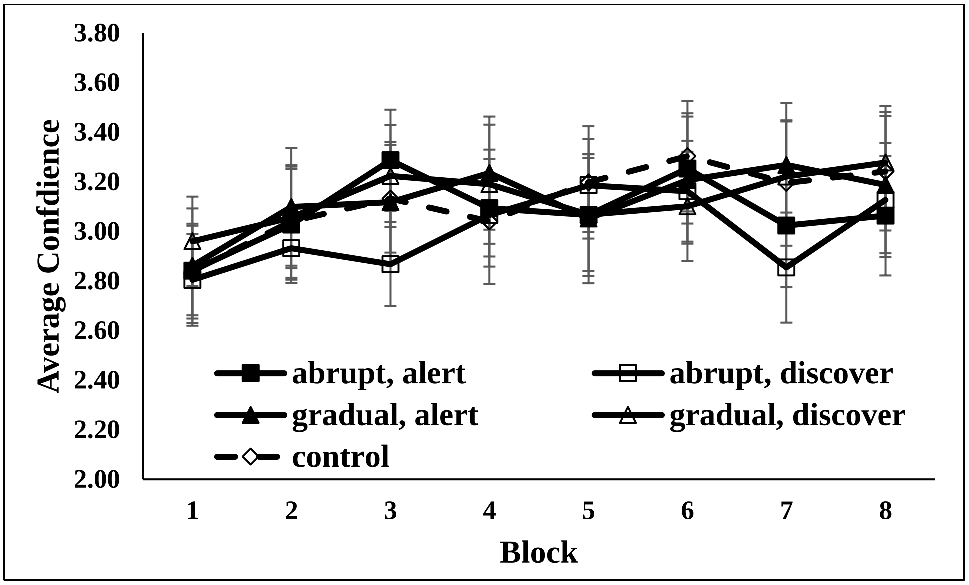

Figure 20. Average confidence by block and condition in Experiment 2. Error bars represent standard error.

<u>Proportion of Category A Responses</u>

There were no differences among conditions with respect to the average proportion of Category A responses in pre-change blocks (all *p*'s > .403). A subsequent ANOVA run on the five change/post-change blocks (4-8) revealed a main effect of condition on the average proportion of Category A responses, $F(4, 120) = 6.29$, $p < .001$, $\eta_p^2 = .17$. Post-hoc comparisons suggest that abrupt/alert ($p < .001$), abrupt/discover ($p = .003$), and gradual/alert ($p = .001$) conditions had significantly higher average proportions of Category A responses compared to the control condition. The increase in proportion of Category A responses in the gradual/discover condition compared to the control condition across blocks 4-8 was not significant ($p = .208$). However, as a whole, these results suggest that the experimental conditions were sensitive to the base-rate manipulations via increased responding of category A (the high base-rate category) compared to the control condition with equal category base-rates.

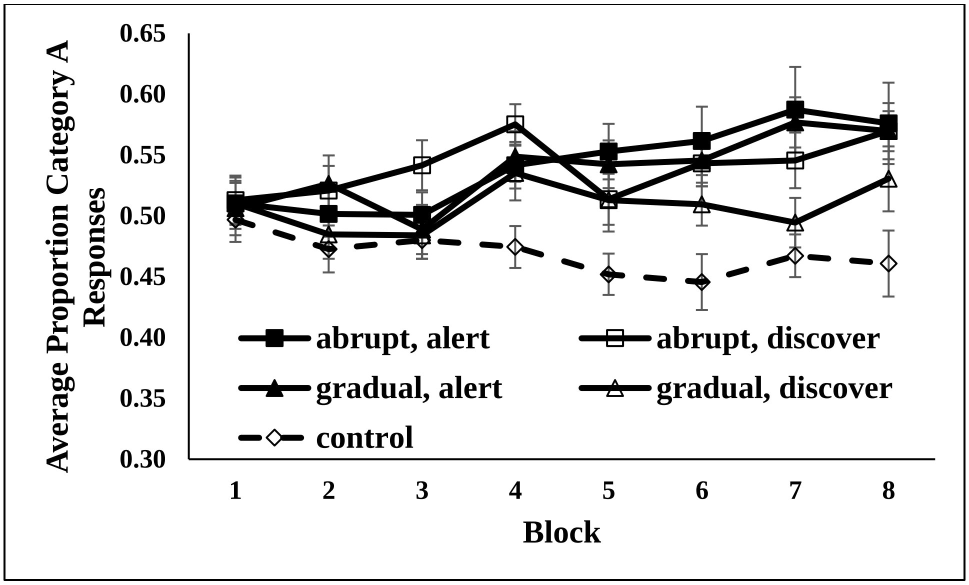

Figure 21. Average frequency of responding with Category A (the high payoff category following the change) by block and condition in Experiment 2.

Change Effects

Criterion Placement (β)

To evaluate the effect of base-rate change on criterion placement, a series of 2 (rate) x 2 (awareness) between-subjects ANOVAs were conducted on each dependent variable with respect to the immediate (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) effects of payoff change. Lnβ-values were again used in the immediate and long-term change effects computations of criterion values to facilitate comparison, and they were calculated as difference scores from $\ln\beta_o$ due to different optimal β values pre- and post-change. There were no main or interactive effects of rate and awareness of change on the immediate (all p's > .079) or long-term (all p's > .136) effect of change on criterion values as measured by $\ln\beta$-$\ln\beta_o$. The immediate and long-term effects of change on each dependent measure can be seen in Table 15 and Table 16, respectively.

A 3 (post-change blocks 6-8) x 2 (rate) x 2 (awareness) mixed factor ANOVA was then conducted on average signal detection β-values in the post-change blocks to evaluate the effects of change characteristics in the blocks following change. The main effect of rate of change on post-change block β-values approached significance, $F(1, 96) = 3.89$, $p = .051$, $\eta_p^2 = .04$, with average β-values tending to be larger in abrupt compared to gradual change conditions, in accordance with Hypothesis 3 - particularly in the final block. There were no other main or interactive effects of rate and awareness of change on β-values in the post-change blocks (all *p*'s > .089).

Table 15. Mean (SD) immediate effect of change by condition in Experiment 2.

Condition (all *n*'s = 25)	ln(SDT-β) - ln(β$_o$)	ln(GRT-β) -ln(β$_o$)	Point Total	Accuracy	Confidence	Proportion Category A Responses
Abrupt, Alert	-1.08 (0.20)	-0.69 (0.82)	186 (457.5)	0.02 (0.07)	-0.19 (0.59)	0.04 (0.10)
Abrupt, Discover	-1.01 (0.13)	-0.88 (0.80)	296 (568.6)	0.01 (0.10)	0.20 (0.72)	0.03 (0.09)
Gradual, Alert	-1.04 (0.17)	-0.71 (0.89)	-108.9 (533.4)	-0.05 (0.09)	0.09 (0.75)	0.06 (0.09)
Gradual, Discover	-1.10 (0.18)	-0.80 (0.62)	101.5 (583.2)	0.02 (0.10)	-0.12 (0.71)	0.03 (0.11)

Table 16. Mean (SD) long-term effect of change by condition in Experiment 2 (payoff change)

Condition (all *n*'s = 25)	ln(SDT-β) - ln(βo)	ln(GRT-β) - ln(βo)	Point Total	Accuracy	Confidence	Proportion Category A Responses
Abrupt, Alert	0.21 (0.77)	0.39 (1.17)	-0.04 (0.10)	-0.04 (0.10)	-0.03 (0.79)	0.01 (0.13)
Abrupt, Discover	0.01 (0.11)	-0.26 (1.00)	-0.01 (0.09)	-0.01 (0.09)	0.06 (0.79)	-0.01 (0.08)
Gradual, Alert	0.08 (0.22)	0.21 (1.01)	0.03 (0.08)	0.03 (0.08)	-0.02 (0.54)	0.02 (0.11)
Gradual, Discover	0.02 (0.30)	0.42 (0.93)	-0.03 (0.07)	-0.03 (0.07)	0.18 (0.85)	0.02 (0.14)

Decision Bound Modeling Criterion (GRT-β estimate)

A pair of 2 (rate) x 2 (awareness) between-subjects ANOVAs were conducted to evaluate the immediate (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) effects of change with respect to decision criteria (GRT-β's) estimated by the free boundary model. As with the signal detection β's, GRT-β values were analyzed in terms of the difference from optimal, after taking the natural log of each β- value [ln(GRT-β)-ln(β_o)]. There were no main or interactive effects of rate and awareness of change on the difference in the immediate effect of payoff change on GRT-β's (all *p*'s > .397). With respect to long-term effects, there was an interaction of rate and awareness of change on average GRT-β estimates, $F(1, 96) = 4.38$, $p = .039$, $\eta_p^2 = .04$, such that alerting made little difference in gradual change conditions but led to greater increases in GRT-β across post-change blocks in abrupt change conditions. There were no main effects of rate or awareness of change on the long-term effects of payoff change on average GRT-β estimates (*p*'s > .240).

Lastly, there was a main effect of block on GRT β-values, $F(1.84, 176.20) = 3.93$, $p = .024$, $\eta_p^2 = .04$, with average β estimates increasing across blocks 6, 7, and 8. There was also a main effect of awareness of change, $F(1, 96) = 4.98$, $p = .028$, $\eta_p^2 = .05$, with higher average β estimates in alert than discover conditions (supporting Hypothesis 2). No other main or interactive effects were found (all *p*'s > .303).

Reward (Point Total)

Next, the immediate (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) effects of payoff change were evaluated with respect to point total (reward). Since optimal point total varied by pre- (3450) and post- (3813) change blocks, reward was analyzed in terms of difference from optimal. A pair of 2 (rate) x 2 (awareness) between-subjects ANOVAs were run to evaluate differences in the immediate effect of payoff change on average point totals among conditions. There was a main effect of rate of change, $F(1, 96) = 5.17$, $p = .025$, $\eta_p^2 = .05$, with closer to optimal point totals in abrupt compared to gradual conditions (providing support for Hypothesis 3) immediately following payoff change. Contrary to Hypothesis 2, there was no main effect of awareness of change, $F(1, 96) = 2.22$, $p = .140$, $\eta_p^2 = .02$, or interaction of rate and awareness of change, $F(1, 0.22)$, $p = .642$, $\eta_p^2 < .01$, on the immediate effect of payoff change on average point totals. A subsequent ANOVA run on long-term effects of payoff change on point totals revealed no main or interactive effects of rate and awareness of change, (all *p*'s > .140). The immediate and long-term effects of payoff change on average point totals, in terms of deviation from optimal, can be seen in Table 15 and Table 16. Finally, a 3 (post-change blocks 6-8) x 2 (rate) x 2 (awareness) mixed factor ANOVA revealed no main or interactive effects with respect to average point totals across post-change blocks (all *p*'s > .142).

Accuracy

A pair of 2 (rate) x 2 (awareness) between-subjects ANOVAs were conducted on the immediate (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and long-term (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) effects of payoff change on accuracy. Since payoff manipulations do not affect the theoretical optimal accuracy of 69% (despite reward maximization being associated with a small decrease in accuracy), accuracy cost and recovery was analyzed using raw accuracy values. There was no main effect of rate of change, $F(1, 96) = 2.41$, $p = .124$, $\eta_p^2 = .02$, or awareness of change, $F(1, 96) = 2.55$, $p = .114$, $\eta_p^2 = .03$ on average accuracy cost, but there was a significant interaction between rate and awareness of change on average accuracy cost, $F(1, 96) = 5.11$, $p = .026$, $\eta_p^2 = .05$, with no difference in accuracy cost in abrupt conditions, but greater cost in alerted than discovered gradual change rate conditions (evidence against Hypothesis 2 but in unexpected direction).

A similar pattern was found for accuracy recovery, with no main effects of rate of change, $F(1, 96) = 1.50$, $p = .224$, $\eta_p^2 = .02$, or awareness of change, $F(1, 96) = 0.88$, $p = .351$, $\eta_p^2 = .01$, but a significant interaction between rate and awareness of change on average accuracy recovery, $F(1, 96) = 8.17$, $p = .005$, $\eta_p^2 = .08$. Accuracy recovery was greater in alerted than abrupt conditions when payoff change occurred gradually, but worse when change occurred abruptly. Average accuracy cost and recovery can be seen in Table 15 and Table 16.

Finally, a 3 (post-change blocks 6-8) x 2 (rate) x 2 (awareness) mixed factor ANOVA was conducted on average accuracy in the final 3 blocks of the experiment. There were no main or interactive effects of block, rate, and awareness of change on average accuracy (all p's > .125).

Confidence

Raw confidence averages were analyzed in terms of the immediate (the immediate impact of payoff change on confidence; block 4 – block 3 in abrupt change, block 6 – block 3 in gradual change) and long-term (the change in confidence following exposure to new category payoffs; block 8 – block 4 in abrupt change, block 8 – block 6 in gradual change) effects of payoff change. There was no main effect of rate of change, $F(1, 96) = 0.02$, $p = .892$, $\eta_p^2 < .01$, or awareness of change, $F(1, 96) = 0.42$, $p = .521$, $\eta_p^2 < .01$ on the immediate effect of payoff change on average confidence. The interaction of rate and awareness of change was significant, $F(1, 96) = 4.71$, $p = .032$, $\eta_p^2 = .05$, with an increase in confidence post-change in the gradual/alert condition but a decrease in confidence post-change in the abrupt/alert condition, with the opposite pattern in discover conditions. An analogous analysis of the long-term effects of payoff change on average confidence found no main or interactive effects of rate and awareness of change (all p's > .339). The immediate and long-term effects of payoff change on average confidence can be seen in Table 15 and Table 16.

An analysis of confidence in the final three blocks of the experiment (post-payoff change) revealed a significant interaction between block and rate of change, $F(1.91, 183.69) = 3.63$, $p = .030$, $\eta_p^2 = .04$, such that average accuracy was no different between abrupt and gradual conditions at block 6 but higher in gradual conditions in blocks 7 and 8. There were no other main or interactive effects (all p's > .298).

Proportion of Category A Responses

There were no main or interactive effects of change on the immediate effect of payoff change on the average proportion of Category A responses (all p's > .332), nor were there on long-term effects of payoff change (all p's > .389). The immediate and long-term effects of payoff change on the average proportion of Category A responses can be seen in Table 15 and Table 16. The average proportion of Category A responses post-payoff change tended to be higher in alert than discover conditions, although this difference was not significant, $F(1, 96) = 3.18$, $p = .078$, $\eta_p^2 = .03$. No other main or interactive effects of block, rate, or awareness of change were found on the proportion of category A responses during blocks 6-8.

