THE ROLE OF CATEGORICAL AND NUMERICAL REINFORCERS IN CATEGORY LEARNING

A Dissertation

by

ASTIN C. CORNWALL

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Darrell Worthy
Committee Members,	Brian Anderson
	Joseph Orr
	Zhangyang Wang
Head of Department,	Heather Lench

August 2021

Major Subject: Psychology

ABSTRACT

Real-world learning signals often come in the form of a continuous range of rewards or punishments, such as receiving more or less money or other reward. However, in laboratory studies, feedback used to examine how humans learn new categories has almost invariably been categorical in nature (i.e. Correct/Incorrect, or A/Not-A). Whether numerical or categorical feedback information leads to better learning is an open question. On one hand, numerical feedback could give more fine-grained information about a category, but may be more uncertain in early learning. On the other, categorical feedback is more dichotomous, potentially leading to larger error signals and more certainty about the outcome. In a series of three studies, the impact of categorical and numerical information was assessed via a multitude of differing category reward structures. To gain a basic understanding of the role that different feedback types have in category learning, Study 1 gave categorical feedback, variable numerical feedback, discrete numerical feedback, and feedback that combined both numerical and categorical information simultaneously to participants who were asked to categorize line stimuli which varied based on two prominent category learning rules. Study 2 expanded on these results and incorporated a basic reward learning manipulation into the task design. In this task, to understand how reward interacts with stimulus similarity, different category clusters were rewarded at different magnitudes with the idea that differences in behavior may arise based on a participant's sensitivity to either reward magnitude or stimulus similarity. Using a similar paradigm, Study 3 instead altered the rate at which different category clusters were observed. The category and reward learning literatures detail a bias towards stimuli that are more frequent, so this study attempted to determine the potential changes in behavior when stimulus frequency was congruent, or incongruent, with stimulus similarity. The results from each study detail that overall, people seem to learn better from feedback that contains categorical information or rewards that are discrete in magnitude. Further, based on fits of a connectionist model to the behavioral data, people are likely to rely more on stimulus similarity than to any difference in reward magnitude or observational frequency.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Worthy, and my committee members, Dr. Anderson, Dr. Orr, and Dr. Wang, for their guidance and support throughout the course of this research.

Thanks also go to my friends who provided welcomed non-work discussions and activities over the past few years. I would also like to thank my colleagues and the department faculty and staff for the support they have provided.

Finally, thanks to my mother and family for their encouragement throughout this process, and thanks to my wife and daughter for their patience, understanding, and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Darrell Worthy, Dr. Brian Anderson, and Dr. Joseph Orr of the Department of Brain and Psychological Sciences at Texas A&M University and Dr. Atlas Wang of the Department of Electrical and Computer Engineering at the University of Texas at Austin.

The contents of Study 1 were influenced by multiple discussions with Dr. Darrell Worthy, Dr. Tyler Davis of Texas Tech University, and Dr. Kaileigh Byrne of Clemson University. Revisions of the contents of Study 1 were based on the recommendations of the anonymous reviewers of the Journal of Cognition where this study was submitted for publication.

All other work conducted for the dissertation was completed by the student independently.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Dissertation Motivation and Structure

Learning is an integral part of the human experience. From the start, experiences we have are etched into memory as we learn what decisions in what instances lead to either positive or negative outcomes. Thorndike's law of effect (1898, 1927) details that behavior followed by a positive outcome has a higher likelihood of being repeated, whereas behavior followed by negative outcomes will be less likely to be repeated. Real world instantiations can include being told that you are correct in naming an animal a dog, to receiving a shiny new dime for helping an elderly person across the street. We consider these to be positive outcomes since they often have the effect of having us repeat our actions in similar situations in the future. Negative outcomes, such as touching the heating element of a stove, can also produce the desired effect of us withdrawing from pain and being more careful in the future (e.g. Herrnstein & Hineline, 1966). Even without someone explicitly telling you not to touch a hot stove again, the negative experience alone would probably suffice. However, these outcomes can take on a variety of forms outside of just being 'positive' or 'negative'. Outcomes can be discrete and categorical where the outcomes are fairly definitive: 'Correct/Incorrect' or 'Ice cream/No Ice cream' as two examples. Here it can immediately be understood whether the choice was 'good' or not as the outcomes are exclusive. Outcomes can also lie on a continuous range: exam grades, degree of sensation, pleasure felt, payment etc. With these outcomes, unless you have something to compare the outcome against, it may be difficult to immediately ascertain whether you made the right decision.

While we know that feedback, as defined here as a controlled outcome in a psychological paradigm, plays a critical role in learning, knowledge acquisition, and rate of learning (e.g. Herrnstein, 1970; Mory, 2004; Thorndike, 1927), we also know that certain types of feedback can have a differential impact on learning performance (Herzog & Fahle, 1997). Two such types, as mentioned prior, are categorical and numerical feedback. Together, both types of feedback encapsulate most of the main forms of outcomes we receive from the environment when learning. As such, it may be critical to understand the impact of each type of feedback on learning, and if there are any situations where one form of feedback is more beneficial. Currently, the differential impact of each type of feedback has been understudied as research typically focuses on whether or not feedback in general is effective (Pashler et al., 2005), or the schedule/timing of reinforcement (Behrens et al., 2007; Ludvig et al., 2011; Petter et al., 2018).

In the three experiments discussed in this dissertation, I will examine how aspects of categorical and numerical feedback impact how well novel categories are learned under different conditions. In the following parts of this section, I will give an overview of the prominent theories of category learning, discuss the facets of numerical reward learning and how it has been shown to impact decision-making, and finally I will detail the theoretical implications and impact of using both categorical and numerical reinforcers in category learning paradigms. As such, in this section, I will also give a brief overview of the use of formal learning models in category learning. This section, and Section 2, will provide the general background and methods for the experiments detailed in Sections 3 through 5. Section 6 at the end will present a discussion of the overall experimental results and implications.

1.2. Category Learning

Categorical feedback can be defined as a discrete number of mutually exclusive outcomes. Whether it be a bell tone, verbal commendation, or food, we can be certain that the feedback received leads to an outcome that is strictly defined (i.e. food or no food, ‘yes’ or ‘no’, etc.). Thus, the outcome of our decision can be immediately understood in most cases. In supervised category learning research, researchers predominantly employ different forms of discrete categorical feedback to promote learning (Ashby & Maddox, 2005). In these paradigms, participants typically view novel stimuli, attempt to classify the stimuli into a discrete number of categories, then receive feedback detailing the outcome of their choices.

The same process can be said to occur in everyday life as we are often inundated with categories. People, object, events, and ideas are all subject to being classified (Markman & Ross, 2003). Some things we observe may already be known and reinforce our concept of what constitutes a category, and other novel observations may still need to be categorized. We can come across instances where we must decide whether a long skinny object on the ground is a stick or a snake, or whether or not some berries are safe to eat or not. To make effective categorizations, the outcomes of the decisions need to be observed or generalized. In line with the above scenario, two discrete outcomes that could occur when reaching down to the long skinny object are that it was safe and I picked it up, or it was a snake and I got bit. Both outcomes are categorical, mutually exclusive, and would likely be reference for future decisions in similar situations.

Through multiple decisions and subsequent processing of feedback, we are able to more finely discriminate observed stimuli into one category or another, and become quicker as the categories are learned (Herrnstein, 1970). Understanding this process, and how we categorize and create categories, is the subject of a long and deep line of research (e.g. Ashby & Maddox, 2005, 2011). This, however, has led to the creation of differing theories of categorization and novel experimental paradigms. Below, I will define what a category is, then briefly detail a few of the leading theories of category learning along with how each utilizes feedback in the learning process.

1.2.1. What is a Category?

A category, in general, can be described as a division or group of things which show some form of commonality or are considered equivalent (e.g. Markman, 1989; Medin, 1989; Rosch et al., 1976). For example, when given a set of differing objects, different people may categorize the same items in different ways without given direction. As the number of objects increases, so too does the number of possible ways they can be categorized (e.g. Heit & Bott, 2000; Medin et al., 2004). Some may focus on the physical dimensions of color, shape, surface type, or other physical features, and others may have more subjective focus based on the way the objects make them feel among other things.

More definitively, according to Smith and Medin (1981), our ability to categorize is essentially a pattern recognition device. If we have the knowledge about the features that encapsulate a category, we should be able to categorize an object as a member, or non-member of that specific category with relative ease. This process also works in reverse as well. If we are instead told that a novel object belongs to a specific category, we can infer with some precision that the object features will be reminiscent of the features we believe make up the category. Since we are able to learn the general features that comprise categories, we do not need to remember every single instance of a scene or object to be able to associate a novel object to a known category. As such, the use of categories, and the features they are comprised of, allows us to extract the maximum amount of information we can from the environment with the smallest impact to cognitive processing (e.g. Barsalou, Lawrence, 1983; Markman, 1989; Rosch, 1978).

Interestingly, one object can belong to both a multitude of broad categories and a few precise categories. This is important as it highlights the idea that categories are not necessarily rigid constructs, but moreover the product of an ever-entangling taxonomy of features that allows us to quickly classify novel objects and recognize the already known (Rosch, 1978; Rosch et al., 1976). These features are, in a sense, modular verbal descriptions that can be applied to innumerous categories. For example, the category 'Animal' has a very extensive list of features. If we were to categorize all animals based on the fact that they include the feature 'fur' or not, we will have created two very broad sub-categories that will include most, if not all, animals combined. As more features are added to the category description, two things occur in an inverse relationship: the number of distinct animals in the category decreases and the degree of similarity between the animals remaining in the category increases (Tversky, 1977). Thus, objects in the same category will have a higher degree of similarity as compared to the similarity of objects in differing categories (Pothos & Chater, 2002).

Furthermore, as specificity in a category increases, given the inclusion of more defining features, the heightened similarity between category members allows a more intuitive understanding of what is and isn't a member of the category (e.g. Pothos & Chater, 2002, 2005). In certain forms of decision making, this would be invaluable. Take medical decision making for example. When viewing a tumor on a screen, two possible categorical judgements could be 'Tumor' or 'Not a Tumor'. Having a defined representation of what an object looks like, given prior knowledge, would make the difference between a correct and incorrect outcome (e.g. Reyna, 2008; Reyna & Farley, 2006). Thus, in general, how well we categorize our knowledge and learn from our experiences has an impact on how we make decisions that pertain to those categories (e.g. Fryer & Jackson, 2008; Varshney et al., 2011).

1.2.2. Theories and Models of Category Learning

In terms of how categories are learned, there are three main fields of inquiry: supervised learning (e.g. Ashby & Maddox, 2005; Medin & Schaffer, 1978; Nosofsky et al., 2019; Smith, 2014), unsupervised learning (Clapper & Bower, 1994; Love, 2003; Pothos & Chater, 2005; Zeithamova & Maddox, 2009), and semisupervised learning (Lake & McClelland, 2011; Vandist et al., 2009, 2019; Vong et al., 2016). Generally, the paradigms used to study each of the above types of category learning are consistent: subjects first view a stimulus and then make a decision about category membership (i.e. Ashby & Maddox, 2005), but other paradigms do exist. The main difference between each type of category learning reduces simply to the form of the feedback each subject is given. In supervised learning paradigms, subjects are typically given some form of discrete feedback as to whether their categorization was correct or incorrect, and in some cases, given additional information about what the correct category was (Maddox et al., 2008). Conversely, unsupervised learning gives no feedback at all. Participants, in general, are simply asked to separate the stimuli as they see fit (Ashby et al., 1999; Pothos & Chater, 2005). As it is believed that category learning in the real world is not wholly supervised or unsupervised, semisupervised paradigms were created to assess how categories are learned when feedback is sparse or sporadic (Vandist et al., 2009). However, in the effort to compare discrete and continuous feedback, I will only be discussing category learning in the purview of supervised learning as discrete feedback is the main form of reinforcement.

In category learning, there are four main views that have been held through the years that attempt/attempted to explain how categories are learned. These views are as follow: a classical view of categories that can be traced back to the philosophical works such as Aristotle's Categories (Aristotle, trans. 1975) and Kant's Critique of Pure Reason (Kant, trans. 1988), and more modernly to Fisher (1916), Hull (1920), and Vygotskiĭ (ed. 2012); the prototype view (Mervis & Rosch, 1981; Mervis et al., 1976; Reed, 1972; Rosch & Mervis, 1975); the exemplar view (Medin & Schaffer, 1978; Nosofsky, 1984, 1986); and the view that categorization relies on multiple systems working in tandem (e.g. Ashby et al., 1998). Each of which, utilizes discrete feedback in some form to learn the underlying category structures to determine category membership when given novel objects.