Base-rate vs. Payoff Change Adaptation

Previous work has found more optimal criterion placement in base-rate than payoff conditions, likely due to the interaction of accuracy and base-rates/payoffs (payoff maximization with equal base-rates requires accuracy sacrifice, whereas accuracy & payoff can be maximized together in base-rate conditions). To evaluate whether change in either category base-rates or payoffs affects human category learning differently, a series of 2 (condition: base-rate, payoff) x 2 (rate: abrupt, gradual) x 2 (awareness: alert, discover) between-subjects ANOVAs were conducted on signal detection and GRT β's, accuracy, average confidence, and the proportion of Category A responses (the category associated with higher base-rate/payoff post-change).

Signal Detection Decision Criterion Shifting (β)

A first ANOVA was conducted on lnβ values with respect to the immediate and long-term effects of change. There was a main effect of condition on the immediate effect of change on average lnβ, $F(1, 207) = 213.01$, $p < .001$, $\eta_p^2 = .51$, suggesting that participants in base-rate change conditions were better able to immediately shift their decision criteria than participants in payoff change conditions. There was also a main effect of rate of change on the immediate effect of change on lnβ, $F(1, 207) = 113.27$, $p < .001$, $\eta_p^2 = .35$, such that there was greater immediate criterion change in abrupt than gradual change rates. The interaction of condition and rate of change was also significant, $F(1, 207) = 166.29$, $p < .001$, $\eta_p^2 = .45$, suggesting rate of change did not impact immediate criterion shifting in response to base-rate change, but for payoff changes, immediate criterion shifting was better in response to abrupt than gradual change. Finally, long-term criterion adaptation was greater in base-rate than payoff conditions, $F(1, 207) = 7.28$, $p = .008$, $\eta_p^2 = .03$, with no main or interactive effects of rate and awareness of change (all p's > .242).

Decision Criterion Shifting (GRT-β Model Estimates)

A second set of ANOVAs was conducted on ln-GRT-β-values with respect to the immediate and long-term effects of change (ln values were selected to match signal detection β analyses). There was a main effect of condition on the immediate effect of change on ln-GRT-β, $F(1, 207) = 19.29$, $p < .001$, $\eta_p^2 = .09$, with greater criterion shifting in base-rate than payoff conditions. There was also an interaction of condition and rate of change on long-term effects of change on ln-GRT-β, $F(1, 207) = 6.46$, $p = .012$, $\eta_p^2 = .03$, such that abrupt change led to greater criterion shifting than gradual change in response to base-rate change, but rate of change had no effect on long-term adaptation to payoff change.

Accuracy

A third set of ANOVAs was conducted on accuracy values with respect to the immediate and long-term effects of change. There was a main effect of condition on the immediate effect of change on average accuracy, $F(1, 207) = 27.00$, $p < .001$, $\eta_p^2 = .12$, such that average accuracy increased to a greater extent following base-rate than payoff change. There was also an interaction between condition and awareness of change, $F(1, 207) = 3.90$, $p = .050$, $\eta_p^2 = .02$, such that the benefit of alerts was greater in base-rate than payoff conditions. There were no significant main or interactive effects with respect to long-term adaptation to change in terms of accuracy (all p's > .067). However, when accuracy adaptation was recalculated in terms of maximum accuracy achieved pre-change vs. post-change, a significant main effect of condition was found, $F(1, 207) = 48.67$, $p < .001$, $\eta_p^2 = .19$, suggesting that across all change types, base-rate change leads to greater increases in maximum accuracy than payoff change.

Confidence

A fourth set of ANOVAs was conducted on the immediate and long-term effects of change on average confidence. There was a main effect of condition on the immediate effect of change on confidence, $F(1, 207) = 5.00$, $p = .026$, $\eta_p^2 = .02$, with larger increases in confidence following base-rate change than payoff change. There was an interaction of rate and awareness of change, $F(1, 207) = 4.47$, $p = .036$, $\eta_p^2 = .02$, such that participants who were alerted were more confident in gradual than abrupt change conditions, while participants who were not alerted were more confident in abrupt change conditions compared to gradual change conditions. There were no main or interactive effects on long-term effects of change on average confidence (all *p*'s > .116).

Proportion of Category A Responses

A fifth set of ANOVAs was conducted on the immediate and long-term effect of average proportion of category A responses. There was a main effect of condition on the immediate effect of change on the proportion of category A responses, $F(1, 207) = 90.25$, $p < .001$, $\eta_p^2 = .30$, such that the proportion of Category A responses increased to a greater extent following base-rate than payoff change. There was also a main effect of condition, $F(1, 207) = 9.28$, $p = .003$, $\eta_p^2 = .04$, such that increases in the proportion of Category A responses were greater in base-rate than payoff conditions. There was also a main effect of rate of change on the long-term effect of change on average proportion of Category A responses, $F(1, 207) = 4.83$, $p = .029$, $\eta_p^2 = .02$, with greater increase in the average proportion of Category A responses in abrupt than gradual change rate conditions. Finally, there was an interaction of condition and rate of change, $F(1, 207) = 10.94$, $p = .001$, $\eta_p^2 = .05$, such that abrupt change led to larger increases in the proportion of Category A responses across post-change blocks compared to gradual change in base-rate conditions but not payoff conditions.

Discussion

Payoff Change Adaptation

As in Experiment 1, participants in Experiment 2 learned to categorize bar graphs simulating the results of a medical test into one of two disease categories. This time, instead of a shift in category base-rates, there was a change in the payoffs (i.e., the point values for correct responses) mid-experiment. On each trial, participants were awarded points for correct Category A and Category B responses. These point values were equivalent (i.e., 1:1 payoff ratio) for the first three blocks but shifted after block 3 to favor Category A (i.e., 3:1 payoff ratio). There were again four experimental conditions (where change occurred abruptly or gradually, and participants were either alerted or not to the change) and one control condition with no change in payoffs. There were three hypotheses of interest: 1) decision criterion placement following payoff change would be conservative relative to that of base-rate change in the first experiment, 2) alerting one to changes in payoffs would lead to improved performance relative to discovery conditions, and 3) decision criterion shifting would be better (i.e., closer to optimal) in abrupt than gradual change rate conditions.

Criterion shifting was much more conservative in response to changes in payoffs in Experiment 2 than base-rates in Experiment 1. In fact, there was no significant increase in criterion values post-payoff change. These results ultimately support Hypothesis 1, given more conservative β-values in Experiment 2 than Experiment 1. However, the overall pattern of decision criterion placement in Experiment 2 is also more conservative than what has been found in previous studies using stationary, 3:1 payoff ratios (Bohil & Maddox, 2001; Maddox & Bohil, 2000), although is similar to corresponding payoff condition in (Maddox & Bohil, 2005). Future research might consider the possibility that payoff changes are more difficult to adapt to than base-rate changes in cases of low perceptual sensitivity.

Hypothesis 2 received some support in the present experiment. There was an interaction of rate and awareness of change on the long-term effect of change on average GRT β estimates, such that alerting made no difference in gradual change conditions but led to greater increases across post-change blocks (i.e., criterion shifting in optimal direction) in abrupt change conditions. Within the post-change blocks (blocks 6-8), alerting was associated with larger GRT β-values than discovery conditions. There was a greater immediate accuracy cost followed by recovery in gradual, alerted than gradual, discover conditions, while alerting made no difference in abrupt change conditions. Finally, confidence increased post-change in gradual, alert but decreased in abrupt, alert. Alerting was associated with (non-significant) increases in the frequency of category A responses during blocks 6-8. Post-change GRT β-values larger in gradual/alert compared to control, and more unbiased in discover than control via model comparisons in block 5 (with blocks 6 and 7 similar).

Hypothesis 3 also received some support in the present experiment. GRT models suggested that in block 4, more participants were best fit by an unbiased model in abrupt than gradual conditions. However, this is expected since the payoff change occurred in full prior to block 4 in abrupt conditions, whereas the payoff change was only beginning in the gradual condition (and was associated with a much more conservative optimal criterion; $\beta_{opt} = 1.5$). The GRT β-estimates post-change (blocks 6-8) were also larger in abrupt compared to gradual change conditions. Finally, the analysis of the immediate effect of change suggested that abrupt change was associated with closer to optimal point totals following change than gradual conditions (in other words, abrupt change had an immediate improvement in point totals not seen in gradual change conditions). Therefore, despite large increases in β or accuracy in Experiment 2, the results support the hypothesis that abrupt payoff change is easier to adapt to than gradual payoff change.

Summary of Base-rate vs. Payoff Change Comparison

Adaptation to change within base-rate and payoff change conditions generally followed what is expected based on the criterion placement in stationary base-rate and payoff conditions. Criterion shifting was conservative in both conditions, but was closer to optimal in base-rate than payoff conditions. Participants shifted their decision criteria earlier, and farther, in response to base-rate change than payoff change. Gradual change was expected to support base-rate adaptation more than payoff adaption; however, this hypothesis was not supported. Instead, abrupt change supported adaptation in both base-rate and payoff conditions more than gradual change. It was also anticipated that alerting would be more beneficial in payoff than base-rate change conditions. This hypothesis was not supported either. Contrary to expectations, alerting supported greater base-rate adaptation compared to payoff adaptation. Alerting actually reduced the decreased accuracy change costs to a greater extent in base-rate than payoff conditions. While generalizations from these findings must be cautioned given the lack of payoff adaptation in Experiment 2 – likely due to insufficient motivation to maximize points (e.g., they were not paid) - the results do suggest that the hypothesized distinction between implicit and explicit learning in base-rate and payoff learning needs to be reconsidered.

CHAPTER 5: EXPERIMENT 3: RB ANALOGICAL TRANSFER

Analogical Transfer

Analogical transfer is the ability to transfer knowledge despite large differences in surface features (Casale et al., 2012). Thus, this type of generalization is often considered in category learning. In the present work, as in Casale et al. (2012), analogical transfer is taken to mean a change in the average category features without a change in the optimal decision criterion. In other words, category distributions shift to a new area of stimulus space along the same (learned) criterion. Thus, while the average features may change (e.g., lines may appear longer and angles may appear more acute), application of the rule learned prior to transfer will result in positive transfer performance (see Figure 9).

Analogical transfer has received some attention in the COVIS literature (e.g., Casale et al., 2012; Helie et al., 2015), but the problem has not been studied from the perspective of concept drift. More specifically, there has not been a systematic exploration of change characteristics that might affect transfer adaptation. In particular, gradual change rates have not been considered nor compared to abrupt change rates. Experiment 3 provides an initial exploration of analogical transfer in perceptual category learning and a comparison between explicit and implicit-based category structures (following Experiment 4).

Perceptual categories in the natural environment are subject to change (i.e., display concept drift). Because of the distinction made earlier between categories and concepts, a more general term for the phenomenon of concept drift in perceptual category learning might be categorical change. Going back to a previous example, the symptoms that underlie a particular illness might change over time, as is often the case when diagnostic measures improve.

Sometimes these new symptoms might signify a change in the illness classification, while other times it might simply be a different symptom profile for the same category of illness.

Given the understanding that i) perceptual categories can change over time, and ii) widespread evidence suggests the role of multiple learning systems involved in perceptual category learning (Ashby & Maddox, 2005), it makes sense to explore the way(s) in which categories dominated by one learning system or another influence change performance. Experiments 3 and 4 should provide a replication of the work by Casale et al. (2012) and others finding shift costs related to II but not RB category structures. The manipulation of change rates and alerting to change may uncover conditions in which II analogical transfer is more successful. Experiments 3 and 4 will lay the groundwork for a fuller understanding of human change detection, adaptation, and transfer.

In machine learning, analogical transfer is often referred to as population drift (Hand, 2006; Kelly, Hand, & Adams, 1999), covariate shift (Shimodaira, 2000; Yamazaki, Kawanabe, Watanabe, Sugiyama, & Müller, 2007), or sample selection bias (Hand, 2006; Huang, Gretton, Borgwardt, Schölkopf, & Smola, 2007). Analogical transfer occurs when the input distribution (or initial category distribution) differs from later distributions, despite conditional probabilities staying the same (Alaiz-Rodríguez & Japkowicz, 2008). Alaiz-Rodríguez and Japkowicz (2008) note that this type of change can be seen in banking, medical diagnosis, and bioinformatics. Experiments 3 and 4 do not explore class definition change - the case where $p(x)$ remains but $p(C|x)$ changes, resulting in a new mapping of features to category labels (Alaiz-Rodríguez & Japkowicz, 2008). Rather, Experiment 3 and 4 address analogical transfer in which $p(x)$ changes (i.e., average category features change) with $p(C|x)$ remaining the same (i.e., same classification boundary). Experiment 3 addresses analogical transfer in rule-based (RB) categories, while Experiment 4 evaluates analogical transfer in information-integration (II) category structures.

Method

Experimental Design

In Experiment 3, four experimental conditions resulting from the factorial combination of the independent variables (all between-subjects) rate of change (abrupt, gradual) and awareness of change (alert, discover) were investigated, along with a control condition:

1. RB: Abrupt change, alerted in instructions
2. RB: Abrupt change, no alerting (i.e., discovered)
3. RB: Gradual change, alerted in instructions
4. RB: Gradual change, no alerting (i.e., discovered)

One additional condition serving as an RB control was examined with stationary category distributions.

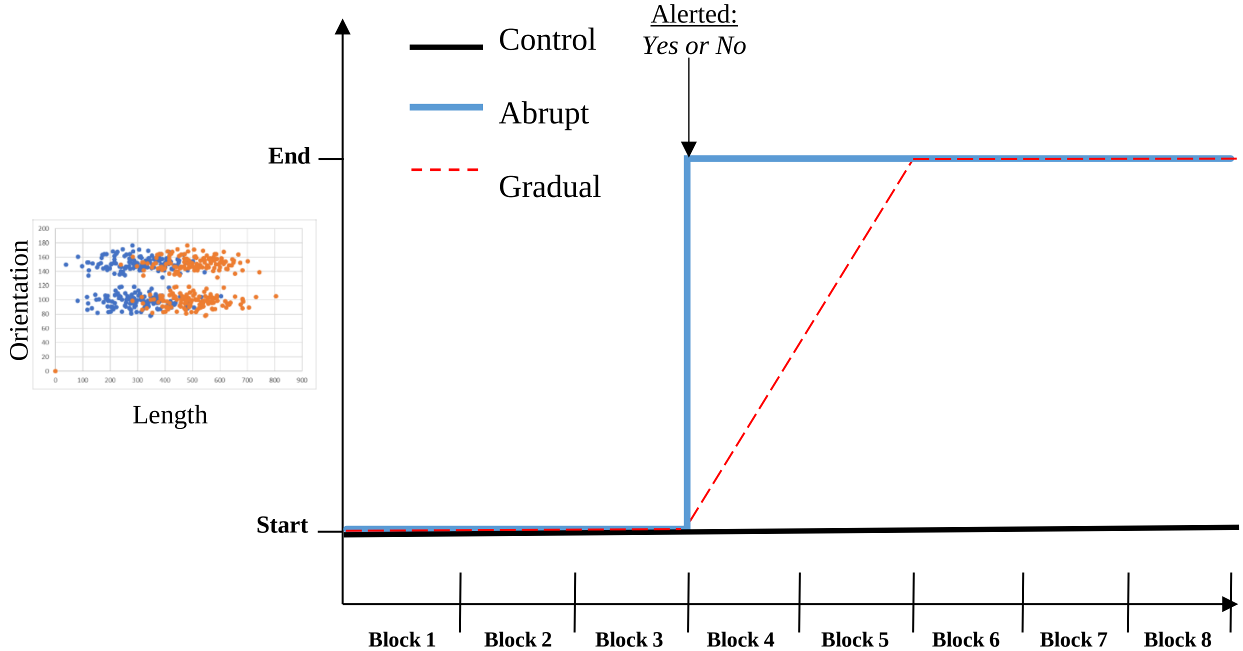

Figure 22. Experimental design for Experiment 3.

Participants

One hundred and twenty-five students from the University of Central Florida and volunteers from the community participated in the experiment. Most participants (students) signed up for and participated in the experiment through the SONA systems participation tool in exchange for course credit. All participants were at least 18 years of age and reported normal or corrected-to-normal vision (see Table 17). Each participant completed only one condition of the experiment, with 25 participants per each of the 5 conditions. This sample size is similar to previous work in analogical transfer (Cantwell et al., 2015; Helie et al., 2015), larger than (Casale et al., 2012), and fulfills the suggestion of an a priori power analysis conducted using G*Power 3.1.9 (Faul et al., 2007; power = .98, f = .40, alpha = .05, numerator df 1,# of groups = 2) using a large effect size from Experiment 2 of Casale et al. (2012) evaluating the interaction of block and condition (RB/II).

Table 17. Participant demographics for Experiment 3.

	Age		Gender Count	
Condition (*n*'s = 25)	*M*	*SD*	Male	Female
Abrupt, Alert	18.36	0.86	12	13
Abrupt, Discover	18.52	1.53	10	15
Gradual, Alert	18.64	1.82	12	13
Gradual, Discover	18.84	2.21	9	16
Control	18.20	0.50	5	20
Overall (*n* = 125)	18.51	1.51	48	77

Stimuli

Two-dimensional stimuli (i.e., lines) were created that varied in length and orientation (angle). Two categories (A and B) were sampled from overlapping normal distributions with different mean lengths and orientations but identical standard deviations (see Table 18). RB stimuli were created via the Ashby and Gott "randomization technique" (Ashby & Gott, 1988) using the Matlab GRT toolbox (Alfonso-Reese, 2006). RB stimuli were created with a *d'* of 6 based on previous work showing appropriate learning over the course of an experiment this length, and the categories were defined on the orientation dimension (i.e., length provided no important information for discriminating categories). Three RB data sets were created: 1) a control dataset with stationary RB category distributions throughout the 400 trials, 2) an abrupt change RB dataset with 150 pre-change trials and 250 post-change trials with new category means on the irrelevant dimension (i.e., length), and 3) a gradual change RB dataset with the same 150 pre-change trials, followed by 100 trials with category means incremented appropriately, and finally 150 post-change trials matching the other datasets. Stimuli were presented at a 1:1 base-rate with no payoffs provided. All participants saw identical stimuli in the first 150 trials (i.e., the pre-change blocks) with all change blocks taking the stimulus values in the control dataset and incrementing category means appropriately based on the magnitude/rate of change.

Table 18. RB Category statistics for Experiment 3.

Condition	Category	μ_x	μ_y	σ_x	σ_y	Cov_{xy}
RB Start	A	300	152	100	9	0
	B	300	98	100	9	0
RB End	A	500	152	100	9	0
	B	500	98	100	9	0

All conditions in Experiment 3 used equal base-rate and payoff ratios (although payoffs were not included). The only parameter of change was the mean value of the categories on the irrelevant, length dimension. The relationship of category A to category B never changed; only the stimulus space of the selected stimuli changed. In gradual change conditions, category stimulus values were equivalent to the respective stimuli in the control condition but the mean value of the critical dimension(s) increased incrementally based on the total amount of change from pre- to post-change. For example, if the mean line length was 200 pixels before the change and 400 pixels after the change, the mean line length would increase by (400-200)/100 change trials = 2 pixels on each change trial. Note that the definition of "long" versus "short" remains the same.

Procedure

Participants provided verbal informed consent prior to study participation. They were then directed to the research lab and seated at a computer with a high-resolution monitor. The categorization task was employed through MATLAB via the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Participants were provided instructions on how to complete the task using the 'A' and 'B' labeled keys (the 'z' and 'm' keys, respectively) on the keyboard on each trial. They were also instructed that on some trials they would be asked to provide a rating of confidence on a scale from 1 (very unsure) to 5 (very confidence) using these respective keys. Participants were instructed to read through the instructions on the screen before beginning the task.

On each trial, one randomly selected stimulus was presented on the center of the screen. The stimulus remained onscreen until the participant made a response, or for 10 seconds. If the participant did not respond within 10 seconds, they were notified that they took too long and the trial repeated. On 17 random trials per 50-trial block (i.e., 1/3 of all trials; randomized across participants), the categorization response was followed by a prompt asking for a confidence rating prior to receiving correct or incorrect feedback on their categorization response. The participant again had 10 seconds to respond with a confidence rating, at which point categorization feedback was displayed for 750ms. The inter-trial interval was 1000ms.

Participants completed eight 50-trial blocks. Participants could take short breaks between blocks as needed when an instruction screen was presented as a reminder of the task. The category distributions shifted upon completion of the third block of trials, either abruptly or gradually depending on the experimental condition. At the end of the experiment, the participant was thanked, awarded credit for participation, and dismissed.

Predictions

Experiment 3 investigates rule-based category transfer by sampling from category distributions that are shifted along the optimal decision criterion (i.e., category members are drawn from a new region of perceptual space while the optimal criterion maintains the same slope and intercept values). Because use of the same decision criterion pre- and post-change would result in the same level of performance, it was not expected that change rates (abrupt or gradual) or awareness of change (alerted or discovered) will impact performance, at least in terms of accuracy or rule usage. However, it is possible these variables have an impact on confidence. Confidence was not investigated in previous relevant examples of analogical transfer in RB (and II) category learning.

Hypothesis 1: RB-change conditions will exhibit analogical transfer (i.e., performance following change will be no different than control condition).

Hypothesis 2: Awareness of change (alert, discover) will have no effect on performance in RB-change conditions.

Hypothesis 3: Rate of change (abrupt, gradual) will have no effect on performance in RB-change conditions.

Results

Comparison to Control

Accuracy

A 3 (pre-change blocks 1-3) x 5 (condition; including control) mixed factor ANOVA was conducted on average accuracy in the pre-change blocks. While average accuracy increased across pre-change blocks, $F(1.21, 145.23) = 106.14$, $p < .001$, $\eta_p^2 = .47$, there was no main effect of condition on average accuracy, $F(4, 120) = 0.40$, $p = .807$, $\eta_p^2 = .01$, or interaction of block and condition, $F(4.84, 145.23) = 0.53$, $p = .747$, $\eta_p^2 = .02$. (As a side note, there was again no effect of condition on *d'*, $F(4, 120) = 1.14$, $p = .343$, $\eta_p^2 = .036$).

Next, a 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA was conducted on average accuracy to evaluate the influence of analogical transfer on accuracy in the experimental conditions relative to accuracy in the control condition. There was a main effect of block, $F(3.85, 462.44) = 4.97$, $p < .001$, $\eta_p^2 = .04$, with significantly lower average accuracy in blocks 4 ($p = .022$) and 5 ($p = .008$) compared to block 8. There was also a main effect of condition, $F(4, 120) = 5.65$, $p < .001$, $\eta_p^2 = .16$. Average accuracy was significantly higher than control in abrupt/discover ($p < .001$), gradual/alert ($p = .012$), and gradual/discover ($p = .001$) conditions, with the increase in accuracy in abrupt/alert approaching significance ($p = .068$). However, there was also an interaction of block and condition, $F(15.42, 462.44) = 1.69$, $p = .049$, $\eta_p^2 = .05$, such that accuracy was initially higher in block 4 in discover conditions compared to alert conditions, and average accuracies fluctuated across the remaining blocks by condition. Average accuracy can be seen below by block and condition in Figure 23.

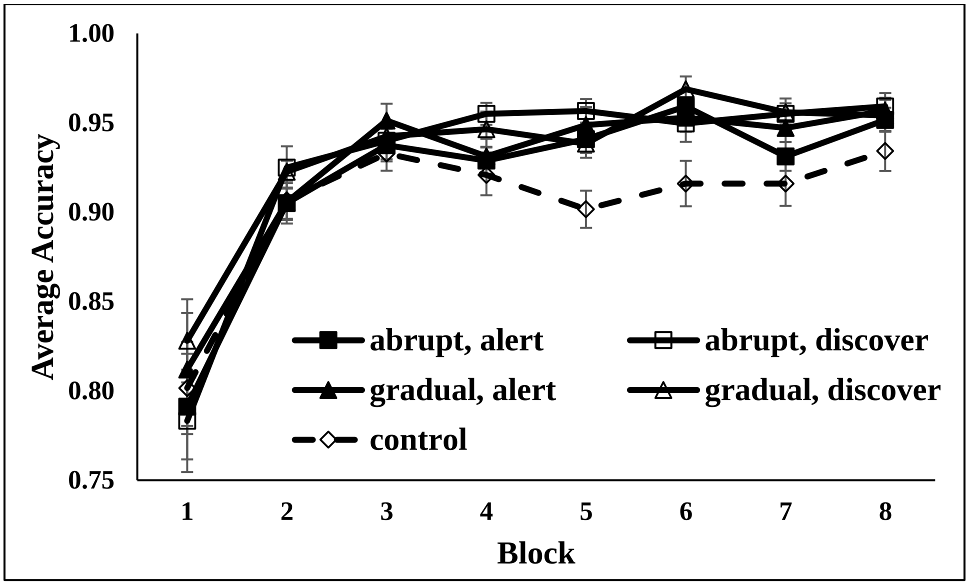

Figure 23. Average accuracy by condition and block in Experiment 3. Error bars represent standard error.

Confidence

An analogous set of ANOVAs was conducted on average confidence ratings. First, there was a main effect of block on pre-change average confidence, $F(1.75, 209.91) = 86.35$, $p < .001$, $\eta_p^2 = .42$, with average confidence increasing across pre-change blocks 1-3. Critically, there was no difference in confidence among conditions prior to the analogical transfer, $F(4, 120) = 0.62$, $p = .650$, $\eta_p^2 = .02$. A similar result was found when comparing average confidence in the change/post-change bocks 4-8. There was a main effect of block, $F(3.51, 420.94) = 13.07$, $p < .001$, $\eta_p^2 = .10$, with average confidence continuing to increase across post-change blocks. There was no main effect of condition on average confidence in post-change blocks, $F(4, 120) = 1.60$, $p = .178$, $\eta_p^2 = .05$, nor was there an interaction of block and condition, $F(14.03, 420.94) = 1.43$, $p = .135$, $\eta_p^2 = .05$. Therefore, despite the shift in RB categories (i.e., analogical transfer), average confidence levels in the experimental conditions are no different than that of the control condition. Average confidence can be seen by block and condition in Figure 24.

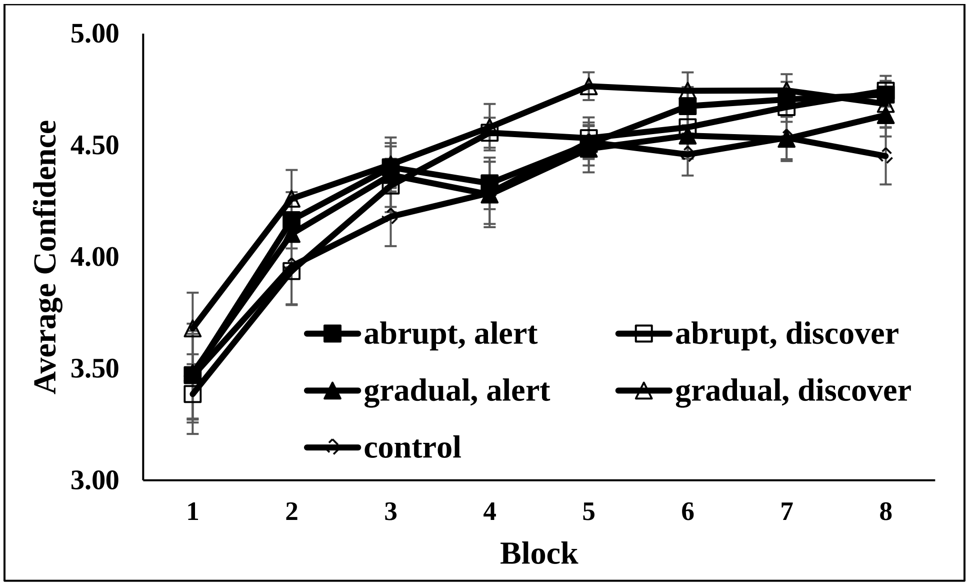

Figure 24. Average confidence by condition and block in Experiment 3. Error bars represent standard error.

Change Effects

Analogical Transfer within Conditions

Two-tailed paired-samples *t*-tests were conducted within each of the four experimental conditions to evaluate evidence of analogical transfer. Average accuracy in the final pre-change block (block 3) was compared to the final block of the experiment (block 8) to assess evidence of analogical transfer (i.e., if participants returned to at least pre-change accuracy levels following the category change). There were no significant differences between average accuracy in blocks 3 and 8 for any of the four experimental conditions (see Table 19), thus providing evidence of analogical transfer in every RB condition. Additional paired-samples *t*-tests were conducted to evaluate significant accuracy costs (block 4 – block 3 in abrupt conditions; block 6 – block 4 in gradual conditions) and recovery (block 8 – block 4 in abrupt conditions; block 8 - block 6 in gradual conditions) within each condition. Cost and recovery are also summarized in Table 19. The terms "cost" and "recovery" are used here in the context of RB (and in Experiment 4, II) analogical transfer, matching the literature (Maddox et al., 2010). First, there was a significant "cost" in the gradual/discover condition, $t(24) = 2.51$, $p = .019$, such that average accuracy actually increased following completion of the gradual change. There was also significant "recovery" in the abrupt/alert condition such that participants in this condition saw no accuracy costs from the change and continued to increase their accuracy across the blocks. As a whole, it is evident that analogical transfer did occur, and transfer was not impacted negatively by differences in rate or awareness of change.