1.2.2.1. The Classical View

Early, categories were determined to be a system for which all things could be described. Stemming from the simple question of "What is this?", a judgment can be made as to what the object is based on existence within a particular category system. While these systems have been debated and evolved over the centuries, one common strand between them is that they sought to uncover the essential categories that form our cognition of phenomena. In a departure from the philosophical questioning of what a category is, early researchers such as Fisher (1916) and Hull (1920) sought to understand how categories were learned and how they could be applied. Using simple stimuli, it was discovered that novel stimuli could become associated with a certain category if given arbitrary labels as feedback. Later researchers such as Vygotskiĭ (ed. 2012), built upon the work of Hull and developed three stages of category learning: novel categorization is impressionable, categories then become grouped based on the relationship of other members, and finally true categories are developed. However, these early studies defined categories by a rule. For example, for an object to be categorized, it must fit within a strict boundary of an existing category: the category of triangle must have three points and three sides; category X includes all blocks of Y color. Essentially, to belong to a category, category members must be an exact match.

1.2.2.2. Prototype and Exemplar Views

Unfortunately, the classical view often failed to explain naturally occurring categories, categories that often exist in the real world, since some members of real-world categories do not adhere to the strict boundaries required by a rule (Mervis & Rosch, 1981; Rosch et al., 1976; Rosch & Mervis, 1975). Going back to the previous example of the category ‘Animal’, this would be considered a natural category. Under the classical view and its rule-based approach, we would have issues developing one unitary description that applies to all animals due to the inherent variation within the category. To solve this, two theories, which assumed that the learning of categories was more based on similarity rather than matching features, were proposed. The first of which, the prototype theory, details that people develop a representation of a category that reflects the features most often seen (Homa & et al., 1973; Mervis et al., 1976; Reed, 1972; Rosch et al., 1976). The second, the exemplar theory, states that we hold in our memory representations of each stimulus we have ever encountered. The previously seen stimuli represented in memory for each category are exemplars to which all other potential category members are compared (e.g. Medin & Schaffer, 1978). Interestingly, the only difference between both views is in the way categories are represented. The paradigms, use of discrete feedback, and the updating of representations are all fairly similar between the two.

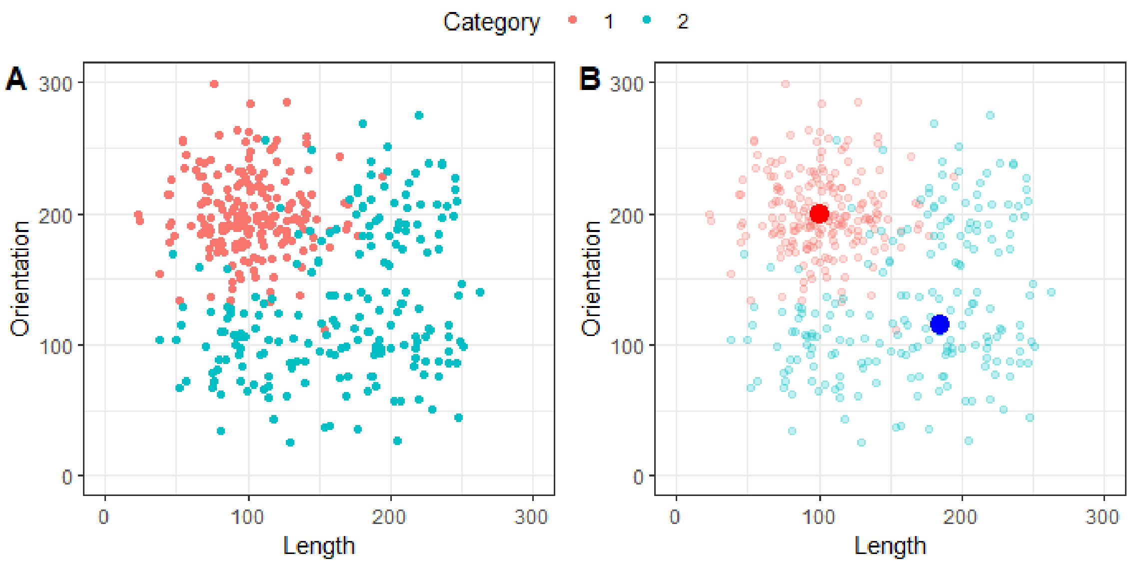

Figure 1.1 Visualization of the differences between exemplars and prototypes. A.) Each individual point observed is an exemplar of the category it is a member of. B.) There is a singular prototype for each category which is created from the amalgamation of all previously seen category members, denoted by the larger, darker points.

To clearly demonstrate the differences in category representation between each view, I will use the category of 'Dog' as an example. Provided that we have, over the course of our life, had multiple observations and experiences with dogs, we would probably have a pretty good idea of what features constitutes a 'Dog'. However, when a new animal is observed, how do you determine if the animal is a member of the 'Dog' category or not? The prototype view assumes that if the new animal is similar enough to what we believe is a prototypical dog, where a prototype can be thought of as the amalgamation of all dogs in memory, the new animal is also a dog. Conversely, the exemplar view assumes that the new animal is a dog as long as the new animal has a high degree of similarity with all other dog exemplars in memory. In each of these views, category membership is based on how similar the new stimulus is to either the prototypical stimulus or the exemplar stimuli stored in memory as compared to another category. If the new stimulus is more similar to one category than another, it is assumed to have a higher likelihood of being a member of that category. In each case, provided that the categorization attempt is deemed to be 'correct' via feedback, the representations for the category of dog will be updated and reinforced.

1.2.2.3. Multiple Systems View

To that end, Ashby and colleagues (1998) developed a theory that assumed category learning is governed by two systems: an explicit verbal system which is consciously controlled; and an implicit nonverbal system which utilizes procedural learning. This "competition between separate verbal and implicit category systems" model (COVIS) details that the system used is dependent on the category structure. For rule-based structures, not to be confused with the classical view of rule-based categorization, a verbal rule can be created that leads to optimal categorization performance. This rule can be simple such as all green items are category A, to more complex rules such as all big green circles are Category A, but small green circles are not. In these situations, COVIS details that the explicit system would dominate. For information-integration structures, where rules are typically difficult to verbalize (Ashby et al., 1998) and multiple dimensions need to be integrated at a pre-decisional stage (Ashby et al., 2003; Ashby & Gott, 1988), the procedural system would likely be more dominant.

The COVIS model has also been extended as a neurobiological model detailing that implicit and explicit systems of COVIS functionally mirror systems in the brain (Ashby et al., 2011). Much like the prior theories of category learning, categorical feedback is utilized in the learning process. However, in these theories, learning of categories from feedback is assumed to occur since performance drastically changes dependent on whether feedback is present or not (Homa & Cultice, 1984). Additionally, using COVIS, researchers were able to show that particular regions of the brain were correlated with either 'correct' or 'incorrect' categorizations and that the processing of feedback differed based on whether the responses are correct or not (Milton & Pothos, 2011; Nomura et al., 2007).

1.3. Reinforcement Learning

Numeric feedback can be defined as a set of outcomes that include a range of number values or have numeric properties. Two prominent examples include monetary values and points. Dependent on the paradigm, monetary and point-based feedback could consist of range where positive outcomes are any number greater than zero, and negative outcome are zero and negative values; or the feedback could consist of discrete values such as gaining or losing a dollar or point. Reward and punishment via numerical values are outcomes that we often encounter in our daily lives. Whether the outcomes are positive or negative, or implicit or explicit, we are able to utilize the information gained to learn about the outcomes of our actions and better predict what will occur in similar situations in the future. However, depending on the amount of uncertainty surrounding the point values, the values being either more discrete or continuous for example, the rate at which we learn from numeric values is impacted (Walker et al., 2019) as the value of the reward may not be as readily understood.

In reinforcement learning research, numerical rewards are used to demonstrate how people/agents learn to associate certain actions to situations in their effort to maximize a numerical signal (Sutton & Barto, 2018). By using trial and error, and utilizing feedback information, people are able to learn to predict, and optimally acquire, reward (Gershman & Daw, 2017). Should we be rewarded for our actions or decisions, we are more likely to replicate what worked in the same or similar situations (Schultz, 2016; Thorndike, 1927). Should we be punished for our choice instead, by way of losing points or money for example, we may be more hesitant to make the same decision again. Importantly, this differs from prediction learning where outcomes are predetermined regardless of an agent's response (Dayan & Balleine, 2002). In essence, reinforcement learning, in humans, is the research of optimal behavior: how agents make decisions that maximize positive outcomes and minimize any negative outcomes. Like category learning, there are multiple approaches researchers have taken in an effort to understand how people learn from reward. Unlike category learning however, reward values can differ in a variety of ways. Below, I will briefly detail some of the ways reward values are manipulated as well as briefly detail a couple of the main approaches used in the study of reward learning.

1.3.1. What is Reward?

Reward is a term that has a broad definition. In general, a reward can be described as any form of outcome, or reaction, in response to a stimulus that is perceivably positive (e.g. Wise, 1989). As such, whether fortunately or unfortunately, this means that anything can be regarded as a reward. Thus, the notion of what is rewarding can differ between people, but it can also be learned and extinguished (Haruno & Kawato, 2006; Rose & Behm, 2004). Other definitions of reward may refer to the increase in striatal dopamine (Bódi et al., 2009; Glimcher, 2011; Moustafa & Gluck, 2011) or anything that elicits a reward prediction error (e.g. Schultz, 2016) in the neurobiological domain, and it could also refer to a positive numerical signal as used in reinforcement learning (RL; Sutton & Barto, 2018). In reward learning, reward can take on both discrete and continuous forms. Discrete forms of reward could include food, coins, dichotomous point values etc., whereas more continuous forms of reward can include a wide variety of both explicit and implicit things such as a range of numerical values or the subjective amount of emotion felt.

However, whether a reward is present or not is not the only factor that determines the rate at which learning occurs. The magnitude, probability, and frequency of reward each have a unique impact on determining optimal choice within human reinforcement learning.

1.3.1.1. Reward Magnitude

Numerical reward values can exist with differing orders of magnitude. A value of zero can be understood to be less than a value of 1. However, the comparison between 0 and 1 would likely be different than the comparison between 0 and 100. To showcase this, in both animals and humans, food is a salient reward. When food is given or taken away, a reward signal is produced. Interestingly, this signal can change depending on the magnitude of the reward given. In rodents, while differing amounts of food have been shown not to affect the rate at which a task is learned, the rodents that received more food showed higher extinction rates than rodents who received less food when food was no longer rewarded for task completion (Capaldi, 1966; Skinner, 1938; Tolman, 1948). For humans, the impact of differing reward magnitudes depends on the application. In situations where differing magnitudes of rewards are interspersed within trials, differing values are not shown to have an effect behaviorally or neurobiologically (Bellebaum et al., 2010a). However, when the magnitudes of rewards are consistent within groups or alternatives, larger reward magnitudes have a positive effect on learning (e.g. Weinstein, 1971). Research in visual attention has also shown that the larger a reward is, the more likely it is to capture attention as compared to conditions where smaller, or no, rewards were given (e.g. Anderson et al., 2012; Anderson & Halpern, 2017). Additionally, in tasks such as the Iowa Gambling Task (Bechara et al., 1994), modulation of the differences in reward magnitude between alternatives can lead to overall differing levels of task performance (Van Den Bos et al., 2006) and event-related potentials (Meadows et al., 2016).