Table 19. Average cost, recovery, and analogical transfer by condition in Experiment 3.

Condition	Measure	Value (acc_2-acc_1)	$t(24)$	p	Cohen's d
Abrupt, Alert	Cost	-0.01	0.85	.406	0.17
	Recovery	0.02	2.31	**.030**	0.46
	Transfer	0.01	1.52	.141	0.30
Abrupt, Discover	Cost	0.02	1.36	.186	0.27
	Recovery	0.00	0.42	.676	0.09
	Transfer	0.02	1.77	.090	0.35
Gradual, Alert	Cost	0.00	0.18	.863	0.04
	Recovery	0.00	0.54	.598	0.11
	Transfer	0.01	0.55	.587	0.11
Gradual, Discover	Cost	0.03	2.51	**.019**	0.50
	Recovery	-0.02	1.72	.098	0.35
	Transfer	0.01	1.199	.248	0.24

Accuracy Cost and Recovery by Rate and Awareness of Change

A set of 2 (rate: abrupt, gradual) x 2 (awareness: alert, discover) between-subjects ANOVAs were conducted on average accuracy cost (block 4 – block 3 for abrupt change; block 6 – block 3 for gradual change) and recovery (block 8 – block 4 for abrupt change; block 8 – block 6 for gradual change) as measures of analogical transfer. If RB analogical transfer does indeed occur, there should be no accuracy cost, nor need for recovery (other than any continual increases in accuracy until asymptotic performance). There was a main effect of awareness of change, $F(1, 96) = 7.64$, $p = .007$, $\eta_p^2 = .07$, such that alerting led to a negligible change cost (i.e., accuracy was maintained), while discover conditions led to small increases in accuracy post-analogical transfer (see Table 19). There was no main effect of rate of change on accuracy cost, $F(1, 96) = 0.68$, $p = .410$, $\eta_p^2 = .01$, nor was there an interaction of rate and awareness of change, $F(1, 96) = 0.09$, $p = .763$, $\eta_p^2 < .01$.

Analysis of accuracy recovery (from first block post-change to final block) revealed main effects of both awareness of change, $F(1, 96) = 4.39$, $p = .039$, $\eta_p^2 = .04$, and rate of change, $F(1, 96) = 4.59$, $p = .035$, $\eta_p^2 = .04$. Alerting led to greater accuracy recovery than discover conditions, and there was a greater recovery in accuracy in abrupt change rate conditions compared to gradual change rate conditions. Given no main effect of change rate on accuracy cost, the main effect of change rate on recovery suggests that accuracy increased beyond pre-change levels in abrupt change conditions. Lastly, there was no interaction of rate and awareness of change on average accuracy recovery, $F(1, 96) < 0.01$, $p = .982$, $\eta_p^2 < .01$.

Confidence Cost and Recovery by Rate and Awareness of Change

An analogous set of ANOVAs were conducted on the immediate and long-term effects of change on average confidence. There were main effects of both rate of change, $F(1, 96) = 7.00$, p

= .010, $\eta_p^2 = .07$, and awareness of change, $F(1, 96) = 7.30$, $p = .008$, $\eta_p^2 = .07$, on the immediate effect of change on average confidence. Average confidence increased more post-change in discover than alert conditions and gradual than abrupt conditions. There was no interaction of rate and awareness of the immediate effect of change on average confidence, $F(1, 96) = 0.06$, $p = .815$, $\eta_p^2 < .01$.

Analysis of the long-term effect of change on average confidence again revealed main effects of both rate, $F(1, 96) = 11.71$, $p < .001$, $\eta_p^2 = .11$, and awareness of change, $F(1, 96) = 6.86$, $p = .010$, $\eta_p^2 = .07$. Average confidence increased to a greater extent in abrupt than gradual change rate conditions, as well as when alerted rather than not alerted to the change. There was no interaction of rate and awareness of change on the long-term effect of change on average confidence, $F(1, 96) = 0.18$, $p = .676$, $\eta_p^2 < .01$.

The analysis of the immediate and long-term effects of change on average confidence suggests differing impacts of RB analogical transfer on confidence levels by change characteristics. Yet, confidence by the end of the experiment is equivalent across conditions (meaning that recovery was possible for conditions in which confidence took an immediate hit). It appears that alerting participants to a potential change in the categories, when the same decision rule would apply with the new categories, resulted in a small drop in confidence despite generally no drop in accuracy. Once participants had enough experience with the new categories, rate and awareness of change had little effect on confidence or accuracy.

Decision Bound Modeling

Data from each individual participant were fit by several linear decision bound models to estimate the category rule that participants were using in each block of the experiment. Decision bound models provide an important second look at analogical transfer, as accuracy alone does not tell the whole story due to the possibility that different types of strategies can result in similar accuracy levels. Four decision bound models were examined: 1) an optimal (opt) model that used fixed values for the slope and intercept of the decision bound, assuming that the participant correctly used the optimal rule, 2) a general linear classifier (glc) model with free parameters for both the slope and the intercept, providing an estimate of when the participant was integrating information from both dimensions in their rule, 3) a model (called the rbx model) that assumed the participant used a selective attention rule focusing on the x dimension (i.e., line length), and 4) a model (called the rby model) that assumed the participant used a selective attention rule focusing on the y dimension (i.e., line orientation). The categories in the present experiment were defined by the y dimension (line orientation). Thus, participants who were best fit by the rby model but not the optimal model learned the correct rule but not the optimal criterion location upon which to classify the two categories on line orientation. Of critical interest is the number of participants best fit by a selective attention rule (i.e., rule-based strategy; opt, rbx, or rby models) vs. a dimensional integration rule (i.e., information-integration strategy; glc model). For this reason, the number of participants best fit by either a selective attention or dimensional integration model type is reported in Table 20.

Table 20. Number of participants whose data were best fit by RB vs. II model types in Experiment 3.

n = 25 per condition		Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
Abrupt, Alert	RB	21	20	24	24	18	23	21	21
	II	4	5	1	1	7	2	4	4
Abrupt, Discover	RB	22	21	19	20	23	23	22	25
	II	3	4	6	5	2	2	3	0
Gradual, Alert	RB	22	21	22	23	22	24	25	23
	II	3	4	3	2	3	1	0	2
Gradual, Discover	RB	23	21	24	23	24	25	24	20
	II	2	4	1	2	1	0	1	5
Control	RB	19	23	23	23	21	25	21	21
	II	6	2	2	2	4	0	4	4
Averages (selective attention rule types only)									
	Abrupt	43	41	43	44	41	46	**43**	46
	Gradual	45 #	42 #	46 #	46 #	46 #	49 #	**49** *	43 #
	Alert	43	41	46	47	**40**	47	46	44
	Discover	45 #	42 #	43 #	43 #	**47** *	48 #	46 #	45 #

Note. RB model types, known as selective attention rules, include opt and rby. No one was best fit by rbx model. II model type, known as a dimensional integration rule, includes the glc model. The bottom portion of the table describes the proportion of participants best fit by a selective attention model. Significance of comparisons across rate and awareness of change conditions is reported below the number of participants best fit for average conditions (# p >= .05; * $p < .05$; ** $p < .01$).

From Table 20, it is clear that the majority of participants across all conditions used a rule-based strategy (of the optimal type; no one used sub-optimal rbx rule) to classify the stimuli. It is particularly interesting to explore how many people switched (or did not switch) from an RB to an II strategy type following the category change. From Table 20 it is clear that there are few differences in selective attention vs. dimensional integration rule usage between change conditions. However, two significant differences were found. First, in block 5 (the second gradual change block or the second block post-abrupt change with the new RB categories) a significantly larger number of participants in discovery conditions were best fit by a selective attention rule compared to participants in alerted conditions, z (2-tailed) = 2.08, p = 038. In addition, in block 7 there was a significantly larger number of participants in gradual than in abrupt change conditions that were best fit by a selective attention type rule, z (2-tailed) = 2.21, p = .027. Nonetheless, for the majority of the blocks, and particularly in the final block, there were no differences in rule usage among change conditions suggesting that analogical transfer (i.e., category change) did occur and was minimally impacted by levels of rate and awareness of change.

Discussion

Experiment 3 provides a systematic investigation of the effect of concept drift characteristics (rate and awareness of change) on rule-based (RB) analogical transfer. Participants learned to categorize lines varying in length and orientation into two categories in equal base-rate and payoff conditions. After block 3, category distributions shifted either abruptly or gradually to new areas of the stimulus space defined by the same optimal decision bound (i.e., a sampling shift). While previous research has found evidence of analogical transfer in explicit (RB) learning (Casale et al., 2012; Helie et al., 2015; Smith et al., 2015), gradual change is a novel consideration. In addition, both Casale et al. (2012) and Smith et al. (2015) alerted participants they would be transferring to new categories, while Helie et al. (2015) did not provide an alert. Forewarning has not been systematically manipulated within a single study. The rate and awareness of change variables in the present experiment were used largely as a comparison to the case of II categories (Experiment 4), where these variables were expected to mediate performance detriments in II analogical transfer. Thus, the three hypotheses regarding the present Experiment predicted positive RB analogical transfer (Hypothesis 1), no effect of awareness of change (Hypothesis 2), and no effect of rate of change (Hypothesis 3).

The results of Experiment 3 are consistent with the first hypothesis. Positive analogical transfer was present in all experimental conditions. There was little-to-no change in accuracy from the final pre-change block (block 3) to the first-post change block (block 4 abrupt; block 6 gradual). GRT modeling results support this conclusion in that there were the same, or more, cases of RB-model fits post-change compared to the last pre-change block. In fact, accuracy was higher (and a larger proportion of participants were best fit by selective attention vs. dimensional integration model) in blocks 4-8 in all experimental conditions compared to the control condition, suggesting an unexpected benefit from the change. There were no differences in confidence among any of the experimental conditions or between experimental and control conditions.

Hypothesis 2 was not supported. An analysis of accuracy cost and recovery revealed that the initial impact (i.e., cost) of RB analogical transfer was an improvement in accuracy in discover conditions not observed in alerted conditions. Likely due to this difference in the initial cost of analogical transfer, there was a significant improvement in accuracy across post-change blocks (i.e., recovery) observed in alerted compared to discovery conditions. Therefore while the immediate impact of analogical transfer was better in discovery than alerted conditions, there were no differences in analogical transfer by the final block. In addition, the immediate impact of analogical transfer was an increase in confidence in discovery conditions but a decrease in confidence in alerted conditions (despite generally high confidence ratings in all conditions).

Even though there was evidence of analogical transfer in each experimental condition, Hypothesis 3 – regarding a predicted lack of difference between change rate conditions – was not supported. Despite there being no significant differences in accuracy costs among experimental conditions, accuracy recovery was significantly greater in abrupt than gradual change rate conditions. Since accuracy was similar in the final block, this suggests that accuracy costs must have been more positive (meaning accuracy increased post-change) in gradual than abrupt conditions (although this was not significant). Thus, the gradual condition exhibited a larger increase in accuracy from the initial cost to the final block. Similarly, immediate changes in confidence were more positive in gradual than abrupt change conditions, while long-term increases in confidence were greater in abrupt than gradual conditions. Model fits from GRT models provide additional support for potential benefits of gradual change in RB analogical transfer, with a significantly greater proportion of gradual change participants fit by optimal RB models compared to participants in abrupt change conditions. Nonetheless, asymptotic performance is equivalent in the final block.

One interesting finding was the difference in accuracy between the control and experimental conditions in the change/post-change blocks (blocks 4 through 8). Given that no part of the task changed in the control condition, accuracy should have at the least maintained at the asymptotic level. The fact that average accuracy in the control condition decreased across blocks 5-8 in particular suggests that participants perhaps learned the rule fairly quickly and got bored with the task in the later stages, resulting in a few more incorrect trials. Perhaps then experimental conditions were more engaged throughout the task given that the average categories shifted and new information was provided in some conditions (i.e., alert conditions). Interestingly, accuracy continued to increase slightly in discover conditions who may never have realized that anything changed. Thus, it may not take awareness of change to maintain engagement through task changes. Future work might further explore the effect of change, and awareness of change, on task engagement.

A second interesting finding was that alerting participants to the upcoming category change resulted in small decreases in both accuracy and confidence in the first block following change. It is possible that alerting can have a negative cost on immediate performance in analogical transfer by presenting participants with an invitation to figure out what has changed. Instructions did not say how the categories would change - only that they would change. Performance may likely have been better in discover conditions in block 4 since they could continue to apply their learned rule successfully without thinking they might need to change their strategy.

Despite these interesting findings, of central importance to the question of concept drift in perceptual category learning is the difference in analogical transfer between RB and II category structures, and the extent to which change characteristics may or may not mediate differences in analogical transfer. Experiment 4 then reports the results of an analogous experiment conducted using II category structures.

CHAPTER 6: EXPERIMENT 4: II ANALOGICAL TRANSFER

Experiment 4 provides a direct comparison for Experiment 3. Experiment 4 is identical to Experiment 3 with the exception of the category structure used. II category structures were used instead of RB to evaluate analogical transfer and the effect of change characteristics within the procedural learning system.

Method

Experimental Design

Four experimental conditions resulted from the factorial combination of the between-subjects independent variables rate of change (abrupt, gradual) and awareness of change (alerted, discovered):

1. II: Abrupt change, alerted in instructions
2. II: Abrupt change, no alerting (i.e., discovered)
3. II: Gradual change, alerted in instructions
4. II: Gradual change, no alerting (i.e., discovered)

Again, a fifth between-subjects condition was also run with the II category distributions remaining stationary over the entire experiment. See for a figure of the experimental design.

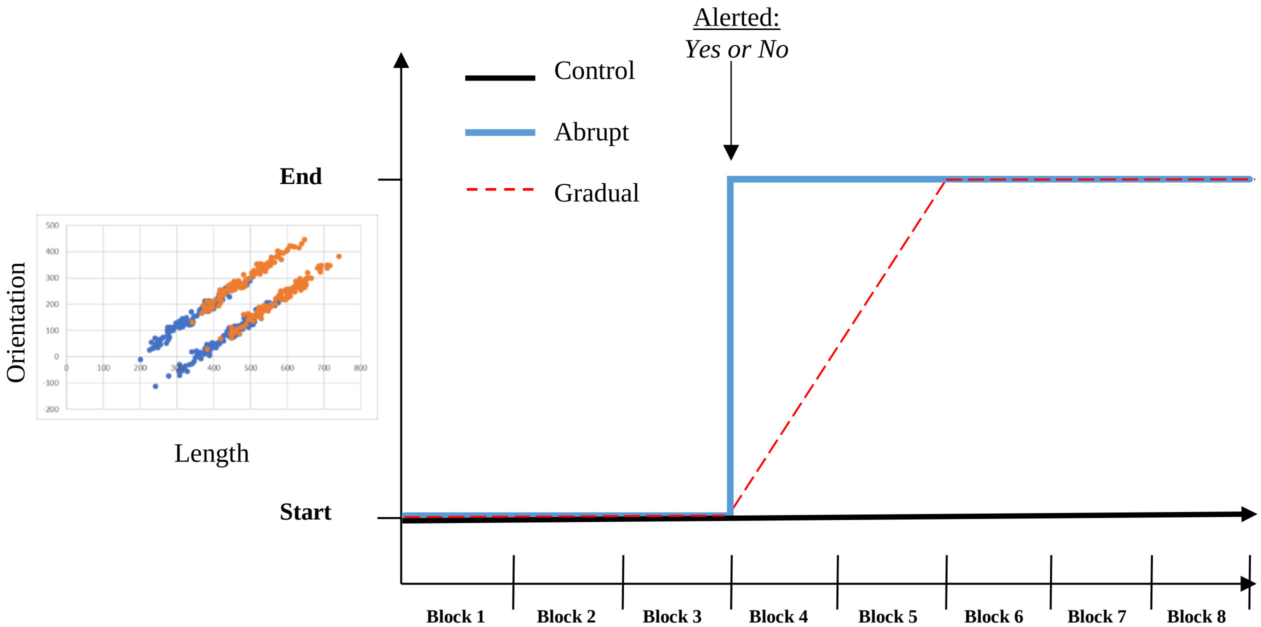

Figure 25. Experimental design for Experiment 4.

Participants

One hundred and twenty-five students from the University of Central Florida and volunteers from the community participated in the experiment. Students who signed up for the study through the SONA systems participation pool received course credit in exchange for their participation. No form of compensation was awarded to volunteers. All participants were at least 18 years of age and reported normal or corrected-to-normal vision (see Table 21). Each participant completed only one condition of the experiment, with 25 participants in each of the 5 conditions. The sample size for Experiment 4 was set to match that of Experiment 3.

Table 21. Participant demographics in Experiment 4.

	Age		Gender Count	
Condition (*n*'s = 25)	*M*	*SD*	Male	Female
Abrupt, Alert	26.80	12.85	12	13
Abrupt, Discover	24.04	7.27	15	10
Gradual, Alert	23.16	9.45	7	18
Gradual, Discover	25.12	11.52	15	10
Control	24.40	11.15	14	11
Overall (*n* = 125)	24.70	10.52	63	62

Stimuli

The stimuli were of the same kind as Experiment 3 (i.e., lines varying in length and orientation) but were created with a different structure. Information integration category structures were created using the Ashby and Gott (1988) "randomization technique", as were the RB stimuli in Experiment 3. Since II category structures require an integration of dimensional information (i.e., both length and orientation are relevant to classification), they are typically harder to learn. Because of this, the path of many other researchers was taken(Ashby, Maddox, & Bohil, 2002; Casale et al., 2012, Exp 2), and the II category *d'* value increased (from *d'* = 6 to *d'* = 13) making categories further separated so as to relatively equate asymptotic pre-change accuracy levels between the RB and II categories (RB_{acc} = .82, II_{acc} = .83, learner average across 400 trials; $p > .05$; based on previous work).

First, the Matlab GRT toolbox by Alfonso-Reese (2006) was used to create an RB stimulus set with a *d'* of 13 (otherwise matching the RB stimulus set in Experiment 3). Then, the II stimuli were then created by shifting the distributions to center on the origin, rotating the horizontal categories counter-clockwise 45 degrees, and shifting these new II categories back to their original stimulus space location (matching the starting stimulus space location in the RB categories). Mid- and post-change category distributions were created by incrementing the category means from the II control data set based on the number of change trials. This process resulted in three data sets: 1) 400 trials of stationary II category distributions (control), 2) 150 trials of II category distributions followed by immediate change to 250 trials of new (ending location) II category distributions, and 3) 150 identical pre-change trials, 100 change trials (same stimuli as control dataset but with category means incremented appropriately), and 150 post-change trials identical to the other two datasets. Stimuli were presented at a 1:1 base-rate with no payoffs provided. II category statistics can be seen in Table 22.

Table 22. II category statistics for Experiment 4.

Condition	Category	μ_x	μ_y	σ_x	σ_y	Cov_{xy}
II Start	A	349.80	157.53	71	71	5041
	B	432.53	74.80	71	71	5041
II End	A	491.22	298.95	71	71	5041
	B	573.95	216.22	71	71	5041

Procedure

The procedure for Experiment 4 was identical to that of Experiment 3.

Predictions

COVIS theory predicts sustained performance in RB conditions but deteriorated performance in II conditions (since these conditions rely on stimulus-response associations). Indeed, previous research has shown good performance in RB conditions in change of this type but generally poor II performance (Casale et al., 2012; Helie et al., 2015; Smith et al., 2015). However, one study (Seger et al., 2015) found positive II transfer when category distributions shifted to flanking distributions with the same category means (see Figure 3). The results of Seger et al. (2015) suggest that under the right conditions, people might be able to handle concept drift in information integration category structure conditions. In particular, since II category learning is mediated by implicit learning, certain characteristics of change (i.e., gradual change) may improve analogical transer. The following hypotheses were developed regarding II analogical transfer in Experiment 4:

Hypothesis 1: Based on previous research, analogical transfer will occur with II categories (i.e., full recovery to pre-change accuracy levels), but there will be an initial performance decrement (i.e., accuracy cost) due to the shift in category distributions. This initial performance decrement is linked to the need to relearn stimulus-response associations in implicit learning.

Hypothesis 2: There will be no effect of awareness of change (alerted, discovery) on II analogical transfer since II categories are learned implicitly.

Hypothesis 3: II analogical transfer will be greater in response to gradual than abrupt change, as it provides the slower implicit learning system a chance to gradually relearn the stimulus-response associations.

Results

Comparison to Control

Accuracy

A 3 (pre-change blocks 1-3) x 5 (condition; including control) mixed factor ANOVA was conducted on average accuracy in pre-change blocks. There was a main effect of block, $F(1.80, 216.36) = 99.60$, $p < .001$, $\eta_p^2 = .45$, with average accuracy increasing across pre-change blocks. There was no main effect of condition, $F(4, 120) = 0.87$, $p = .487$, $\eta_p^2 = .03$, or interaction of condition and block, $F(7.21, 216.36) = 0.67$, $p = .699$, $\eta_p^2 = .02$. Average accuracy by condition across all blocks can be seen in Figure 26.

A subsequent 5 (change/post-change blocks 4-8) x 5 (condition; including control) mixed factor ANOVA was conducted on average accuracy in blocks 4-8 to assess the impact of II analogical transfer on accuracy in the experimental conditions compared to the control condition. There was a main effect of block, $F(2.97, 356.23) = 9.94$, $p < .001$, $\eta_p^2 = .08$, such that average accuracy continued to increase across change/post-change blocks. There was also a main effect of condition, $F(4, 120) = 3.57$, $p = .009$, $\eta_p^2 = .11$, with lower average accuracy in the abrupt/discover condition compared to both the control condition ($p = .022$) and the gradual/alert condition ($p = .046$). The interaction of block and condition was also significant, $F(11.87, 356.23) = 4.59$, $p < .001$, $\eta_p^2 = .13$, such that there are differences among conditions in blocks 4 and 5 (e.g., abrupt/discover lower than other conditions), but there are no differences in blocks 6-8.

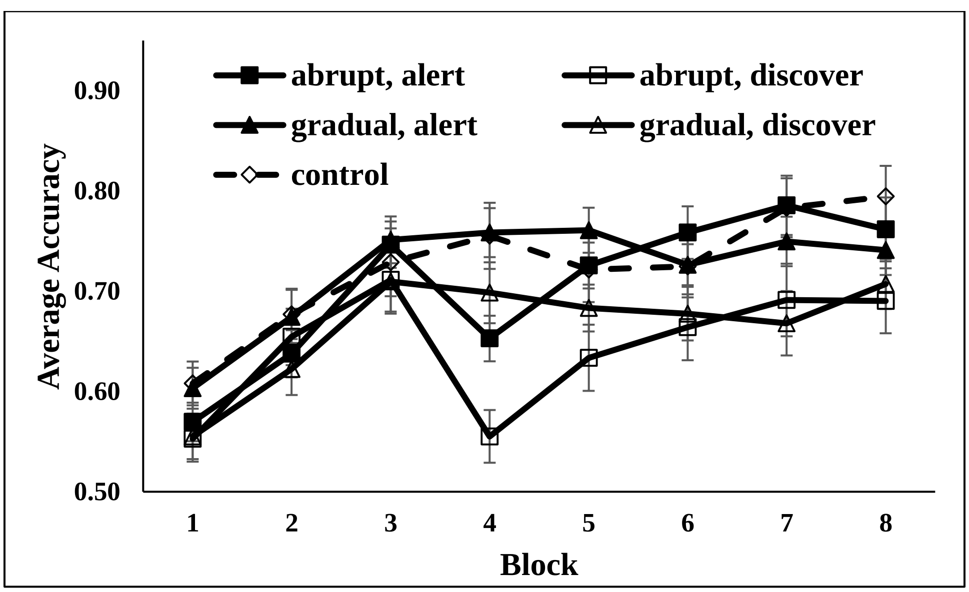

Figure 26. Average accuracy by block and condition in Experiment 4. Error bars represent standard error.

Confidence

Average confidence increased across pre-change blocks (1-3), $F(1.75, 210.42) = 29.49$, $p < .001$, $\eta_p^2 = .20$. There was no main effect of condition on average confidence in pre-change bocks, $F(4, 120) = 0.67$, $p = .614$, $\eta_p^2 = .02$, or interaction of block and condition, $F(7.01, 210.42) = 0.47$, $p = .857$, $\eta_p^2 = .02$.

Additionally, average confidence tended to continue to increase across change/post-change blocks (4-8), $F(2.89, 11.57) = 7.58$, $p < .001$, $\eta_p^2 = .06$. There was no main effect of condition on average confidence in change/post-change blocks, $F(4, 120) = 2.00$, $p = .099$, $\eta_p^2 = .06$, or interaction between block and condition, $F(11.57, 346.98) = 0.83$, $p = .615$, $\eta_p^2 = .03$. Confidence was lowest in the abrupt/discover condition, with similar average confidence in the remaining conditions. Average confidence by block and condition in Experiment 4 can be seen in Figure 27.

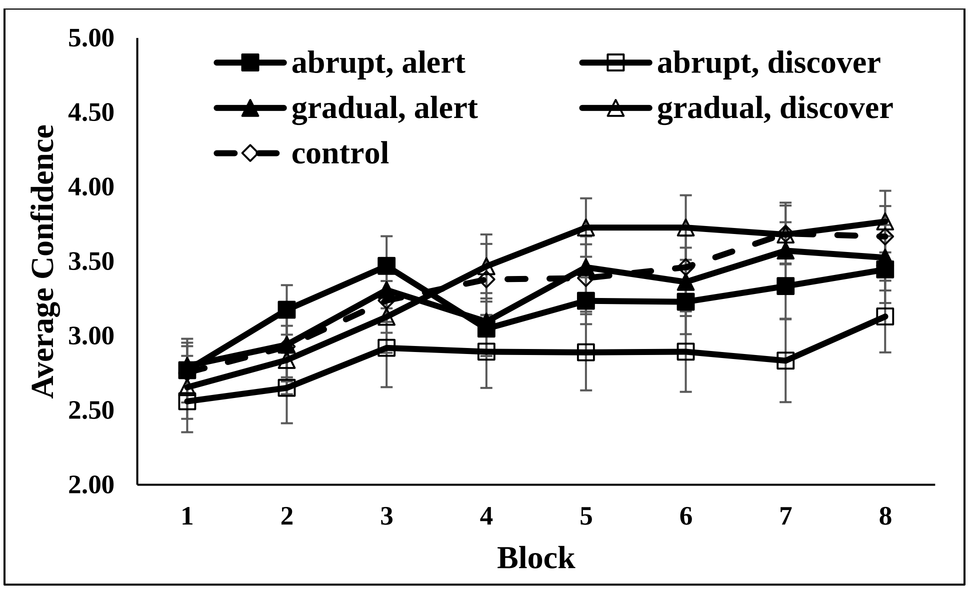

Figure 27. Average confidence by block and condition in Experiment 4. Error bars represent standard error.

Change Effects

Analogical Transfer within Conditions

Paired samples *t*-tests were conducted within each of the experimental conditions to evaluate evidence of analogical transfer. Average accuracy in the final pre-change block (block 3) was compared to the final block of the experiment (block 8). There was no significant difference between block 3 and block 8 for abrupt/alert, $t(24) = 0.58$, $p = .566$, abrupt/discover, $t(24) = 0.71$, $p = .483$, gradual/alert, $t(24) = 0.44$, $p = .661$, or gradual/discover, $t(24) = 0.09$, $p = .933$, thus providing evidence of analogical transfer in every experimental condition. Paired samples *t*-tests were also conducted to evaluate significant differences in accuracy cost (block 4 – block 3 in abrupt conditions; block 6 – block 4 in gradual conditions) and recovery (block 8 – block 4 in abrupt conditions; block 9 - block 6 in gradual conditions) within each condition. Since optimal accuracy is the same pre- and post-change (100%), raw values are used. Average cost and recovery by condition are summarized in Table 23. There were significant effects of change on accuracy cost and recovery in the abrupt change conditions but not in the gradual change conditions. This suggests that abrupt change is associated with an immediate accuracy cos, but observers are able to recover significantly over the next several blocks.

Table 23. Average accuracy cost, recovery, and analogical transfer by condition in Experiment 4.