1.3.1.2. Reward Uncertainty

Sometimes, rewarding outcomes may be uncertain. When repeatedly making the same decision, the outcomes could vary. The reward could be larger or smaller, or sometimes you may receive a reward whereas sometimes you do not. Learning from probabilistic and uncertain rewards is a hallmark in the reinforcement learning literature (Behrens et al., 2007; Daw et al., 2005; Gershman, 2018). In paradigms such as these, reward values may change over time or certain alternatives have differing probabilities of reward. Thus, creating a dilemma for potential participants: should all of the alternatives be sufficiently explored, or should a perceivably rewarding option be exploited (e.g. Cohen et al., 2007; Gershman, 2018a)? Much like in the real world, the predicted outcomes of our decisions are uncertain until we make a choice. Even then, without anything to compare it against, there still exist a degree of uncertainty. There is the possibility that the choice we made was optimal and we would lose out by exploring other alternatives. Conversely, we could make the same assumption about a suboptimal alternative and, while we may be consistently rewarded, we are still losing out. The same can be said for probabilistic rewards. Without sufficient exploration, we might never know if our choice is the most optimal. Tentative solutions to this dilemma lie in determining the optimal tradeoff between exploring and exploiting (Addicott et al., 2017; Constantino & Daw, 2015). Concisely, how much should we explore before we exploit? Some accounts promote that optimal performance is obtained via random exploration, while other believe that directed exploration produces better results (e.g. Wilson et al., 2014). However, it seems the optimal strategy is a combination of both structured and stochastic exploration before exploitation (Gershman, 2018b; Krueger et al., 2017). Relatedly, if the degree of certainty and uncertainty are believed to exist on a range, the more certain the environment is, the more structured exploration is predicted to be. Conversely, as the degree of uncertainty increases, the more stochastic exploration becomes in search of the optimal strategy (Cohen et al., 2007) and the less optimal the responses become (Walker et al., 2019).

1.3.1.3. Reward Frequency

Rewarding occurrences can also occur at varying frequencies. In conjunction with reward uncertainty as described above, decisions can sometime result in a rewarding outcome or an outcome that is less so or non-existent. A real-life example could be in the choice between two restaurants. Picking one could possibly lead to either a great meal or a poor one. By repeatedly trying both restaurants, you may find that going to one may lead to far more great meals. As such, choice may be biased towards the more frequently rewarding decision. Despite the amount of surprise the reward from the infrequent option may give, the infrequency of reward may have a negative impact on the association between the choice and the reward (Balleine & Dickinson, 1998; Schultz, 2006). Further, the rate at which options are observed, or rewards are obtained, also leads to a differing perception of the choices themselves. In probability learning tasks, choices that are observed more frequently tend to have their actual probability value overestimated (Estes, 1976). Based on the above discussion, we know that when given the choice between two alternatives that differ in magnitude, the alternative with the largest reward is preferable. However, recent work has shown that this preference for the largest magnitude of reward can be overridden by the manipulation of how frequent the rewards are given. When given a choice between a suboptimal option (lower reward magnitude) that has been more frequently rewarded and an optimal option (higher reward magnitude) that has been rewarded less frequently, choice is biased towards the suboptimal, more frequently presented option (Don et al., 2019). In cases such as these, while the expected value for the infrequent options may be larger overall, the more frequently rewarding options are often associated with a larger cumulative reward over time due to how often rewards are actually received. In a comparison between reward magnitude and frequency, differences between gender can also be observed. When given the choice between options that lead to probabilistically more frequent rewards, or less frequent larger rewards, females tend to choose the most frequently rewarding options whereas the males tend to choose the options with the perceivably higher rewards despite the infrequency in actually receiving them (Cornwall et al., 2018; O'Brien & Hess, 2020).

1.3.2. Approaches of Reinforcement Learning

When in a situation where we must choose between various alternatives, such as what restaurant to eat at, which pond leads to the biggest fishing haul, or what slot machine is likely to lead to the most payouts, we often do not know which option is the most optimal. However, through repeated sampling of each alternative, we can begin to make some associations between both the magnitude and frequency of the rewards obtained and the alternatives we chose. Per Sutton and Barto (2018), reinforcement learning is a process of selecting various alternatives, association of outcomes, finding the optimal solutions, and connecting actions to certain situations. Much like how category learning has multiple views of how a category is learned, there are a few different approaches the reinforcement learning literature has taken in algorithmically defining behavior.

1.3.2.1. Expected Value

When making decisions based on rewarding outcomes, it is often a reasonable strategy to determine the expected value of all alternatives in order to make an informed decision (Dayan & Abbott, 2001; Rolls et al., 2008). As such, when confronted with numerous options, we can repeatedly sample each, and begin to understand what the average reward, or outcome, is for each. Depending on our current goals, we can decide which option is the best. However, in most cases, it would be the option with the largest outcome that is deemed optimal. A model for this type of behavior can be found in the delta rule (Rescorla & Wagner, 1972; Widrow & Hoff, 1960; Williams, 1992). When first encountering a set of alternatives, the expected value of each is assumed to be equal. Over time, as each alternative is selected and rewarded, the expected values for each fluctuate. Should the reward received be greater than the current expected value, a positive prediction error occurs and the expected value for the rewarded alternative increases proportional to a learning rate. Relatedly, if the reward does not meet the expectation, the expected value will decrease instead. While this model can account for behavior in discrete situations, in situations where the alternatives are presented at differing frequencies (Don et al., 2019), the delta rule can make errant predictions.

1.3.2.2. Markov Decision-Making

In the study of determining optimal decision-making, Markov Decision Processes (MDP) efficiently model decision-making tasks that include multiple, sequential timesteps (Bellman, 1957; Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 2018). A MDP is assumed to begin with a blank slate (i.e. no prior knowledge) and consist of a series of discrete states, actions, rewards that culminate in the probability value that an action will occur given a particular state and time (Gershman & Daw, 2017). A state can be any situation where we are confronted with a choice: do you pick the right or left option? After making our decision we make a choice (action), receive a reward, and find ourselves in the next state where we must choose between the options again. However, this time, we have the information we learned from the previous choice that we can use to make all future choices.

1.3.2.3. Model-Free and Model-Based Learning

Often, when it comes to making decisions, we can proceed based on two reinforcement learning methods: model-based or model-free decision-making. Simply, in model-based decision-making, we can learn a model of the structure of the given task that we can use to predict the outcomes of any given state. While it can lead to accurate decision making, it is more computationally intensive and effortful. In model-free decision-making, we forego the model and learn the utility of each action given a particular state and tend to choose whichever option has led to the most reward in the past (e.g. Thorndike, 1927). In contrast to model-based, model-free decision-making is computationally more efficient, but inflexible to changes. For a clear example of the distinction of both, we can look to an example given by Dayan and Niv (2008) where we are confronted with a decision of which route to take home. Given a certain time of the day, should we take the freeway or take the back road? Using model-based decision-making, we could check out a traffic app on our phone to make an informed decision. Conversely, using model-free decision-making, we would check the time and understand that during this same time in the past, the freeways were packed, and the backroads are much more efficient. However, as some researchers have determined, solely using one method or the other is not practical. Moreover, both forms of decision-making are thought to occur in parallel and each type dominates in certain situations (c.f. Gershman & Daw, 2017).

1.4. Theoretical Application of Categorical and Numerical Reinforcers

In general, supervised category learning paradigms include discrete feedback in the form of categorical (Yes/No or Correct/Incorrect) or cognitive (Green Light/Ding Sound = Correct; Red Light/Buzz Sound = Incorrect) feedback to aid the classification of novel stimuli (e.g. Ashby & Maddox, 2005, 2011; Nosofsky et al., 2019; Salatas & Bourne, 1974; Smith & Medin, 1981; Smith, 2014). More recently, category learning research has drawn on some of the methods of reinforcement learning to determine how numerical reward impacts category learning performance (Apitz & Bunzeck, 2012; Daniel & Pollmann, 2010), how reward magnitude or losses contribute to how well category memberships are learned (Abohamza et al., 2019; Moustafa et al., 2015; Schlegelmilch & von Helversen, 2020), and how reward guides attention to categories (e.g. Hickey et al., 2015). However, the reward values used in these paradigms are typically static numerical values or images of currency. Similarly, reinforcement learning tasks typically have a discrete number of choices where participants must iteratively learn which choices are more valuable based on either discrete or continuous numerical feedback (Daw et al., 2006; Erev & Barron, 2005; Frank & Claus, 2006; Kool et al., 2017; Niv, 2009; Sutton & Barto, 2018). The flexibility afforded by RL paradigms in the use of either discrete or continuous numerical values has allowed researchers to assess multiple facets of how differing forms of reinforcers and reinforcement schedules modulate learning and motivation. Though the framing of the tasks and types of typically given feedback differs, as mentioned by Radulescu et al. (2019) and detailed in the prior sections, feedback is critical in changing future behavior in both category and reinforcement learning paradigms. Collectively, both supervised category and reward learning are feedback-dependent processes that shape how individuals learn new information. Figure 1.2 below details some example visual differences that people may see when taking part in either type of task.

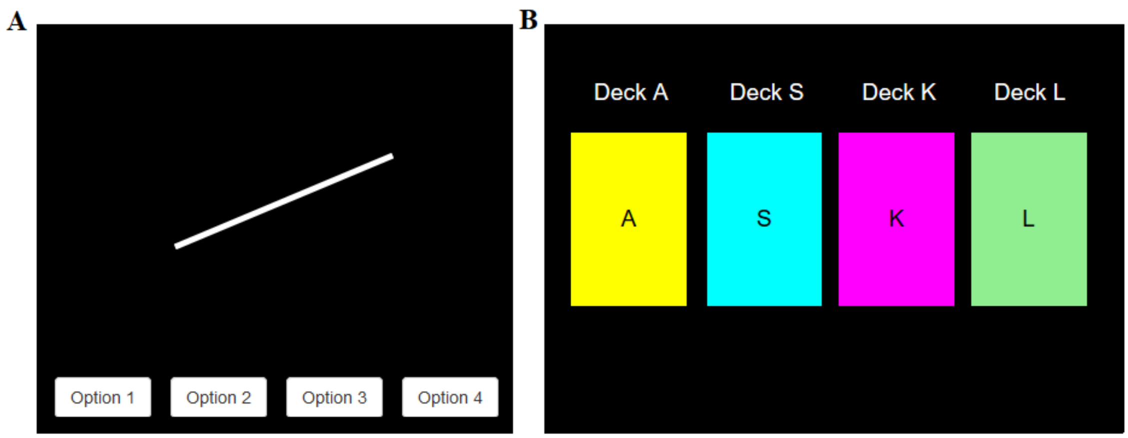

Figure 1.2 Example of the visual differences between a category learning task (A) and a reward learning task (B). Category learning tasks typically have a stimulus with potential options that the stimulus can be categorized into, and reward learning tasks typically have a set of options that participants can free choose from to reveal the potential rewards.

As mentioned, in both supervised category and reward learning, the feedback given is crucial in determining future responses and both require a degree of trial and error before becoming proficient. Generally, for category learning, relevant features are learned through repetition and reinforcement via corrective (i.e. categorical) feedback. Similarly, in reward learning, the optimal choices are determined through multiple sampling attempts and the maximization of positive reinforcement (and minimization of punishment or negative outcomes) most often via a numerical reward signal. However, the processing of feedback and subsequent determination of choice may not be entirely dissimilar between categorical and numerical feedback.

1.4.1. Similarities in Computation

Algorithmically, models of category and reinforcement learning assume that learning is based on past experiences (e.g. Gershman & Daw, 2017; Kruschke, 2012) and the minimization of the error between observed and expected outcomes, commonly known today as the prediction error (e.g. Glimcher, 2011; Schultz, 2017). In category learning, minimization of the prediction error results from adequately learning which features predict categorization as the expectation is that each decision is correct. Similarly, for reward learning, the error is minimized by learning which options lead to a reward, or reward value, that is equal to, or greater than, the expected reward. The prediction error formulation itself can be traced back to Bush and Mosteller’s model of simple learning (1951) and the Widrow-Hoff learning rule (Widrow & Hoff, 1960), but it is more recognized as part of the Rescorla-Wagner model (Rescorla & Wagner, 1972). In these models, and a variety of category and reward learning models, similar error computations are used to update the weights of certain options or states (i.e. delta rule models), or update the weights of category exemplars (i.e. ALCOVE).

Overall, for both sets of models, as the expectations for each alternative are met, and the prediction errors therein are minimized, learning is assumed to have occurred. With each type of model having the goal to not only learn what is more rewarding or correct, but also to sufficiently predict choice given the learned information, each type of model also utilizes a similar method of determining the choices of agents at any given time step. The models do so through a type of ‘winner take all’ component such as Luce’s choice rule (Luce, 1977) or related softmax function (e.g. Daw & Doya, 2006; Knox et al., 2012), or simply which choice results in the highest value (argmax). In reinforcement learning models, this is typically based on a weighted comparison between the expected values of each choice, or which alternative has the largest potential value based on the model. For category learning models, rather, it is often the weighted comparison of the summation of the similarity or distance values of each category based on the currently viewed stimuli. In each type of model, the choice, or category, with the highest value is predicted to be the most likely choice. Essentially, models of each form of learning create a formal representation of what is likely occurring ‘under the hood’ and attempt to model behavior in mathematical terms.

1.4.2. Similarities in the Brain

Neurobiologically, research has shown that the midbrain dopaminergic system shows signs of heightened activity during supervised category learning (e.g. Knutson et al., 2001; Shohamy et al., 2008) and reward learning in general (Gershman & Uchida, 2019; Steinberg et al., 2013). This area of the brain is believed to be responsible for learning via a biological equivalent of the reward prediction error, the modulation of striatal dopamine (Bayer & Glimcher, 2005; Engelhard et al., 2019; Nomura et al., 2007). Through multiple instances of trial and error, phasic spikes of dopamine are assumed to be associated with positive predictions, and likewise phasic dips are associated with negative prediction errors. Interestingly, many of the areas in the brain associated with reward learning and dopaminergic processes are also believed to be part of the COVIS explicit system of category learning (Ashby et al., 2011). As described prior, COVIS assumes two systems to be present in category learning, the explicit and conscious system which utilizes rules to learn, and the implicit system which infers category membership based on experience.

Studies involving motivation in primates showed that as the contingencies surrounding reward were learned, the rewarding outcome itself was no longer the source of heightened dopamine neuron activity, rather that the conditioned stimulus, and the act of predicting reward, showed the largest amount of neuronal activity (Schultz et al., 1993). This shift in dopamine activity also occurs in conjunction with the rate at which responses are made. Thus, as the outcomes are learned, the speed at which decisions are made increases, and the dopamine activity associated with the reward signal decreases. This, again, relates to the COVIS model that assumes that as categories are learned, there is a shift from the explicit system to the implicit system cooccurring with decrease in the time spent determining category memberships. In sum, neurological activity associated with reward learning is likely similar to the activity associated with category learning.

1.4.3. Application

Thus, given the similarities and prior applications of categorical and numerical feedback in category and reinforcement learning, we can begin to theoretically explore how each type of feedback will impact learning in paradigms where the given type of feedback is not often used. From both a modeling and a neurobiological perspective, we should expect that both discrete and continuous feedback is sufficient for learning in either a category or reinforcement learning paradigm. Looking strictly at the prediction error calculation that is present in most models between each literature, the difference between the expected outcome and actual outcome, we can make some predictions about how each form of feedback would affect learning performance. If categorical and discrete feedback are treated as dichotomous outcomes of 'reward' and 'no reward' (coded 1 and 0; e.g. Ashby et al., 2011), and we have numerical information scaled to the 0 to 1 range, we can begin to theorize about the potential differential impacts of both types of feedback.

With discrete numerical and categorical feedback, the optimal response would always lead to positive prediction error, or at the least never result in a negative prediction error. If the current paradigm is categorization, that would mean that we chose the correct category, similarly for a reward learning task, this would be whether or not a reward was received. Consequently, making the wrong decision, likely resulting in a non-rewarding outcome, would lead to a negative prediction error in nearly all cases. With what we understand about the dopaminergic system, the positive prediction error should result in increased dopamine activity in the brain and thus better association between the reward received and the choice. Thus, when given these types of feedback, performance should theoretically be similar. Conversely, with numerical feedback which consists of a more continuous range of rewards, it is more likely that the magnitude of the prediction errors will be less than that of discrete numerical and categorical feedback in most cases. If you imagine that, during a task, you learn that one response over another leads to more rewards in general. For this most rewarding option, there is an expected reward of 0.75 (scaled to be between 0 and 1). This means that in order to elicit a positive prediction error, the reward value would need to be greater than 75. Thus, since most reward learning tasks with numerical rewards introduce some variability to the rewards given, it is possible that the sign of the prediction error could go in either direction. In addition, given sufficient learning, the magnitude of both prediction errors are likely to become smaller in magnitude, and thus prompt smaller amounts of dopamine activity as compared to larger prediction errors.

1.4.4. Simulation

To visualize the possible differences in the effectiveness of each feedback type in category learning in particular, I have simulated a simple information-integration category learning task (e.g. Ashby & Maddox, 2005; Daniel & Pollmann, 2010) using the ALCOVE category learning model (Kruschke, 1992). ALCOVE can be modified to update exemplar weights using either categorical or numerical feedback information (all model formalisms will be described in detail in Section 2). The task consists of 400 trials and 2 categories. On each trial, an agent observed a novel line stimulus that varies along the two dimensions of line length and orientation. Over the course of the task, the agent will see 200-line stimuli that belong to each category in a randomized manner. The category structure and individual stimuli can be seen in Figure 1.3 below.

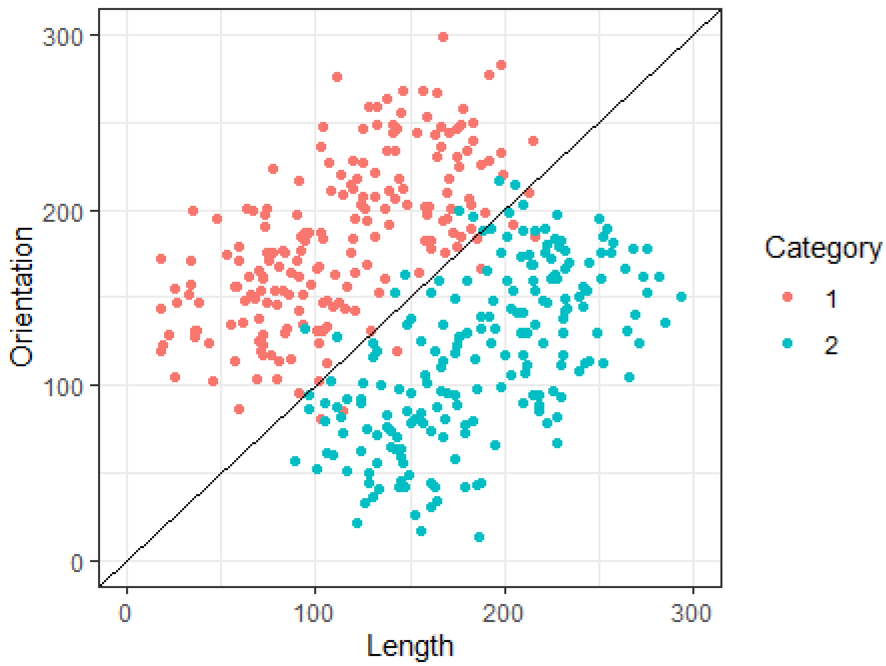

Figure 1.3 Depiction of the category structure for this simulation. Red dots represent stimuli that belong to category 1, and category 2 for blue dots.

When given categorical feedback, the agent is only given information regarding whether a categorization attempt was correct or not (coded as 1 and 0). For numerical feedback, four conditions were created: a condition where the largest rewards were given for the most typical stimuli (based on probabilities derived from the GCM: Nosofsky, 1986), where the largest reward were given for the most atypical stimuli (1 – GCM Probability), a condition where the rewards are uniformly distributed amongst each category, and a final condition where discrete rewards of 100 and 0 are used for correct and incorrect categorizations respectively. In Figure 1.2a below, the predicted proportions of correct categorization and the associated prediction errors, are visualized and separated by feedback type.

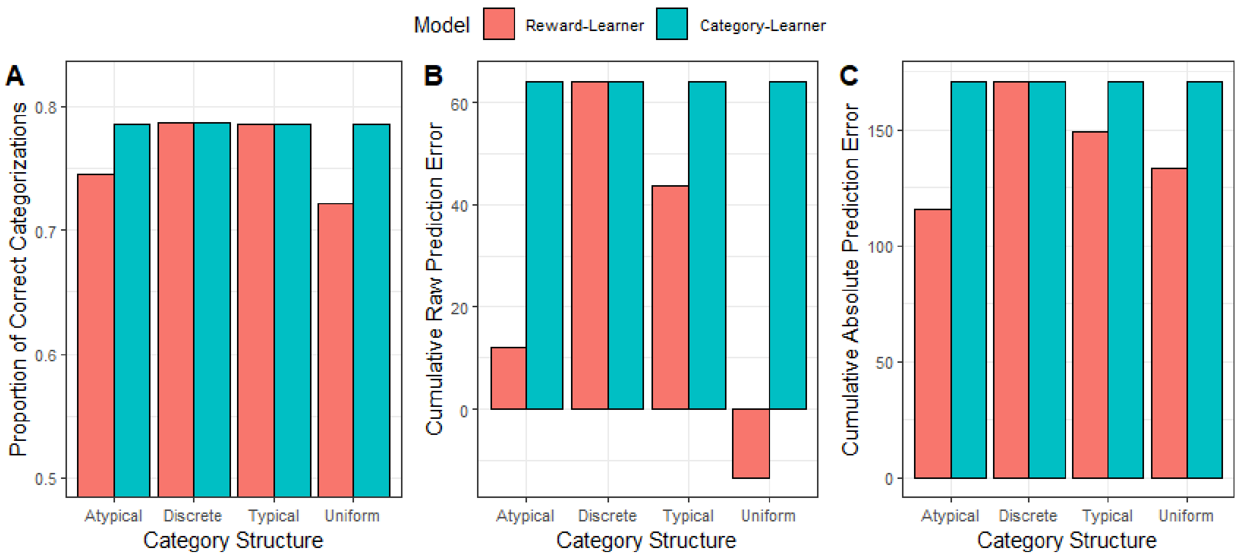

Figure 1.4 Simulated output of a category learning task with 500 simulated agents given categorical and numerical feedback. A.) Proportion of correct categorizations by type of reward feedback and model. B.) Cumulative raw prediction errors for each feedback type. C.) Cumulative absolute values of the prediction errors for each condition.

In Figure 1.4a above, the ALCOVE category learner model predicts that agents overall will learn to categorize the novel stimuli with a fair degree of accuracy. Like described, the predicted accuracy and prediction errors for categorical feedback and discrete rewards were near identical which further reinforces the idea that categorical and discrete numerical feedback are processed in a similar fashion. Interestingly, the reward conditions where the largest rewards are given for the most typical stimuli resulted in predicted accuracy similar to that of discrete rewards and categorical feedback. However, the overall magnitude of the predicted cumulative prediction errors were smaller. This suggests that, even though the rewards were more variable, the large amount of reinforcement given for successful categorizations of the most typical stimuli led to the most representative exemplars in the model having the largest weights. Thus, creating a defined representation of what features are indicative of each category.

Conversely, when the most atypical stimuli were the most rewarded, predicted accuracy is poorer and resulted in a larger amount of negative prediction errors per Figure 2b-c. Finally, when rewards were uniformly randomized, predicted accuracy was the poorest out of each condition, and resulted in more negative prediction errors than positive over the course of the task. Overall, the simulation predicts that both categorical and numerical feedback would likely be effective in promoting learning in a simple category learning task. However, it is clear that different types of numerical reward feedback are predicted to have a differential impact on task proficiency based on the magnitude of the reward and how it is associated to the stimuli.

1.5. Impact of Categorical and Numerical Reinforcers in Category Learning

Based on the simulation above, the type of feedback given has the theoretical potential to differentially impact both the rate at which the optimal responses are learned as well as how well the task is learned. Discrete categorical or numerical feedback gives an immediate idea of whether the decision was correct/optimal. In contrast, numerical feedback consisting of a larger range of values moreover gives an idea of how right the decision was without any indication of the correctness of the choice. Thus, the discrete feedback may more quickly orient people towards the correct outcomes, but they may not learn much about the contingencies surrounding the choice other than it was correct or not. Conversely, more continuous numerical feedback may take a bit longer to learn, but the knowledge about the choices may be more defined. However, the difference in framing here may lead to lasting effects in terms of overall performance. In categorization tasks in particular, the simple process of asking participants to either predict outcomes (as typically asked in reward learning task) or categorize stimuli led to distinct differences in performance (Bott et al., 2007), where categorization showed improved performance.