Condition	Measure	Value (acc_2-acc_1)	*t*(24)	*p*	Cohen's *d*
Abrupt, Alert	Cost	-0.09	3.62	**.001**	0.72
	Recovery	0.11	3.84	**< .001**	0.77
	Transfer	0.02	0.58	.566	0.12
Abrupt, Discover	Cost	-0.16	4.91	**< .001**	0.98
	Recovery	0.14	4.35	**< .001**	0.87
	Transfer	-0.02	0.71	.483	0.14
Gradual, Alert	Cost	-0.03	1.50	.146	0.30
	Recovery	0.01	0.79	.440	0.16
	Transfer	-0.01	0.44	.661	0.09
Gradual, Discover	Cost	-0.03	1.26	.220	0.25
	Recovery	0.03	1.54	.137	0.31
	Transfer	0.00	0.09	.933	0.02

Accuracy Cost and Recovery by Rate and Awareness of Change

A set of 2 (rate) x 2 (awareness) between-subjects ANOVAs was conducted on the average accuracy cost and recovery to evaluate any differences with respect to change characteristics. There was a main effect of rate of change on average accuracy cost, $F(1, 96) = 14.32$, $p < .001$, $\eta_p^2 = .13$, such that there was a larger decrease in accuracy immediately following change in abrupt compared to gradual change conditions. There was no main effect of awareness of change on average accuracy cost, $F(1, 96) = 1.87$, $p = .175$, $\eta_p^2 = .02$, nor was there an interaction between rate and awareness of change, $F(1, 96) = 1.17$, $p = .281$, $\eta_p^2 = .01$.

Table 24. Mean (SD) accuracy and confidence cost and recovery by condition in Experiment 4.

	Accuracy		Confidence	
	Cost	Recovery	Cost	Recovery
Abrupt, Alert	-0.09 (0.13)	0.11 (0.14)	-0.42 (0.70)	0.51 (0.64)
Abrupt, Discover	-0.16 (0.16)	0.14 (0.016)	0.11 (0.43)	0.24 (0.61)
Gradual, Alert	-0.03 (0.08)	0.01 (0.09)	0.05 (0.63)	0.16 (0.85)
Gradual, Discover	-0.03 (0.13)	0.03 (0.10)	0.60 (1.17)	0.13 (0.46)

Analysis of average accuracy recovery also revealed a main effect of rate of change, $F(1, 96) = 16.16$, $p < .001$, $\eta_p^2 = .14$, such that accuracy recovered to a greater extent in abrupt than gradual change rate conditions. There was no main effect of awareness of change on average accuracy recovery, $F(1, 96) = 0.70$, $p = .405$, $\eta_p^2 = .01$, nor was there an interaction of rate and awareness of change, $F(1, 96) = 0.05$, $p = .822$, $\eta_p^2 < .01$.

Average accuracy recovery was considered in a second way to understand how much recovery is possible despite normal fluctuations in performance across blocks (e.g., often observer's highest accuracy does not fall in the final block of the experiment). Accordingly, recovery was recalculated as the difference between the maximum accuracy level obtained in pre-change blocks (1-3) compared to maximum accuracy obtained in post-change blocks (6-8). An analogous ANOVA was conducted that revealed a nearly significant main effect of awareness of change, $F(1, 96) = 3.80$, $p = .054$, $\eta_p^2 = .04$, but no main effect of rate of change, $F(1, 96) = 0.32$, $p = .575$, $\eta_p^2 < .01$, or interaction of block and condition, $F(1, 96) = 2.60$, $p =$

.110, $\eta_p^2 = .03$. Analyzed in this way, average accuracy recovery was found to be higher when alerted to change as opposed to when not (particularly when change was abrupt).

Finally, a 3 (post-change blocks 6-8) x 2 (rate) x 2 (awareness) mixed factor ANOVA was conducted on average accuracy in post-change blocks. There was a main effect of awareness of change, $F(1, 96) = 7.10$, $p = .009$, $\eta_p^2 = .07$, such that participants who were alerted to the category change had higher average accuracy in post-change blocks than participants who were not alerted. No other main or interactive effects were significant (all *p*'s > .121).

Confidence Cost and Recovery by Rate and Awareness of Change

There was a main effect of rate of change on the immediate effect of change on average confidence, $F(1, 96) = 14.36$, $p < .001$, $\eta_p^2 = .13$, with a larger initial decrease in confidence post-change in abrupt compared to gradual change rate conditions. There was also a main effect of awareness of change on the immediate effect of change on average confidence, $F(1, 96) = 7.64$, $p = .007$, $\eta_p^2 = .07$, with larger initial decreases in confidence post-change when alerted to change compared to when not alerted. There was no interaction of rate and awareness of change, $F(1, 96) = 0.55$, $p = .462$, $\eta_p^2 = .01$. In addition, there were no main or interactive effects of rate and awareness of change on the long-term effects of change on average confidence (all *p*'s > .086). The immediate and long-term effects of change on average confidence can be seen in Table 24.

Finally, a 3 (post-change blocks 6-8) x 2 (rate) x 2 (awareness) mixed factor ANOVA was conducted to evaluate the role of rate and awareness of change on confidence levels post-category change. There was a main effect of block on average confidence, $F(1.95, 187.47) = 3.89$, $p = .023$, $\eta_p^2 = .04$, such that confidence tended to increase across post-change blocks, with significantly higher accuracy in block 8 compared to block 6 ($p = .036$). There was also a main effect of rate of change on average confidence in post-change blocks, $F(1, 96) = 4.55$, $p = .035$, $\eta_p^2 = .05$, with higher average confidence in gradual compared to abrupt change conditions. No other main or interactive effects were significant.

Decision Bound Modeling

The same decision bound models from Experiment 3 were evaluated here, with the addition of a free-boundary model that assumed the participant used the optimal slope of the decision bound but left the intercept value free to vary (giving the model two free parameters: intercept and noise). The free-boundary model was not explored in Experiment 3 because it was identical to the rby model in that case. Here, the free-boundary model provides an estimate of when a participant uses the optimal rule (dimensional integration) but sub-optimal criterion value. Participants were again compared in terms of whether they were best fit by a selective attention (i.e., RB) or a dimensional integration (i.e., II) model type in each block. Dimensional integration models in the present experiment are encompassed in the optimal (opt), free-boundary (frb), and general linear classifier (glc) models, while selective attention models include the rbx and rby selective attention models. The number of participants best fit by models of either type across the eight blocks of the experiment can be seen in Table 25.

Table 25. Number of participants best fit by RB vs. II model types by condition and block in Experiment 4.

n = 25 per condition		Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
Abrupt, Alert	RB	8	9	6	10	5	7	5	7
	II	17	16	19	15	20	18	20	18
Abrupt, Discover	RB	8	5	2	6	6	7	7	7
	II	17	20	23	19	19	18	18	18
Gradual, Alert	RB	3	5	1	2	6	5	6	4
	II	22	20	24	23	19	20	19	21
Gradual, Discover	RB	10	10	3	4	6	6	9	3
	II	15	15	22	21	19	19	16	22
Control	RB	4	6	3	2	2	5	2	2
	II	21	19	22	23	23	20	23	23
Average (dimensional integration [II] model types only)									
	Abrupt	34	36	42	**34**	39	36	38	**36**
	Gradual	37	35	46	**44**	38	39	35	**43**
		#	#	#	**	#	#	#	*
	Alert	39	36	43	38	39	38	39	39
	Discover	32	35	45	40	38	37	34	40
		#	#	#	#	#	#	#	#

Note. RB model types include rbx and rby models. II model types include opt, frb, and glc models. The bottom portion of the table describes the proportion of participants best fit by a dimensional integration model. Significance of comparisons across rate and awareness of change conditions is reported below number of participants best fit for average conditions (# p >= .05; * $p < .05$; ** $p < .01$).

Table 25 reveals there was generally a drop in the number of participants in each condition best fit by a dimensional integration model starting in block four, but in most cases showed evidence of "recovery" towards block 8. Participants in the abrupt/alert condition in block 4 were best fit more often by a sub-optimal selective attention model than the control condition, and this difference was significant, z (1-tailed) = 2.65, p = .004. In block 7, the gradual/discover condition was best fit more often by a selective attention model than the control group, z (1-tailed) = 2.39, $p = .008$. Finally, in block 8, a greater proportion of both abrupt conditions were best fit by selective attention models compared to the control condition, z (1-tailed) = 1.84, $p = .033$. Averaging across rate and awareness of change conditions (the bottom half of Table 25) revealed a significant cost in analogical transfer in block four in abrupt compared to gradual conditions, z (1-tailed) = 2.41, $p = .008$, suggesting that abrupt conditions switched to a selective attention rule more often immediately following the analogical transfer. The same pattern was found in the final block where a greater proportion of participants in abrupt conditions were best fit by a sub-optimal selective attention rule compared to gradual change conditions, z (1-tailed) = 1.72, $p = .043$, providing evidence of improved analogical transfer with gradual change.

Comparing RB and II Analogical Transfer

To evaluate the differential effect of analogical transfer in RB and II category structures, a series of 2 (category structure: RB, II) x 2 (rate: abrupt, gradual) x 2 (awareness: alert, discover) between-subjects ANOVAs was conducted on the immediate and long-term effects of change on average accuracy and confidence. Of central importance is the immediate effect of analogical transfer on accuracy within each category structure.

Accuracy

First, the effect of category structure, rate of change, and awareness of change were evaluated with respect to average accuracy costs. There was a main effect of category structure on average accuracy cost, $F(1, 192) = 37.50$, $p < .001$, $\eta_p^2 = .16$, with a greater decrease in average accuracy immediately post-change in II compared to RB category structures. Second, there was a main effect of rate of change, $F(1, 192) = 14.64$, $p < .001$, $\eta_p^2 = .07$, with a greater initial decrease in average accuracy in abrupt compared to gradual change rates. Third, there was an interaction between condition and rate of change, $F(1, 192) = 10.41$, $p = .001$, $\eta_p^2 = .05$, such that abrupt change led to larger initial decreases in accuracy post-change than gradual change within II conditions (no difference within RB category structures). Fourth, the interaction of category structure and awareness of change was significant, $F(1, 192) = 5.18$, $p = .024$, $\eta_p^2 = .03$, such that alerting had a greater impact (i.e., reduced decreases in accuracy relative to discover conditions) in II compared to RB category structures.

Next, the effect of category structure, rate of change, and awareness of change were evaluated with respect to average accuracy recovery. There was a main effect of category structure on average accuracy recovery, $F(1, 192) = 26.43$, $p < .001$, $\eta_p^2 = .12$, with greater recovery in II than RB category structures. There was a main effect of rate of change average accuracy recovery, $F(1, 192) = 20.30$, $p < .001$, $\eta_p^2 = .10$, with greater recovery in abrupt compared to gradual change rates. There was also an interaction between category structure and rate of change, $F(1, 192) = 9.40$, $p = .002$, $\eta_p^2 = .05$, such that the improvement of recovery in abrupt relative to gradual change was greater in II than RB category structures. Finally, when analyzing recovery in terms of difference between maximum accuracy pre and post-change, there was a significant interaction between category structure and awareness of change, $F(1, 192) = 4.89$, $p = .028$, $\eta_p^2 = .03$. Awareness of change made no difference in RB conditions but improved average accuracy recovery in II conditions.

<u>Confidence</u>

Confidence analyses revealed a main effect of rate of change, $F(1, 192) = 21.13$, $p < .001$, $\eta_p^2 = .10$, and awareness of change, $F(1, 192) = 13.90$, $p < .001$, $\eta_p^2 = .07$, on the immediate effect of change on average confidence. Confidence increased post-change to a greater extent in discover than alert conditions, and gradual compared to abrupt change rates. There was also an interaction between category structure and rate of change, $F(1, 192) = 4.12$, $p = .044$, $\eta_p^2 = .02$, such that discover conditions were associated with greater increases in confidence post-change compared to alert conditions to a greater extent in II than RB category structure conditions.

Lastly, there were main effects of both rate of change, $F(1, 192) = 9.76$, $p = .002$, $\eta_p^2 = .05$, and awareness of change, $F(1, 192) = 5.09$, $p = .025$, η_p^2 - .03, on the long-term effects of average confidence. Average confidence increased across post-change blocks to a greater extent in alert than discover conditions, as well as in abrupt compared to gradual change rate conditions.

Summary of RB vs. II Analogical Transfer

As predicted, analogical transfer was evident in both RB and II categories. In addition, there was a greater initial accuracy cost (sometimes associated with a switch to a sub-optimal rule type) in the block immediately following the transfer (again, in line with predictions). Because of the greater initial cost of analogical transfer in II categories, there was also greater recovery. The largest accuracy costs were associated with abrupt change, which affected II category learning more than RB category learning. Finally, alerting an individual to change improved performance in II conditions but made no difference in RB conditions; confidence was higher when alerted and when change occurred gradually. This entire pattern of results follows what was expected based on previous research and the concept drift literature. These results have implications for the transfer of implicit vs. explicit category learning knowledge.

Discussion

II Analogical Transfer

Experiment 4 provides a direct counter-part to Experiment 3. Experiment 3 investigated the effect of concept drift characteristics within RB analogical transfer, while Experiment 4 replicated this experiment using information-integration (II) categories. Participants again learned to categorize lines varying in length and orientation, with a category sampling shift (i.e., analogical transfer) occurring after block 3. As in the previous experiments, three hypotheses were explored: 1) analogical transfer would occur in II learning (i.e., full recovery) after an initial performance decrement (i.e., accuracy cost) due to the shift in the category distributions, 2) awareness of change (i.e., alerting) would not improve II analogical transfer, and 3) analogical transfer would be greater in gradual change conditions than abrupt change conditions.

Analogical transfer did occur, in support of Hypothesis 1. Visual inspection of Figure 26 reveals that average accuracy returned to at least pre-change accuracy levels. Two other tests were conducted to verify the occurrence of analogical transfer. First, a paired-samples *t*-test was conducted comparing average accuracy in the final pre-change block (block 3) to the final post-change block (block 8). There was no significant difference between average accuracy in blocks 3 and 8, $t(99) = 0.35$, $p = .729$. Additionally, analogical transfer was assessed by comparing maximum accuracy in pre-change blocks (blocks -1-3) to maximum accuracy in post-change blocks (blocks 6-8). A paired-samples *t*-test found a significant difference between pre- and post-maximum accuracy, $t(99) = 2.50$, $p = -.014$, cohen's $d = .25$, with higher maximum accuracy in post-change blocks than pre-change blocks. Thus, not only did analogical transfer occur in II category learning, but the analogical transfer categories were learned even better than the original categories.

The second part of Hypothesis 1 predicted that there would be an initial performance decrement due to the category change. This was also supported, as there was an initial accuracy cost. However, it was only significant in abrupt change conditions. This provides support for Hypothesis 3, suggesting that gradual change is easier to adapt to in II categories than abrupt change. Decision bound modeling results corroborate this by highlighting a higher number of participants using an optimal rule types in gradual compared to abrupt change conditions. Finally, Hypothesis 2 was not supported. Contrary to expectations, alerting participants to change in II categories resulted in significantly higher accuracy in post-change blocks. This finding is particularly interesting, as II categories involve a typically non-verbalizable rule, and yet instructions providing a general alert to change (without specifying the manner in which categories would change) improved performance.

CHAPTER 7: GENERAL DISCUSSION

The impact of change on human perceptual category learning was explored through a series of four experiments. Changes occurred mid-experiment to either category base-rates (Exp 1), category payoffs (Exp 2), RB categories (Exp 3), or II categories (Exp 4). Each type of change reflected a component of "virtual concept drift" prevalent in machine learning and data sciences. Concept drift occurs when data distributions change over time, causing the performance of learning models to deteriorate. In the present case, data distributions were represented as normally distributed category distributions in human perceptual category learning tasks.

A thorough review of concept drift in machine learning resulted in the identification of two prominent components of the problem of concept drift. First is the rate at which a data distribution parameter changes over time, ranging from abrupt (instantaneous) change to more gradual change extended over a period of time. Second is a differentiation of approaches taken to handle concept drift. Learning models employing active approaches involve two phases: detection and adaptation. The model must first decide that something in the decision environment has changed if the model is to adjust its parameters to relearn the data distributions. If no change is identified, the same model (trained on previous data) is applied to new inputs. Passive learning approaches, on the other hand, continually adapt to incoming data, with each new input causing the learning model to update its parameters. These differences in approaches lend themselves better to different types of change; active approaches are better suited to handle abrupt concept drift, while passive approaches are better suited to handle gradual concept drift. The idea of active versus passive approaches was employed in the present work by providing an instructional alert to some participants as to an upcoming change (paralleling an active approach that is first made aware of change and then adapts). Other participants received no alert and needed to simply adapt to the change on their own (whether or not they explicitly detected change, paralleling a more passive approach). These two variables – rate (abrupt, gradual) and awareness (alert, discover) of change – were investigated in each of the four experiments.

The selection of parameters of change in each experiment was made to provide insight into the relative impact of change on explicit vs. implicit learning. Experiment 1 (base-rates) and Experiment 2 (payoffs) provide a comparison of two decisional components of categorization that are believed to be learned largely through opposite learning systems. Recent evidence has supported earlier hypotheses (Koehler, 1996; Spellman, 1996) suggesting base-rate sensitivity is mediated largely by the implicit learning system (Bohil & Wismer, 2015; Wismer & Bohil, 2017). Payoffs (i.e., costs and benefits), on the other hand, are believed be learned and reasoned about explicitly. Base-rates and payoffs have a long history of research in category learning and provide an initial venue for exploration of the effect of change in explicit vs. implicit learning.

Additionally, two different category structures studied in the COVIS literature provide an additional set of tasks in which to evaluate the effect of change on performance. Rule-based (RB) category structures are learned via the explicit learning system, mediated by the prefrontal cortex, through conscious hypothesis testing relying on working memory. RB categories typically are defined by a verbalizable rule. Information-integration (II) category structures are learned via the implicit learning system, are mediated by subcortical structures such as the basal ganglia and dopamine reward system, and learn stimulus-response associations through a procedural learning process. II categories are typically defined by a non-verbalizable, or difficult to verbalize, rule. These two category structures - the focus of a wealth of research in the study of multiple systems of category learning - provide an additional layer of investigation into the effect of change within explicit vs. implicit learning.

Experiment 1 Summary

Experiment 1 tested three hypotheses. First, it was hypothesized that people would be sensitive to category base-rates but would display conservatism in the decision criteria used. Second, it was hypothesized that awareness of change would have no effect on change adaptation since base-rates are believed to be learned implicitly. Third, it was hypothesized that gradually changing base-rates may be better incorporated into classification than base-rates that change abruptly. Hypothesis 1 and 2 were supported. People were sensitive to category base-rates but displayed conservative cutoff placement in their decision criteria. Awareness of change also had no effect on change adaptation. Hypothesis 3 was rejected, however. Gradually changing base-rates did not lead to development of an earlier response bias during change or improved criterion placement post-change. Still, evidence of base-rate sensitivity in the gradual/discover condition, coupled with the lowest confidence of all conditions, suggests that implicit learning did occur despite lack of explicit awareness.

The results of Experiment 1 have relevance for various training applications, such as medical training. Oftentimes learners are presented relatively equal numbers of normal and non-normal cases in training, or in the least, consideration of the impact of the distribution of normal vs. non-normal cases in training on future performance is not considered. Training equally on positive and negative cases can actually promote the adoption of an unbiased decision criterion, posing potential issues when practitioners later work with patient populations with different, and possibly quite extreme, base-rates. As noted previously, this is of particular relevance in applications of ambiguous classification, such as radiology or epidemiology, where base-rates should be more vital in the decision making process. Given that simply alerting observers to the base-rate change did not promote more immediately optimal decision criteria placement or increased classification accuracy in the present experiment, additional mitigations may be needed when patient populations, symptom profiles, and/or associated base-rates change. Future work might consider the possible use of simulations in medical (or other) training as a way to provide direct experience with relevant cases matched to population base-rates of relevance in an effort to promote use of optimal decision criteria.

Experiment 2 Summary

Experiment 2 explored additional hypotheses. First, it was hypothesized that people would be sensitive to changes in payoff values, but that decision criteria would be more conservative than analogous base-rate conditions. Second, it was hypothesized that providing an instructional alert about the payoff change would increase criterion shifting to the new payoff values. Third, it was hypothesized that abrupt payoff changes would be easier to adapt to than gradually changing payoffs. Both hypothesis 2 and 3 were based on the assumed role of explicit learning in payoff sensitivity.

Hypothesis 1 was partially supported. Observers were more conservative than observers in corresponding base-rate conditions; however, they exhibited almost no sensitivity to the new payoff values at all in terms of decision criterion placement. There was a slight increase in the frequency of responding with the high value category, as well as bias suggested by GRT model estimates, that suggested some sensitivity to payoffs despite not being reflected in SDT β-values. Hypotheses 2 and 3 also received partial support. Some differences were seen suggesting improved performance in alerted vs. discovery conditions and abrupt vs. gradual change conditions. Both alerting and abrupt change led to larger GRT-β estimates post-change. Alerting improved the long-term effect of change in some conditions. Abrupt change was associated with more immediate increases in point totals (i.e., reward).

Unfortunately, conclusions drawn from Experiment 2 may be limited due to the uncharacteristically low adaptation to the payoff manipulation. Besides the fact that payoff values changed mid-experiment (which is novel in category learning), a few other details differed from previous reports of category payoff learning and may explain the present lack of sensitivity. First, the category payoff values used in the present work were larger than in previous studies. In the past, a 3:1 payoff ratio has typically been introduced by awarding 3 points for one correct response and 1 point for another. In the present work, a 150:50 point ratio was employed. This ratio was selected so that gradually changing payoff values could increment in integer values, starting from a ratio of 100:100. Second, stimuli were randomly selected on each trial determined via category means, variances, and a base-rate parameter (1:1 ratio). Previous studies of category payoff learning have typically preselected stimulus values to ensure perfect 1:1 base-rate ratios within any block of trials. It is possible that the added variability in the present work, coupled with the low category discriminability (d' = 1), resulted in poor payoff sensitivity. Despite the fact that situations of high ambiguity are cases in which payoff values become important, it is obvious that payoffs were not relied on as an alternative to stimulus information. A person who would have chosen to maximize payoffs by responding only category A (the high payoff category) would have received an average of 3750 points in each post-change block – a point total that is higher than any group average across the experiment. Future work might consider using alternative *d'* levels to see if payoff adaptation is dependent on higher levels of discrimination. Third, previous category learning payoff studies usually included two additional lines of payoff feedback onscreen related to the potential amount of points for the completed trial (Bohil & Maddox, 2003; Maddox & Bohil, 2005). The first line lists the number of points that were available to earn ("potential points") while the second line lists the possible total number of potential points earned for the block of trials to that point in time ("potential point total"). However, these lines of feedback were usually presented in studies investigating objective vs. optimal classifier feedback, where possible point totals are linked to either objective accuracies or ideal observer responses. Thus, these lines were chosen to be removed from Experiment 2. Still, it is possible that these lines increased motivation in prior studies, providing a reason for the lower performance witnessed in the present Experiment 2. Finally, no form of compensation was awarded for point maximization (unlike most previous studies of payoff learning), suggesting that participants may have focused on accuracy instead of maximizing reward in Experiment 2.

Experiment 3 Summary

Experiment 3 explored an additional three hypotheses. First, it was predicted that RB analogical transfer would occur, without any initial performance cost to the category change. Second, it was predicted that because positive analogical transfer would occur, awareness of change would not impact performance. Third, for the same reason it was predicted that rate of change would not impact transfer performance.

Hypothesis 1 was supported. Analogical transfer occurred in every experimental condition, regardless of the change characteristics. In fact, change conditions showed increases in accuracy relative to a control condition. This suggests an interesting possibility that something as simple as a sampling shift can increase task engagement in an extended, repetitive task. It is possible that participants in the control condition learned the category rule relatively quickly and became more disengaged later in the experiment, accounting for a small drop in accuracy relative to the analogical transfer conditions.

Hypothesis 2 was rejected. Awareness of change did have some subtle effects on performance in RB analogical transfer. Discovery conditions were associated with greater increases in accuracy immediate following the change as well as increases in confidence. Still, alerted conditions displayed an increase in accuracy to match that of discovery conditions by the final block. This pattern of results suggests that instructionally alerting a person to a category change that does not require a new category rule may actually impede performance improvements initially. This may be due to observers' changing their strategy in anticipation of some larger change in the rule structure. Future studies that evaluate rule structures changes in category learning may find a difference patterns of results with respect to the effect of alerts.

Hypothesis 3 was also rejected; benefits of gradually changing categories were found. In particular, discovery conditions were associated with larger increases in accuracy across the experiment and more frequent adoption of optimal rule types after the change than alerted conditions. Despite these differences in response to change, asymptotic performance in the final block was similar for all conditions. Again, this suggests that change characteristics may have more subtle impacts on the initial effect of analogical transfer in explicit learning that are not long-lasting.

Experiment 4 Summary

Experiment 4 provided a replication of Experiment 3 with information-integration (II) categories, thus providing an exploration of change effects in implicit learning. Three additional hypotheses were explored. First, it was predicted that II analogical transfer would occur, but that there would first be initial performance decrement in response to the change. Second, it was predicted that awareness of change would have no impact on learning in an implicit learning task. Third, it was predicted that performance would be greater in gradual change compared to abrupt change rate conditions. In other words, gradual change might alleviate the predicted cost of analogical transfer and/or lead to increased accuracy recovery across blocks.

Results provided support for Hypothesis 1. There was evidence of analogical transfer in every condition (i.e., final block accuracy returned to pre-change accuracy levels or higher). In fact, the transfer categories were learned to a higher level of accuracy then the initial categories. Additionally, there was a drop in accuracy immediately following analogical transfer. This drop in accuracy was isolated to abrupt change conditions, thus supporting both Hypothesis 1 and Hypothesis 3.

Hypothesis 2 was rejected. Alerting participants to the change in categories resulted in higher accuracy post-change compared to discovery conditions as well as greater accuracy recovery from the cost of analogical transfer than in discovery conditions. In fact, accuracy in discovery conditions was lower than the control condition post-change. The highest accuracy and best example of analogical transfer was the gradual/alert condition. It appears that despite being an implicit learning task, explicit awareness of change may help alleviate analogical transfer costs when those changes occur gradually rather than abruptly. Thus, slower rates of change may enable an observer to make use of the alert and maintain performance with the II sampling shift.

Adaptation in Explicit vs. Implicit Category Learning

Theoretical Implications

Hypotheses concerning null effects in the present work should be considered provisional pending further study and perhaps would be better addressed using Bayesian analysis, which can provide a direct test of the null hypothesis. Nonetheless, the present work has implications for current theories of category learning. First, the translation of concept drift into perceptual category learning provides a theoretical framework to build upon in future research. Studies of categorical change can provide both insights into transfer and generalization, as well as build an understanding of the ways in which humans learn natural categories in the real world where environments are dynamic and change can be expected. This framework can be extended to investigate the effect of other types of change that might provide evidence for single vs. multiple systems theories of category learning.

Second, the first two experiments build on the knowledge base of how decisional elements are utilized in category learning. The results of Experiment 1 and 2 provide additional support for the differential processing of base-rate and payoffs in categorization. Criterion shifting was greater in response to base-rate change than payoff change, mirroring the results of previous work in stationary conditions (Maddox & Bohil, 1998, 2003). In addition, the results provided some evidence in support of enhanced analogical transfer in response to gradual change in base-rate conditions. This pattern of results provides additional support for the implicit learning of base-rates (Wismer & Bohil, 2017) whereas costs and benefits are believed to be reasoned about more explicitly.

Third, the results of Experiments 3 and 4 provide additional qualitative distinctions between what is believed to be implicit and explicit learning in categorization. Analogical transfer occurred in both RB and II categories, although in II categories it was precluded by an initial accuracy decrement in abrupt change conditions. Gradual change and alerting supported II analogical transfer, whereas change characteristics had little effect on RB analogical transfer since no initial accuracy decrement was observed, as expected. This opens up the question of exactly how gradual a change needs to be to support analogical transfer in II category learning, as well as if it can support transfer to other types of II category changes as well.

Practical Applications

Overall, the results of Experiments 1 through 4 suggest that the type of change referred to as "virtual concept drift" in machine learning, or "analogical transfer" in categorization, can be successfully adapted to by human observers. Tasks that involve components believed to be linked to implicit learning (e.g., as with base-rates or II categories) may be enhanced if change occurs slowly and if the learner is made aware to the possibility of change. Alerts do not seem to be necessary in tasks mediated by explicit learning where the change is large enough to be noticed (i.e., unexpected uncertainty; Yu & Dayan, 2005).

The development of decision aids in different contexts might be able to focus support in areas of the most need. With respect to the types of change investigated in this dissertation, this means tasks that require a focus on costs and benefits over accuracy or tasks invoking procedural learning. Particularly when changes of smaller magnitudes can be expected in a decision environment, alerts provided by machine learning algorithms may provide the learner with insight into when a change in strategy is needed. However, more work is needed before this type of information can be reliably provided for decision support.

Future Work

The categorical change framework established in this dissertation opens the door for many future research opportunities. One important consideration in the future will be to understand the effect of other types of change on category learning (i.e., "real" concept drift). Analogical transfer is only one small piece of the puzzle. Additionally, Webb et al. (2016) identified other important quantitative characteristics of concept drift that are of relevance to category learning. The first is the magnitude of change. The magnitude of RB and II category change in the present work was lower than that of Casale et al. (2012), providing one possible explanation for why II analogical transfer was found in the present work but not in Casale et al. (2012). A related metric is the severity of the concept drift, defined as the percentage of the stimulus space in which new stimulus-to-response mappings are created. The greater the severity of change, the more re-learning that needs to occur. Finally, a few researchers have investigated periodic changes (Navarro et al., 2013) or the possibility of multiple changes over time (Summerfield et al., 2011). Each of these components of concept drift provide interesting applications in category learning. The analysis of change effects can further strengthen our understanding of category learning, criterion lability, and a deeper look into the cognitive processes underlying change adaptation.