As discussed prior, recent category learning research has attempted to determine how reward feedback influences categorization performance. Based on the simulation, depending on how reward is utilized, categorization performance can vastly differ, and the category learning literature seems to corroborate this. For instance, studies that included discrete monetary feedback in addition to categorical feedback demonstrated that behavioral performance (i.e. accuracy) did not differ between trials with reward and trials without. Though, the rewarded trials did show increased brain activity in the reward areas of the brain as compared to categorical-only feedback trials (Daniel & Pollmann, 2010; Peterson & Seger, 2013). Further, in line with our simulated example, studies that forewent categorical feedback and solely used a form of value-based reward feedback showed that categories can be learned from numerical information alone (Abohamza et al., 2019; Daw & Shohamy, 2008; Moustafa et al., 2015). Additionally, Kahnt and colleagues (2012) demonstrated that the generalization of categorical knowledge can occur without categorical feedback. After sufficiently learning the categories during training when given discrete numerical rewards only, participants demonstrated neural reward and prediction error responses when shown stimuli that were not seen nor rewarded prior during test. This suggests that not only can people learn to categorize novel objects solely from numerical rewards, but that the reward values were also an effective enough feedback device to prepare people for generalization of categorical knowledge. However, reward has also been shown to promote suboptimal decision-making during category learning. Schlegelmilch and von Helversen (2020) found that the modulation of reward magnitude affected both the rate at which categories were learned as well as the overall accuracy in training and transfer portions of a task. They found that simply rewarding one category exemplar a significant amount more than other exemplars led to no change in learning for the more rewarding exemplar, but poorer learning for all other exemplars. Thus, reward can also have a counterintuitive effect on category learning.

Overall, these studies detail that numerical feedback promotes learning that is either equal or poorer than categorical feedback. Thus, one question this raises is whether or not there are situations when numerical feedback is more beneficial. One possible answer goes back to the framing of both category and reinforcement learning. In category learning, a given stimuli can only be one of a few categories. In this sense, both the choices and the outcomes are discrete and categorical. There is little uncertainty in expectation as to whether the next stimuli will be completely new, as in something that is entirely dissimilar from stimuli already observed. Conversely, reinforcement learning outcomes can include a wide range of reward values which introduces uncertainty into the decision-making process as the next time the same option is chosen, the reward value may differ. However, people are still able to learn to make optimal decisions. In reinforcement learning paradigms, categorical feedback may be counterproductive. For example, what is considered to be a correct choice if all alternatives give rewards? In terms of probabilistic reinforcement, this would be whether a reward is received or not, but it may not be as clear for other forms of continuous rewards. Further, if all you are told is that you are correct, you could consistently choose a rewarding option, but you would have no idea if it is the most optimal, and goes back to the dilemma between exploration and exploitation. In such a task, the information gained from exploring and categorical feedback would be equated amongst all options and exploitation has the possibility of being sub-optimal. Numerical feedback on the other hand, would be able to impart this information. As such, there is the possibility that improved learning could be the result of using the proper feedback in the proper situations.

Using a real-life example, this would be similar to getting an exam grade. Say you studied hard, and the test was over multiple topics, and you were unsure of how well you would perform. You get your test back and it simply says that you passed. While this may be cause for celebration, it does not give any idea as to how well you did or if you need to study more. You could have barely passed, or you could've aced the exam. Rather, a numeric value of 85 gives more fine-grained information. You know that you passed, based on your knowledge of the grading scale, and you also know that you did fairly well, but there is also room for improvement. Similarly, should you be an aspiring radiologic technologist learning to classify tumors in x-rays, being told that what you are viewing is indeed a tumor, or specific injury, is likely more helpful than being told that the image has a score of 85. You would first have to learn what the score means, and how it relates to your actual goal of correctly classifying the image. Thus, in paradigms with numerous and variable outcomes, continuous information may be more useful, whereas for paradigms where the outcomes are more certain, discrete feedback is likely to be preferable.

Similarly, in laboratory category learning paradigms, should the outcomes be uncertain, more continuous numerical feedback may play a larger role than discrete feedback. This uncertainty is most often observed when first learning novel categories. Representations of relevant features are still unknown and similarity determinations may not be effective. Likewise, in reward learning with outcomes that are more certain (i.e. dichotomous outcomes), discrete numerical or categorical feedback information would likely be sufficient and could prompt faster learning. However, in each of these examples, the combination of both forms of feedback may be useful as well. An indication that you are both correct, as well as how correct, may serve to both increase the rate at which the categories or optimal choices are learned, and sharpen the representation we have of the features or expected value for each category or alternative.

1.6. Structure of Dissertation and Open Questions

In sum, while category learning can be explained through multiple theories, the base learning process remains the same: view a stimulus, make a choice, then receive feedback about the choice made. Whether the feedback is verbal (correct/incorrect), cognitive (green/red), or reward-based (coin/no coin), we are still able to develop strategies or rules that will maximize or performance. However, research has also shown that categorization performance can change based on variables such as the length of feedback delay (Dunn et al., 2012; Todd Maddox et al., 2003; Worthy et al., 2013), how much information is given in feedback (i.e. 'Incorrect, answer is B'; Maddox et al., 2008), and the structure of the stimuli (Ashby & Maddox, 2005, 2011). Regardless of which view of category learning is correct, or what manipulations are included, studies have shown that people are able to learn to categorize novel stimuli into complex categories through simple categorical feedback alone (i.e. 'yes', 'no'). Though, one lingering question is why more continuous numerical feedback has not been utilized in category learning. As mentioned prior, discrete feedback gives an immediate sense of whether a choice is correct or incorrect, but numerical feedback has the potential to give an idea of how right a choice is which would be useful in determining how similar a stimulus is to a particular category. In the following sections, a discussion about numerical feedback, as used in reinforcement learning paradigms, may give us an idea.

Overall, reward is a salient motivator of choice. We are able to learn from numerical reward values that are discrete or lie on a more continuous range, and our behavior can change based on how the rewards are presented to us. Further, we can utilize multiple strategies in order to maximize the overall reward we gain or determine what the optimal decisions to make are given our current goals. Through multiple rewarding instances, we can learn to make the most optimal decisions given our current situations and goals, and subsequently use the learned information to make predictions about the outcomes of our future choices. While models of reinforcement learning are able to incorporate both discrete and continuous numerical rewards, one question that remains is whether categorical feedback would lead to improved performance on certain reinforcement learning paradigms.

In both category and reinforcement learning, research has shown that humans are able to learn to make optimal decisions from a variety of feedback types. However, as was discussed, optimal behavior in each type of learning is likely dependent on the type of feedback given relative to the task at hand, and the amount of uncertainty in the value of the outcomes. For most category learning paradigms, categorical feedback is the preferable choice, whereas numerical feedback is more suited to reinforcement learning paradigms. More formally, we can create differing mechanistic assumptions of what promotes better learning under which situations. For situations where the outcomes are certain, we could assume that positive prediction errors are a major driver of learning (Glimcher, 2011; Greve et al., 2017), and since categorical and discrete numerical feedback are more likely to produce positive prediction errors, we could expect better learning. Conversely, for situations where the outcomes are more uncertain, we could assume that the minimization of error (coinciding with reward maximization) is a better driver of learning (e.g. Sohoglu & Davis, 2016) and may prompt us to explore all options more thoroughly as compared to outcomes that are more certain (Speekenbrink & Konstantinidis, 2015). As such, variable numerical feedback would likely be a more effective form of feedback.

Thus, in line with the above and the simulated results, we can create a few predictions. Should people be given more continuous numerical feedback on a category learning task, performance will likely not reach the level of a person who received categorical feedback instead. Similarly, performance should be equal between people who received either discrete categorical or discrete numerical feedback on the same category learning task. Additionally, if the frequency of stimuli observations, or changes in reward magnitude are present, people may be more likely to focus on numerical, rather than categorical, feedback. Finally, the combination of both categorical and numerical feedback would likely allow a unique learning experience in a category learning task where people would be able to take advantage of both whether or not their choice was correct as well as how correct they were. Though, it is unknown how performance given combined feedback would compare to the performance when given only either categorical or numerical feedback. It is possible that the feedback information will be additive in nature, but it is also equally probable that the feedback information will be differentially weighted.

Briefly, in Study 1, I will explore how numerical feedback compares to categorical feedback when either the most typical, or most atypical, stimuli give the largest rewards. Additionally, I will use the same paradigm to explore the difference between discrete, dichotomous, numerical feedback as it compares to categorical feedback. In Study 2, I will extend on the work in Study 1 with the most typical/atypical stimuli being rewarded the most. However, I will use a new training + transfer paradigm where different category clusters are rewarded by different degrees of magnitude based on typicality during training and then assess how each feedback type prepares participants for a transfer phase. In Study 3, I will use the same paradigm as used in Study 2 but modulate the frequency that each individual category is observed. Overall, through each of these studies, I will explore how reward modulates categorical knowledge and the potential situations where numerical reward information may promote better learning as compared to categorical feedback. In the following sections, I will further detail each of the above studies and detail what results I might expect based on the model simulations.

2. GENERAL METHODS

With each of the experiments detailed in this dissertation revolving around similar paradigms, there is a significant overlap in the methods used between each of the forthcoming studies. For conciseness, the general methods will be described here. Methods unique to a study will be described in each respective method section.

2.1. Participants

All participants were recruited for either in-laboratory or online participation from the undergraduate population of Texas A&M University via the SONA recruitment system. For their time, students received partial course completion credit for a course offered by the Department of Psychological and Brain Sciences. Each individual participant was allowed to participate in only one condition of one of the following studies. All participants were asked to review a Texas A&M University Institutional Review Board-Approved Informed Consent form prior to beginning each task and were notified that they could end the study at any time.

2.2. Study Materials and Procedure

Each participant completed a randomly assigned study task served via a computer-based format. Upon registration with the SONA system, in-laboratory participants selected a date of their choosing to appear in the lab, and online participants were given an anonymized URL that led to a web domain where the experiment was hosted. All online studies were self-hosted online using a server running JATOS software (Lange et al., 2015). Almost identical to in-laboratory procedures, upon following their given link, each online participant viewed the consent form and, instead of physically signing a consent form, the participants electronically noted their consent to take part in the study. Upon agreeing to take part, all participants were randomly assigned to experimental conditions in a programmatic fashion. Following consent and condition assignment, participants were asked to complete a demographic brief survey. All demographic questions were optional. In addition, for the online participant, they were additionally asked to detail specifics about their computer hardware (i.e. estimate screen size, OS used, etc.) and their immediate surroundings (i.e. noise level, alone or not, etc.) These additional questions were used to screen the data for possible issues due to the online format.

After completion of the demographic questionnaires and consent, participants were asked to complete the assigned category learning task. In-lab versions of the tasks were programmed in MATLAB, using the Psychophysics Toolbox extensions (Brainard, 1997), and displayed on a computer with a 1920x1080 screen resolution. The online versions of the tasks were programmed to appear, and function, identically to the lab versions using HTML, JavaScript, and the jsPsych JavaScript library (de Leeuw & Motz, 2016) which has been widely used in the creation of online psychological tasks. To attempt to control for the additional variable of screen size that likely varied among participants, each experiment was programmed to automatically scale all stimuli and text to be proportional to the screen size to create a consistent experience. Upon completion of the experimental tasks, participants were debriefed and, depending on the study format, participants were given a paper receipt of participation, or were asked to press a weblink that brought them back to the SONA site which would automatically assign them credit for participation.

2.3. Experimental Task Components

2.3.1. Category Structures

Each study will utilize category structures composed of lines, which differ in the two perceptual dimensions of length and orientation, which have been used in prior categorization work (Filoteo et al., 2010). In each of the studies, different category structures will be used to define category membership: a Conjunctive Rule (CJ) and an Information-Integration Rule (II). Under a CJ rule, category membership is based on a readily verbalizable rule-based structure where the optimal bounds are orthogonal to the axes of the stimulus dimensions (Ashby et al., 1998; Ashby & Maddox, 2005). Below, in Figure 2.1, are three versions of the rule-based category structures that will be used in the following studies. For example, in Figure 2a, the rule is: ‘short and steep lines are category 1, and all other lines are category 2’. For the category structures in Figure 2b and 2c, the rule is dependent on the category cluster.

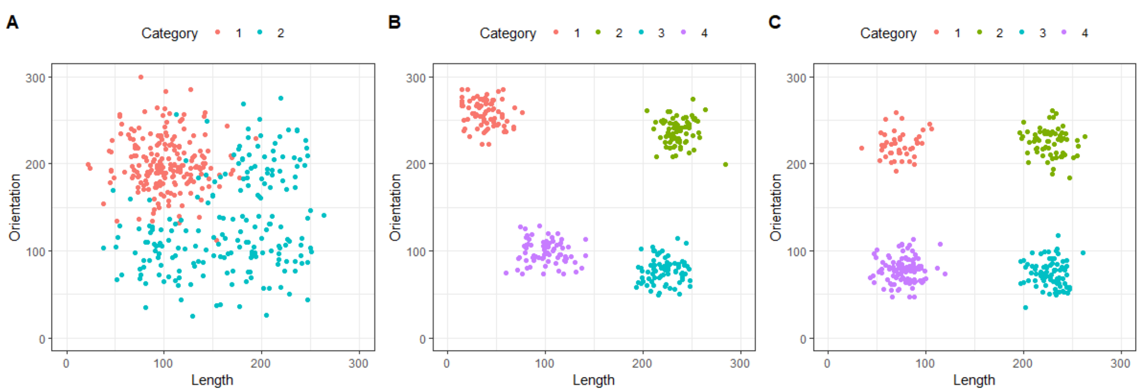

Figure 2.1 A.) Conjunctive rule structure. B.) Rule-based structure which spreads from the center in a radial pattern. C.) Rule-based structure where each category

cluster is equidistant from the center of the stimulus and each of the other category clusters.

In contrast, the optimal bound for the Information Integration (II) structure is oblique to the perceptual dimensions, making it difficult to verbalize an optimal rule. Thus, participants must integrate the features at some predecisional stage to become proficient (Ashby et al., 2003; Ashby & Maddox, 2005; Daniel & Pollmann, 2010). The information-integration category structure used in the following studies is depicted in Figure 2.2 below.

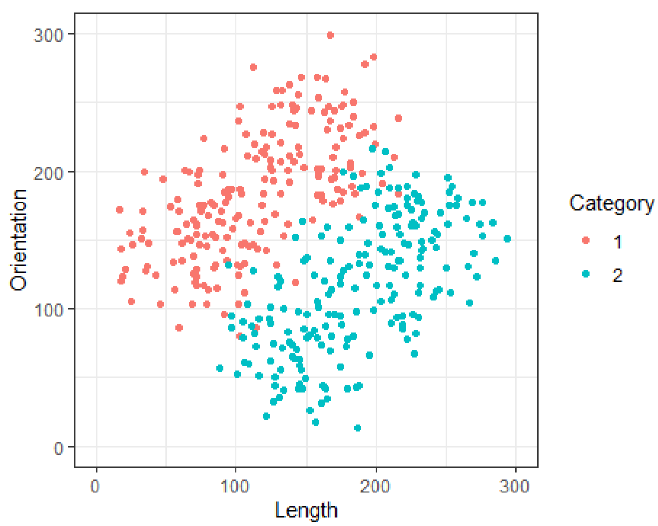

Figure 2.2 Information integration category structure where the decision bound between categories is oblique to the two features.

However, the goal of these studies is not to focus on differences between rule-based and information-integration category learning which have been thoroughly studied (Ashby et al., 2019; Carpenter et al., 2016; Donkin et al., 2015; Ell et al., 2006; Todd Maddox et al., 2004; Zaki & Kleinschmidt, 2014), but instead to simply use two commonly utilized category structures to examine how categorical information interacts with numerical information and determine the value that both categorical and numerical feedback have when categorizing stimuli.

2.3.2. Reward Structures

The reward structures used by each numerical feedback condition of each study will either be based on the probability of category membership as a function of the relative similarity of each stimulus to all the exemplars from each category, stimuli distance, or be uniformly distributed. For rewards based on the probability of category membership, the Generalized Context Model (GCM; Nosofsky, 1986) was used to calculate the classification probability for each stimulus by comparing its similarity to exemplars from its category against its similarity to exemplars from the other category. In the present work, the classification probabilities were calculated by assuming equal weighting to each stimulus dimension and using a sensitivity parameter (ϕ) value of .05, a value that was determined in a semi-arbitrary manner in order to generate reasonable probability gradients for the category structures used in our task. For rewards based on stimuli distance, the reward value for each stimulus is computed based on the rounded and scaled Euclidian distance between the current stimulus and a reference point (i.e. category boundary or center, category cluster, etc.). Finally, for uniform rewards, each stimulus will be assigned a reward value drawn from a uniform distribution of a predetermined range. Figure 2.3 below details a sample of each type of reward structure utilizing each of the previously described category structures.

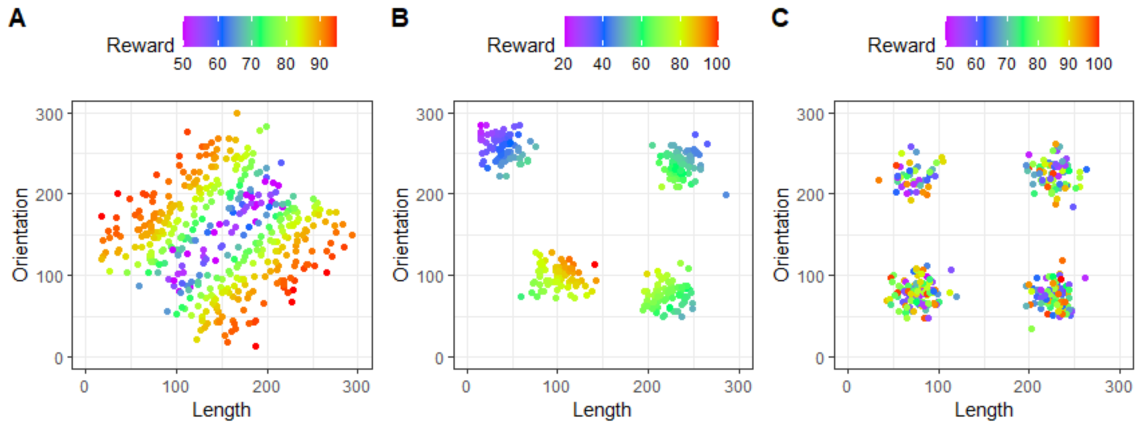

Figure 2.3 Examples of the reward structures used in each study. A.) Reward generated via the GCM; rewards are based on the probability of category membership. B.) Reward based on Euclidean distance from the center of the stimulus space scaled to a defined range. C.) Uniformly random rewards sampled from a defined range.

2.3.3. Trial Procedure and Feedback

In each experimental task, participants were shown a single line stimulus on screen where they were asked to either determine what category the line belongs to in categorical feedback conditions, or to use the line stimuli to determine which option will lead to the largest rewards in the numerical feedback conditions. In each of the following studies, the stimuli shown on screen persisted onscreen until selection. Participants were asked to make their selections using specific keyboard keys, or click on onscreen buttons for online participants, this was consistent throughout all trials. In the categorical feedback conditions, participants were given explicit feedback in the form of text strings of ‘CORRECT’ and ‘INCORRECT’ colored green and red respectively based on the outcome of their decision. For numerical-only feedback conditions, participants were shown the reward value they receive in white text. For the conditions with the combined categorical and numerical feedback, participants were shown the same feedback in the prior two conditions on screen simultaneously. All feedback was shown onscreen for a total of 2s before an intertrial interval of 0.5s where a screen with the text ‘Please wait…’ will be briefly shown on screen before the presentation of the next stimulus. Sample trials screens can be seen in Figure 2.4 below, and more detailed trial diagrams will be displayed within each study.

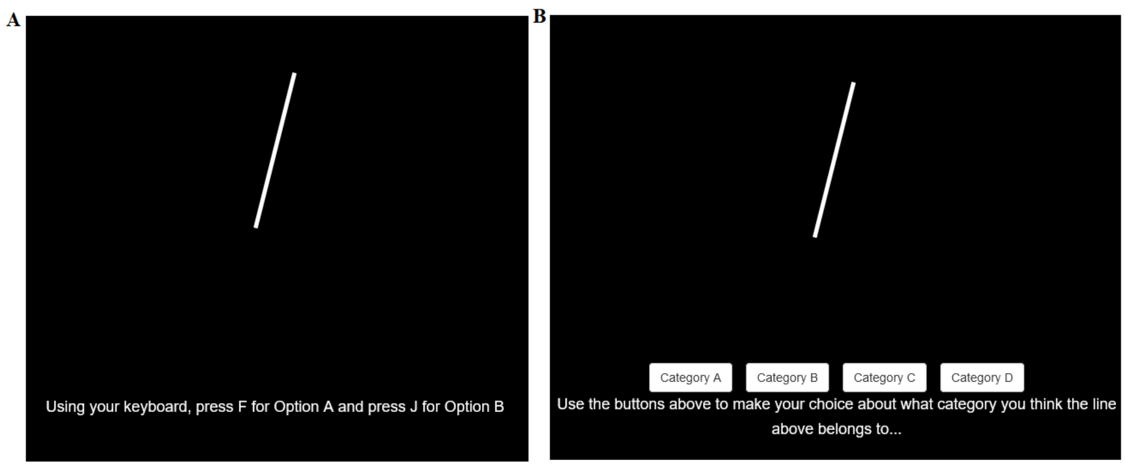

Figure 2.4 Depiction of a sample trial screen for the in-lab and online versions of the study. A.) In-lab version using keyboard responses. B.) Online version using on screen, clickable, button responses.

2.4. Learning Models

To gain a better understanding about the potential processes which occur when categorizing stimuli under each type of feedback, three computational models were derived from the Attention Learning Covering Map (ALCOVE; Kruschke, 1992) category learning model to both simulate and fit behavioral data. ALCOVE is a connectionist model that computes the probability that a stimulus belongs to a given category based on the integration of psychological dimensions and the attention allotted to each.

In ALCOVE, category membership is determined by the attention-weighted similarity between stored category representations (exemplars) and the to-be-classified stimulus. In the covering map version of ALCOVE, category representations are a set of nodes that are distributed evenly across the psychological space spanned by a category learning task, instead of exemplars that have been encountered before. When a new stimulus is observed, the similarity between the stimulus and each node is computed and these similarity values are aggregated across all nodes to determine the probability that the stimulus belongs in a given category (See Kruschke, 1992 for a full description of the model).

In the following studies, I employ three variants of the covering map version of ALCOVE with a few important deviations. Most notably, these model variants do not learn the attentional weights of each psychological dimension; rather, the models use a free parameter to estimate the attention that is given to each psychological dimension since attention learning is beyond the scope of the current study. These models use ALCOVE's equations to compute the similarity between a current stimulus and each of the models' hidden nodes. Each hidden node learns a weight that describes the strength with which its area of psychological space is associated with one category or another. Here, for the covering map, a 21 X 21 grid of 441 hidden nodes evenly spread across the two-dimensional stimulus space is used.

2.4.1. Category-Learner Model

The category-learner model is identical to ALCOVE (Kruschke, 1992) in the computation of the activation values for each j^{th} node as shown in Equation 1:

$$A_j = exp\left[-s\left(\sum_i \alpha_i \left|h_{j,i} - x_{i,t}\right|^r\right)^{q/r}\right] \tag{1}$$

The Activation values (A) for each node (j) are computed on each trial (t) for each psychological dimension (h). In the present model, we used only the psychological dimensions of line length and orientation for $h_{1:2,j}$. For example, on each trial, the absolute value of the error between the nodes' length value ($h_{i,1}$) and the actual observed length ($x_{i,t}$) is multiplied by the attentional weight (α_i) free parameter given to stimulus length. This process is completed again for orientation and summed together before being modified by a specificity constant ($s;\ s > 0$). The exponent of the resultant becomes the activation value for the j^{th} node. Nodes that are more similar to the current stimulus that must be classified will have higher activation than nodes dissimilar to the current stimulus. Additionally, $\alpha_i$values of 0 or 1 indicate exclusive attention given to one dimension. The similarity metric and gradient values, r and q respectively, are both set to 1 in the current model.