APPENDIX A: IRB APPROVAL LETTERS

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Andrew J. Wismer**

Date: **December 01, 2016**

Dear Researcher:

On 12/01/2016 the IRB approved the following human participant research until 11/30/2017 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review
Project Title:	Non-stationary Category Learning
Investigator:	Andrew J. Wismer
IRB Number:	SBE-16-12432
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 11/30/2017, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Kamille Chap

Signature applied by Kamille Chaparro on 12/01/2016 04:11:39 PM EST

IRB Coordinator

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Andrew J Wismer**:

Date: **November 02, 2017**

Dear Researcher:

On 11/02/2017 the IRB approved the following human participant research until 11/01/2018 inclusive:

Type of Review:	IRB Continuing Review Application Form Expedited Review Category # 6 & 7 This approval includes a Waiver of Written Documentation of Consent
Project Title:	Non-stationary Category Learning
Investigator:	Andrew J Wismer
IRB Number:	SBE-16-12432
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 11/01/2018, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Renea C Carver on 11/02/2017 03:15:01 PM EDT

Designated Reviewer

APPENDIX B: INFORMED CONSENT FORM

Characterization of Perceptual Category Learning

Informed Consent

Principal Investigator: Andrew Wismer, M.A.

Faculty Advisor: Corey Bohil, Ph.D.

Investigational Site(s): University of Central Florida, Department of Psychology

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being invited to take part in a research study which will include about 1000 people at UCF. You have been asked to take part in this research study because you are a college age individual. You must be 18 years of age or older, and have normal or corrected to normal vision, to be included in the research study.

The person doing this research is Andrew Wismer (*University of Central Florida – Department of Psychology*). Because the researcher is a doctoral student, he is being guided by Dr. Corey Bohil, a faculty advisor in the Department of Psychology.

What you should know about a research study:

- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this study is to explore how people can learn to classify patterns into different categories. This is important because it is the kind of judgment made on a daily basis by people such as medical personnel and baggage x-ray screeners. The cognitive abilities involved in this particular task are common to a wide range of tasks.

What you will be asked to do in the study: Participants will view a set of stimuli (e.g., lines) on a computer monitor and on each trial will be asked to make a judgment based on the features of the pattern. Participants will learn to determine what the patterns indicate (e.g., Category A or B). The entire experiment should take no more than 60-120 minutes to complete, depending on the amount of SONA credit you signed up for. If you signed up for 1 credit, the study will take no longer than 60 minutes. If you signed up for 2 credits, the study will take no longer than 120 minutes.

Location: The study will be conducted in the psychology building in PSY305, PSY303L, or PSY303K.

Time required: We expect that you will be in this research study for one hour (per SONA credit) and will require only one session in the lab.

Risks: There are no reasonably foreseeable risks or discomforts involved in taking part in this study.

Compensation or payment: Students will receive course credit through the SONA system.

Anonymous research: This study is anonymous. That means that no one, not even members of the research team, will know that the information you gave came from you.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt you, talk to Andrew Wismer, Graduate Student, Department of Psychology at andrew.wismer@knights.ucf.edu, or Dr. Corey Bohil, Faculty Supervisor, Department of Psychology at (407) 823-2755 or by email at corey.bohil@ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

2 of 2

UCF University of Central Florida IRB
IRB NUMBER: SBE-16-12432
IRB APPROVAL DATE: 11/02/2017
IRB EXPIRATION DATE: 11/01/2018

REFERENCES

Ackermann, J. F., & Landy, M. S. (2015). Suboptimal decision criteria are predicted by subjectively weighted probabilities and rewards. *Attention, Perception, & Psychophysics, 77*(2), 638-658.

Alaiz-Rodríguez, R., Guerrero-Curieses, A., & Cid-Sueiro, J. (2007). Minimax regret classifier for imprecise class distributions. *Journal of Machine Learning Research, 8*(Jan), 103-130.

Alaiz-Rodríguez, R., & Japkowicz, N. (2008). Assessing the Impact of Changing Environments on Classifier Performance. In S. Bergler (Ed.), *Advances in Artificial Intelligence* (Vol. 5032, pp. 13-24): Springer Berlin Heidelberg.

Alfonso-Reese, L. A. (2006). General recognition theory of categorization: A MATLAB toolbox. *Behavior Research Methods, 38*(4), 579-583. doi:10.3758/BF03193888

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442-481.

Ashby, F. G., & Ell, S. W. (2002). Single versus Multiple Systems of Learning and Memory. *Stevens' Handbook of Experimental Psychology*.

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

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol., 56*, 149-178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30*(5), 666-677.

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). 4 COVIS. *Formal approaches in categorization*, 65.

Ashby, F. G., & Soto, F. A. (1992). Multidimensional Signal Detection Theory. In (pp. 13-34).

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154-179. doi:10.1037/0033-295X.93.2.154

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat Neurosci, 10*(9), 1214-1221. doi:10.1038/nn1954

Bland, A. R., & Schaefer, A. (2011). Electrophysiological correlates of decision making under varying levels of uncertainty. *Brain research, 1417*, 55-66.

Bohil, C. J., & Maddox, W. T. (2001). Category discriminability, base-rate, and payoff effects in perceptual categorization. *Perception & psychophysics, 63*(2), 361-376.

Bohil, C. J., & Maddox, W. T. (2003). On the generality of optimal versus objective classifier feedback effects on decision criterion learning in perceptual categorization. *Memory & Cognition, 31*(2), 181-198.

Bohil, C. J., & Wismer, A. J. (2015). Implicit learning mediates base rate acquisition in perceptual categorization. *Psychonomic bulletin & review, 22*(2), 586-593.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.

Brown, S., & Steyvers, M. (2005). The Dynamics of Experimentally Induced Criterion Shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 587-599. doi:10.1037/0278-7393.31.4.587

Brown, S., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive psychology, 58*(1), 49-67. doi:http://dx.doi.org/10.1016/j.cogpsych.2008.09.002

Cantwell, G., Crossley, M. J., & Ashby, F. G. (2015). Multiple stages of learning in perceptual categorization: evidence and neurocomputational theory. *Psychon Bull Rev, 22*(6), 1598-1613. doi:10.3758/s13423-015-0827-2

Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory & Cognition, 40*(3), 434-449.

Ditzler, G., Roveri, M., Alippi, C., & Polikar, R. (2015). Learning in nonstationary environments: a survey. *IEEE Computational Intelligence Magazine, 10*(4), 12-25.

Drummond, C., & Holte, R. C. (2006). Cost curves: An improved method for visualizing classifier performance. *Machine learning, 65*(1), 95-130.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*(2), 107.

Franks, B. A., & Hicks, J. L. (2016). The reliability of criterion shifting in recognition memory is task dependent. *Memory & Cognition, 44*(8), 1215-1227.

Gama, J., Medas, P., & Rodrigues, P. (2004). *Concept drift in decision-tree learning for data streams.* Paper presented at the Proceedings of the Fourth European Symposium on Intelligent Technologies and their implementation on Smart Adaptive Systems, Aachen, Germany, Verlag Mainz.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR), 46*(4), 44.

Gifford, A. M., Cohen, Y. E., & Stocker, A. A. (2014). Characterizing the impact of category uncertainty on human auditory categorization behavior. *PloS Comput Biol, 10*(7), e1003715.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics Wiley.[arLEK, DDD. NAM, MT].

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical science*, 1-14.

Hassall, C. D. (2013). Learning in Non-Stationary Environments. *Thesis*.

Helie, S., Ell, S. W., Filoteo, J. V., & Maddox, W. T. (2015). Criterion learning in rule-based categorization: Simulation of neural mechanism and new data. *Brain and cognition, 95*, 19-34.

Hoens, T. R., Polikar, R., & Chawla, N. V. (2012). Learning from streaming data with concept drift and imbalance: an overview. *Progress in Artificial Intelligence, 1*(1), 89-101.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*(6), 418.

Huang, J., Gretton, A., Borgwardt, K. M., Schölkopf, B., & Smola, A. J. (2007). *Correcting sample selection bias by unlabeled data.* Paper presented at the Advances in neural information processing systems.

JASP Team. (2018). JASP (version 0.8.6)[Computer Software]. Retrieved from https://jasp-stats.org/

Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237.

Kantner, J., Vettel, J. M., & Miller, M. B. (2015). Dubious decision evidence and criterion flexibility in recognition memory. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.01320

Kelly, M. G., Hand, D. J., & Adams, N. M. (1999). *The impact of changing populations on classifier performance.* Paper presented at the Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining.

Kendler, H. H., & Kendler, T. S. (1970). Developmental processes in discrimination learning. *Human Development, 13*(2), 65-89.

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What's new in Psychtoolbox-3? Perception, 36 (ECVP Abstract Suppl.). In.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*(1), 1-17.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*: Psychology press.

Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 288.

Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1459.

Maddox, W. T., & Bohil, C. J. (2000). Costs and benefits in perceptual categorization. *Memory & Cognition, 28*(4), 597-615.

Maddox, W. T., & Bohil, C. J. (2001). Feedback effects on cost-benefit learning in perceptual categorization. *Memory & Cognition, 29*(4), 598-615.

Maddox, W. T., & Bohil, C. J. (2003). A theoretical framework for understanding the effects of simultaneous base-rate and payoff manipulations on decision criterion learning in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(2), 307.

Maddox, W. T., & Bohil, C. J. (2004). Probability matching, accuracy maximization, and a test of the optimal classifier's independence assumption in perceptual categorization. *Attention, Perception, & Psychophysics, 66*(1), 104-118.

Maddox, W. T., & Bohil, C. J. (2005). Optimal classifier feedback improves cost-benefit but not base-rate decision criterion learning in perceptual categorization. *Memory & Cognition, 33*(2), 303-319.

Maddox, W. T., Glass, B. D., O'Brien, J. B., Filoteo, J. V., & Ashby, F. G. (2010). Category label and response location shifts in category learning. *Psychological Research PRPF, 74*(2), 219-236.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 30*(37), 12366-12378. doi:10.1523/JNEUROSCI.0822-10.2010

Navarro, D. J., Perfors, A., & Vong, W. K. (2013). Learning time-varying categories. *Mem Cognit, 41*(6), 917-927. doi:10.3758/s13421-013-0309-6

Norton, E. H., Fleming, S. M., Daw, N. D., & Landy, M. S. (2017). Suboptimal criterion learning in static and dynamic environments. *PLoS computational biology, 13*(1), e1005304.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 87.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of" multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin and Review, 7*(3), 375-402.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*(1), 53.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*(4), 247-255.

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual review of psychology, 68*, 73-100.

Parsons, K. S. (2001). *Changes in signal probability and response bias in vigilance.* University of Cincinnati,

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews, 32*(2), 197-205.

Poldrack, R. A., Selco, S. L., Field, J. E., & Cohen, N. J. (1999). The relationship between skill learning and repetition priming: Experimental and computational analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 208.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology*, *77*(3p1), 353.

Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine learning*, *42*(3), 203-231.

Ratini, M. (2017). How Parkinson's Disease Progresses. Retrieved from https://www.webmd.com/parkinsons-disease/guide/parkinsons-disease-progression#1

Saerens, M., Latinne, P., & Decaestecker, C. (2002). Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure. *Neural computation*, *14*(1), 21-41.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics, 78*(1), 355-367.

Seger, C. A., Braunlich, K., Wehe, H. S., & Liu, Z. (2015). Generalization in category learning: the roles of representational and decisional uncertainty. *The Journal of Neuroscience, 35*(23), 8802-8812.

Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference, 90*(2), 227-244.

Smith, & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience & Biobehavioral Reviews*, *32*(2), 249-264.

Smith, Zakrzewski, A. C., Johnston, J. J. R., Roeder, J. L., Boomer, J., Ashby, F. G., & Church, B. A. (2015). Generalization of category knowledge and dimensional categorization in humans (Homo sapiens) and nonhuman primates (Macaca mulatta). *Journal of Experimental Psychology: Animal Learning and Cognition, 41*(4), 322.

Spellman, B. A. (1996). The implicit use of base rates in experiential and ecologically valid tasks. *Behavioral and Brain Sciences, 19*(1), 38-38.

Summerfield, C., Behrens, T. E., & Koechlin, E. (2011). Perceptual Classification in a Rapidly Changing Environment. *Neuron, 71*(4), 725-736. doi:http://dx.doi.org/10.1016/j.neuron.2011.06.022

Tsymbal, A. (2004). The problem of concept drift: definitions and related work. *Computer Science Department, Trinity College Dublin, 106*.

Vickers, D., & Leary, J. N. (1983). Criterion control in signal detection. *Human Factors, 25*(3), 283-296.

Warm, J. S. (1984). An introduction to vigilance. *Sustained attention in human performance,* 1-14.

Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L., & Petitjean, F. (2016). Characterizing concept drift. *Data Mining and Knowledge Discovery, 30*(4), 964-994.

Webb, G. I., Pazzani, M. J., & Billsus, D. (2001). Machine learning for user modeling. *User modeling and user-adapted interaction, 11*(1-2), 19-29.

Wickens, T. (2001). Elementary signal detection theory: Oxford University Press. *New York*.

Widmer, G., & Kubat, M. (1993). *Effective learning in dynamic environments by explicit context tracking.* Paper presented at the European Conference on Machine Learning.

Wismer, A. J., & Bohil, C. J. (2017). Base-rate sensitivity through implicit learning. *PloS one, 12*(6), e0179256.

Wixted, J., & Lee, K. (n.d.). Signal detection theory. Retrieved from http://www.kangleelab.com/sdt-d-prime-calculation---other-tips.html

Yamazaki, K., Kawanabe, M., Watanabe, S., Sugiyama, M., & Müller, K. (2007). *Asymptotic bayesian generalization error when training and test distributions are different.* Paper presented at the Proceedings of the 24th international conference on Machine learning.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681-692. doi:10.1016/j.neuron.2005.04.026

Žliobaitė, I. (2010). Learning under concept drift: an overview. *arXiv preprint arXiv:1010.4784*.

Žliobaitė, I., Pechenizkiy, M., & Gama, J. (2016). An overview of concept drift applications. In *Big Data Analysis: New Algorithms for a New Society* (pp. 91-114): Springer.