The category-learner model slightly departs from the original ALCOVE model in the calculation of the activation values of the output nodes (A_k^{Out}):

$$A_k^{Out} = \sum_j \frac{A_j}{\sum A_j} \cdot w_{j,k} \tag{2}$$

where for each response (k), a vector containing the weights between each jth node and each k response node ($w_{j,k}$) is multiplied by the normalized activation values (A_j) for each node that result from dividing each activation value by the sum of activation across all nodes. The normalization of activation values is a slight modification from the original instantiation of ALCOVE (Kruschke, 1992), and is done here to be consistent with the reward-learner model we present below.

The expected response node values ($w_{j,k}$) for each node in the *chosen* response are modified on every trial according to Equation 3 below:

$$\Delta w_{j,k} = d(\Psi_t - w_{j,k})\, A_j \tag{3}$$

$$where\ \ \Psi_t = C_t = \begin{cases} Correct & 1 \\ Incorrect & 0 \end{cases}$$

where d is a learning rate parameter, and Ψ_t is a teacher signal that will take on values based on the feedback given in each model. In the category-learner, Ψ_t will equal C_t, a binary value, which represents categorical feedback. Correct categorizations result in a value of 1, and 0 for incorrect categorizations. Effectively, this means that whenever a category learning trial is correct, the $\Psi_t\ - w_{j,k}$ error computed in Equation 3 will always be positive and result in an improved association weights between nodes and the chosen category. Conversely, on an incorrect trial, the error value will always be negative and should result in a poorer association between category and hidden layer nodes.

2.4.2. Reward-Learner Model

The reward-learner model was designed to learn from continuously valued reward information only. This model is functionally identical to the category-learner model; however, categorical feedback, used in Equation 3, is no longer used. In place of the binary categorical feedback (C_t), we used scaled, continuous, reward information (R_t;

$r \in (0{,}100)$) as the teacher signal Ψ_t. The calculation of the scaled reward value can be seen in Equation 4:

$$\Psi_t = R_t = \frac{r_t - r_{min}}{r_{max} - r_{min}} \tag{4}$$

where r_t is the actual reward value that is observed on any particular trial. In Equation 4, to calculate the scaled reward value (R_t), the minimum and maximum reward values observed across all trials are recorded, denoted by r_{min} and r_{max} respectively, and used in normalizing the current observed reward as shown in Equation 4. This function ensures that the observed reward values are scaled to a value between 0 and 1, but also constrained to the range of known outcomes. Once the reward values of 0 and 100 are observed, the scaled reward value calculation reduces to $r_t/100$.

2.4.3. CatRwd-Learner

Another variant we used was a hybrid catrwd-learner model which simply weights how much information from both types of feedback is used in updating. The model itself is structurally identical to the reward-learner model aside from a modification of the teacher signal computation in Equation 4. This modification is detailed in Equation 5 below:

$$\Psi_t = qR_t + (1 - q)C_t \tag{5}$$

where $q \in (0{,}1)$ is a free parameter which represents the weight given to the scaled reward value (R_t: as calculated in Eq. 4), with (1-q) representing the weight to categorical feedback information (C_t). Importantly, depending on the value of the q parameter, the model will make increasing similar predictions to either the category- or reward-learner model as q approaches 0 or 1, respectively.

2.4.4. ALCOVE-Decay Model Variants

In the prior variants of ALCOVE, I've discussed, the prediction error calculation in Equation 3 is incredibly similar to the Delta rule model (Don et al., 2019; Rescorla & Wagner, 1972). In ALCOVE, the exemplar weights are updated via a learning rate and activation modulated prediction error between an expected outcome and the actual outcome received, or in other terms, the exemplar weights are adjusted proportionally to the error gradient to minimize the future error (Kruschke, 1992). As shown in Equation 6 below, in the Delta rule, the same error-based updating process occurs where the expected value for a chosen alternative is updated based on a learning rate modulated prediction error between expected reward and reward obtained.

$$\Delta EV_k = \alpha \cdot (r_t - EV_{k,t}) \tag{6}$$

In this equation, *EV* is the expected value for *k* alternative, α is a learning rate (α; $\alpha \in (0,1)$) and *r* is the outcome received. Like mentioned prior, this equation is similar to Equation 3 in ALCOVE minus the use of the activation values. Remember that these activations values ensure that largest updates to exemplar weights are given to the exemplars that are most similar to the observed stimulus. The Delta rule model assumes that each alternative has an expected value that is learned over time. As alternatives are selected and feedback is received, the expected value for the chosen alternative is updated based on a prediction error. Concisely, the Delta rule model learns, through multiple iterations, what the average reward is for each alternative.

In comparison, the Decay rule model (Erev & Barron, 2005; Yechiam & Busemeyer, 2005), which learns in a similar fashion, associates each given alternative with their learned cumulative value. Rather than learning via prediction error, the Decay model (Equation 7) assumes that the expected value of all alternatives decay over time and that the raw reward values are more important than the prediction error.

$$EV = EV \cdot \alpha \tag{7}$$

$$EV_{k,t+1} = EV_{k,t} + r_t$$

In Equation 7 above, EV is the expected value for the chosen alternative, α is a decay factor ($\alpha \in (0,1)$), and *r* is the observed outcome. On each trial *t*, the expected values for each alternative are decayed by a factor of α. Then, based on an agent's decision, the raw reward value *r* directly increases the expected value of the chosen alternative.

According to recent research by Don and colleagues (2019), both the Delta and Decay models can make opposing predictions when the frequency, and magnitude, of rewarding outcomes differs between alternatives. As such, since the ALCOVE model utilizes a similar updating method as the delta model, a decay variant to each of the previously described category learning models was created. Exemplar weights for each ALCOVE-Decay variant are computed based on Equation 8 below and will be substituted for Equation 3 above in the previous models.

$$w = w \cdot d \tag{8}$$

$$w_{j,k} = w_{j,k} + \Psi_t \cdot\ A_j$$

Similar to the Decay model, at the beginning of each trial, the exemplar weights (w) are assumed to decay by a factor of d. Upon determining which category the observed stimulus belongs to, the Ψ_t values, based on what feedback type the participant is given, is modified by the activation values (A_j) and directly added to the exemplar weights ($w_{j,k}$) for the chosen category k. As such, this model will learn the cumulative outcomes for each category.

2.4.5. Predicted Choice in Models

The predicted response (K) for any given trial and model is denoted by a probability value computed using Equation 9:

$$Pr(K) = exp(\phi A_{\mathrm{K}}^{Out}) / \sum_k exp(\phi A_k^{Out}) \tag{9}$$

where (ϕ; $\phi \in (0,5)$) is an inverse temperature parameter and *k* represents individual response options. Low inverse temperature parameter values typically lead to random decision-making, whereas higher parameter values indicate more consistent responses for the most probable category predicted by the model.

2.4.6. Simulation Method and Predictions

For each of the following studies, the potential effects and hypotheses were be based on data simulated from the prior models. In each study condition, 500 simulated agents were created to represent the simulated behavior of a theoretical participant with a unique set of free parameters. For each agent, the parameter representing attention given to the psychological dimensions of length and orientation (α) will be set to 0.5 across all simulations to reflect equal attention to both dimensions; the specificity parameter (*s*) will be set to 0.05; the learning rate (*d*) and feedback weighting (*q*: used in catrwd-learner variants) parameters will be independently drawn from a uniform distribution, *U*(0,1); the inverse temperature parameter (ϕ), will be drawn from *U*(0,5). For consistency in comparisons between model predictions, each study was simulated utilizing the same 500 agents for each combination of task condition and computational model. To accommodate the randomized stimulus presentation within the task, each simulated agent naively completed each task 10 times and had their data aggregated at the trial level.

3. STUDY 1

3.1. Introduction

"Hot or Cold?" "Healthy or Unhealthy?" Categorization enables us to make judgments about the value of different actions as well as inferences about future events, such as whether or not a jacket is needed based on the weather (Markman & Ross, 2003). The ability to categorize objects and events allows us to generalize much of our knowledge about the world and reduce it to manageable proportions (Barsalou, Lawrence, 1983; Rosch, 1978). Similarly, learning from rewarding outcomes allows us to understand the potential values of any given alternatives so we can make decisions about how to best proceed (Berridge, 2000; Berridge & Robinson, 1998; O'Doherty et al., 2007). Through subjectively rewarding events, we can determine which restaurants are good to eat at, whether a test score was passing, or what driving speeds won't result in a speeding ticket. It is possible that category and reward learning are two sides of the same coin: on one side we discretely categorize an individual stimulus/event, on the other we categorize the outcomes of a given stimulus/event.

Category learning paradigms make it possible to study how people acquire new categories in a laboratory setting. Typical category learning paradigms have participants learn to classify stimuli into one or more categories, while receiving feedback about whether their classifications are correct or incorrect (Ashby & Maddox, 2005; Markman & Ross, 2003; Medin & Schaffer, 1978). Thus, this type of learning depends on a process of observation, choice, and feedback in order to classify novel stimuli into a discrete number of categories (Ashby & Maddox, 2011; Nosofsky et al., 2019; Smith, 2014). However, category learning research has primarily employed feedback in the form of either discrete outcomes such as ‘Correct/Incorrect’, or similar discrete numerical values (i.e. 0 or 1 point).

In contrast to category learning paradigms, reinforcement learning paradigms often utilize either discrete or continuous numerical rewards. Reinforcement learning tasks are characterized by choosing between a discrete number of choices and iteratively learning which choices are more valuable based on feedback (Daw et al., 2006; Erev & Barron, 2005; Frank & Claus, 2006; Kool et al., 2017; Niv, 2009). While category and reinforcement learning paradigms differ in their framing of ways to influence learning (Radulescu et al., 2019), reward feedback is critical to changing future behaviors in both category learning (e.g. Abohamza et al., 2019; Daniel & Pollmann, 2010; Moustafa et al., 2015) and reinforcement learning (e.g. Montague et al., 2006; Sutton & Barto, 2018; Thorndike, 1927) paradigms. Collectively, both category and reinforcement learning are reward-dependent processes that shape how individuals learn new information and make future choices. Despite their similarities, the type of feedback that is typically used to examine these paradigms has largely differed; category learning has relied on categorical feedback, but reinforcement learning has relied on both discrete and continuous numerical feedback. Consequently, the role of categorical feedback, as compared to continuous numerical feedback, on category learning outcomes remains a relatively under-studied area of research.

Although category learning paradigms have favored categorical feedback, rather than continuous feedback that encompasses a more continuous range of values, such dichotomous outcomes do not necessarily mirror the graded feedback one may receive in a real-life situation. As an illustration, when making decisions about what to wear based on the weather, we can make a prediction about the temperature, put on some clothes, and walk outside to where we will be able to gauge the efficacy of our prediction. If our choice of clothes was relatively congruent with the outside temperature, we would most likely opt to wear those clothes again in similar weather. Ergo, predictions that are congruent with the category representation will be reinforced (Ross, 2000). However, imagine if we predict that the weather is warm and wear light clothing. If the weather turns out to be a little cool, we might feel a bit of discomfort, but if it is frigid outside, we may find ourselves in a potentially dangerous situation. To prevent a similar occurrence in the future, we would have to understand that our prediction was incorrect by a certain degree and update our representations of weather types accordingly. Thus, category learning tasks may benefit from the inclusion of feedback that falls on a continuous scale as it may be able to better confer a degree of correctness in the response to each decision.

It is, however, currently unknown whether categorical, continuous, or a combination of both types of feedback would promote better category learning. Learning from continuous feedback, as defined by a variable range of numerical values or reward-based feedback, such as the magnitude of discomfort felt due to the weather or a low amount of payment received for work, is often attributed to the amount of surprise one receives from the outcome based on prior expectation (e.g. Schultz, 2016, 2017), or the prediction error. Decisions that result in positive prediction errors, where the outcome received was greater than expected, are more likely to be made again as they may become predictive of future rewards. Similarly, repeating the same decisions is less likely when the outcomes, or lack thereof, fall below expectation, leading to negative prediction errors and a reduced likelihood of choosing the same action again. Thus, continuous feedback may facilitate learning by giving participants a sense of how right their choice was. In other words, discrete feedback provides gross-level information, such as “Pass/Fail”, but does not specify how close or far a behavior is from the correct response. Continuous feedback, on the other hand, provides more fine-grained information. For example, a “Fail” score of 69% versus 19% on a test indicates vastly different degrees of revision that would be needed to achieve a passing score. However, because continuous feedback provides a broader range of information, it may take multiple observations before a reliable expectation of reward is learned because ‘correct’ choices are more variable and therefore more ambiguous.

Conversely, categorical feedback immediately gives an expectation, in terms of correct or incorrect, but no information is given regarding how correct the response was. Additionally, recent work suggests that prediction error magnitude may have an impact on the rate at which categories are learned (Lohse et al., 2020). From this perspective, categorical feedback may facilitate better learning than continuous feedback because initial prediction errors will tend to be larger. For example, if categorical feedback is enumerated as ‘1’ for correct and ‘0’ for incorrect per Ashby et al. (2011), continuous feedback is scaled from 0 to 1, and predicted probabilities of category membership also range from 0 to 1, then categorical feedback will lead to larger prediction errors than continuous/variable feedback. Combining both types of feedback may provide both an immediate expectation and an indication of how right the decision was. A combined approach therefore provides information about the magnitude of discomfort or incorrectness felt while simultaneously receiving information about the outcome of your decision in categorical terms. Thus, it is possible that receiving both forms of feedback may promote better performance on category learning tasks than categorical or continuous feedback alone.

Additionally, category learning in particular may benefit from the inclusion of continuous feedback when discriminating stimuli based on their representativeness. Often, in category learning tasks, stimulus classification is defined by a perceptual boundary that distinguishes two categories, such as Category ‘A’ and Category ‘B’. Stimuli that are easier to classify are defined as being more representative of that category; these stimuli are usually farther away from the perceptual boundary. In contrast, stimuli that are more difficult to classify are less representative of that category and are typically closer to the perceptual boundary that divides category members. It is possible that continuous feedback may differentially affect learning rates for more or less representative stimuli.

3.1.1. Simulation

To explore these potential impacts of the differing forms of feedback on category learning, simulations were conducted utilizing CJ and II category structures detailed in Figure 2.1-2, the ES and DI reward structures detailed in Figure 2.3, and four feedback types: Categorical (Cat), Continuous Numerical (Rwd), Discrete Numerical (Dis), and a

Hybrid Cat-Rwd (CatRwd) feedback which displayed continuous numerical feedback alongside categorical feedback. The combination of the aforementioned structures and feedback types resulted in 10 total simulation conditions with 8 conditions having some form of reward-based feedback and 2 with only categorical feedback due to the lack of reward structure in these conditions.

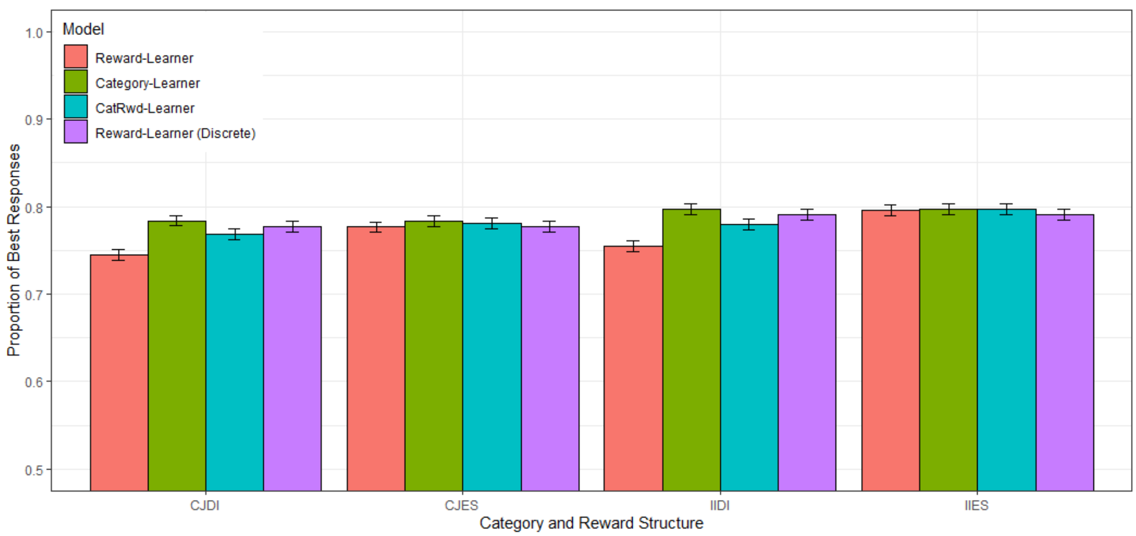

Figure 3.1 Simulation results for each feedback type and reward/category structure condition. Higher values denote a greater degree of best responses.

Ultimately, the main difference between discrete and continuous feedback is the type of information that is given in response to a choice. Theoretically, in addition to categorical feedback, both continuous and combined feedback have the potential to promote the learning of an underlying category structure. However, as we described above, the comparative efficacy of each feedback type is under-researched. To test how each form of feedback impacts learning, we use a computational modeling framework to simulate learning based on the four forms of feedback: categorical only, variable numerical rewards, discrete numerical rewards, and combined categorical and variable reward feedback. In the simulations output above in Figure 3.1, we show that categorical, and discrete numerical, feedback does indeed lead to larger magnitude prediction errors than continuous feedback, and this leads to better predicted learning. The simulations also predict that when only variable numerical feedback is given, optimal learning depends on reward magnitude and category representativeness. Specifically, learning is optimized when high-magnitude rewards are given for correctly classifying stimuli that are highly representative of a category compared to less representative stimuli.

In the following sections we first present a basic overview of the task and the category and reward structures. We then detail the procedures of this specific paradigm before we finally report results from an experiment with human participants to evaluate their behavior as compared to our model-based predictions.

3.2. Method

3.2.1. Participants

Students from Texas A&M University participated in the study in partial fulfillment of an Introductory Psychology course requirement. We created a twelve-group experimental design, identical to the groups described in the simulations above, which consisted of 10 groups who were given reward information, and 2 groups that were only given categorical feedback.

Our goal was to collect data from 80 participants for each condition. This provides over 80% power to detect even small effects (η_p^2 = .02). We also wanted relatively large sample sizes to avoid having to draw conclusions from insufficient data which could lead to Type 1 errors. In total, we recruited 760 participants and randomly assigned each to one of the following conditions. For reward feedback, the sample sizes per condition were as follows: 78 in the CJDI condition, 77 in the CJES condition, 79 in the IIDI condition, 78 in the IIES condition. For catrwd feedback, the sample sizes were: 84 in the CJDI condition, 79 in the CJES condition, 80 in the IIDI condition, 78 in the IIES condition. In the discrete reward feedback conditions: 71 in DisCJ and 73 in DisII. Finally, the sample sizes for the categorical feedback conditions were: 62 in the CJ condition, 64 for the II condition. These two conditions were run as comparison conditions at a later date than the other eight conditions and had slightly smaller sample sizes due to being run near the end of an academic term.

3.2.2. Task and Structure

3.2.2.1. Experimental Task

As described in the General Method section, and prior simulations, participants were shown one of 400 unique line stimuli on each trial in a randomized order. The line varied in both its length and orientation, and participants could make one of two responses to indicate what category they thought the line was in. Depending on what between-subjects condition participants were in, they received one of three types of feedback after making each choice: continuous numerical feedback (Rwd) that varied between 0 and 100 points, discrete numerical feedback that delivered either 0 or 100 points, categorical feedback (Cat), or both types of feedback (CatRwd).

For the category and catrwd feedback conditions, participants categorized the lines into two categories of A or B. They were told whether they were correct for each choice, and participants in the catrwd condition also received either 0 points for an incorrect classification, or between 50-100 points for a correct classification. For the numerical feedback conditions, participants were not told anything about the stimuli belonging to categories. Instead, they were told that on each trial they would pick from either option 1 or 2, and that the line on the screen would aid them in predicting which option would result in a reward, similar to the procedure detailed in Kahnt, Park, Burke, & Tobler (2012). To mimic real-life scenarios where an underlying category structure must be learned from non-categorical feedback, participants would then receive either 0 points or between 50-100 points; participants were not told whether they were 'correct' or not, but they could draw this inference from whether they received points or not on each trial. Similarly, the discrete numerical feedback participants were simply given 0 points for incorrect feedback, and 100 points for correct feedback in an effort to emulate categorical feedback.

Thus, the major distinction between the conditions with numerical feedback, and the conditions which included categorical feedback, was that participants were told to use the line as a reference for which option will be more rewarding as compared to being explicitly told to categorize the line stimuli. Ultimately, the tasks only differed in framing (Radulescu et al., 2019) with the tasks with categorical feedback following typical category learning procedures, and the reward-only task following the procedure of a basic reinforcement learning or decision-making task.

3.2.2.2. Category and Reward Structures

As described in by the simulations, two category structures were used in the experimental tasks: a conjunctive rule and an information-integration rule. In both structures, each category is close in proximity in terms of location in the stimulus space. This means that the boundary between categories may seem fuzzy to participants, and the categories may be less deterministic as a result. A depiction of each category structure can be seen in Figure 3.2a below.

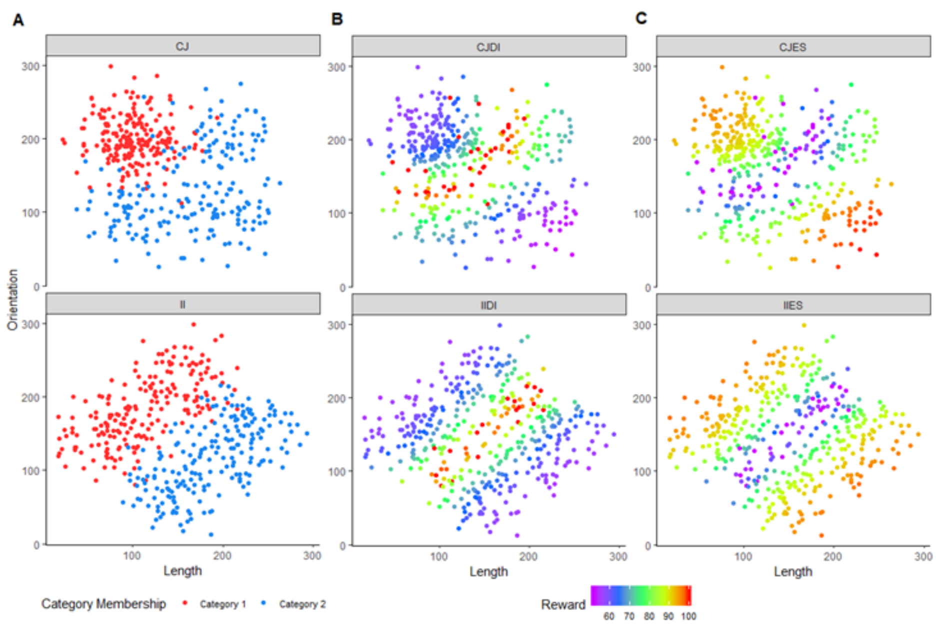

Figure 3.2 Plots of stimuli used in each condition by the psychological dimensions of length and orientation. Each individual dot is a single stimulus. A.) The II and CJ category structures used in this task. B.) The DI reward structures where the largest rewards are at the bounds of the category. C.) The ES reward structures where the largest rewards are for the most typical stimuli.

As mentioned in the general method section, the Generalized Context Model (GCM; Nosofsky, 1986) is used to calculate the classification probability for each stimulus. In the ES conditions, the points were a direct function of the classification probability of the stimulus for the correct category from the GCM model. The probability that the current stimulus is in the correct category, according to the model, was simply multiplied by 100 points. In the DI conditions, points given for correct classification of response/category k were given by: 100 * (1-Probability(k)). All reward values were scaled to the range of 50-100. Thus, in the ES condition, more points were given for correctly classifying the easier, high probability stimuli, whereas in the DI condition more points were given for correctly classifying the more difficult, low probability stimuli. The category structures with rewards associated for correct classifications of each stimulus are shown in Figure 3b-c above.

3.2.3. Procedure

Each participant completed an experimental task on a computer in a laboratory environment after signing an Institutional Review Board-approved consent form. The instructions and stimuli were presented onscreen using Matlab and PsychToolbox version 2.54 (e.g. Brainard, 1997). Participants were told that they would be shown images on a screen, which consisted of white lines that varied in length and orientation. Depending on the assigned feedback type, there were slight differences in the task as detailed below and in Figure 3.3.

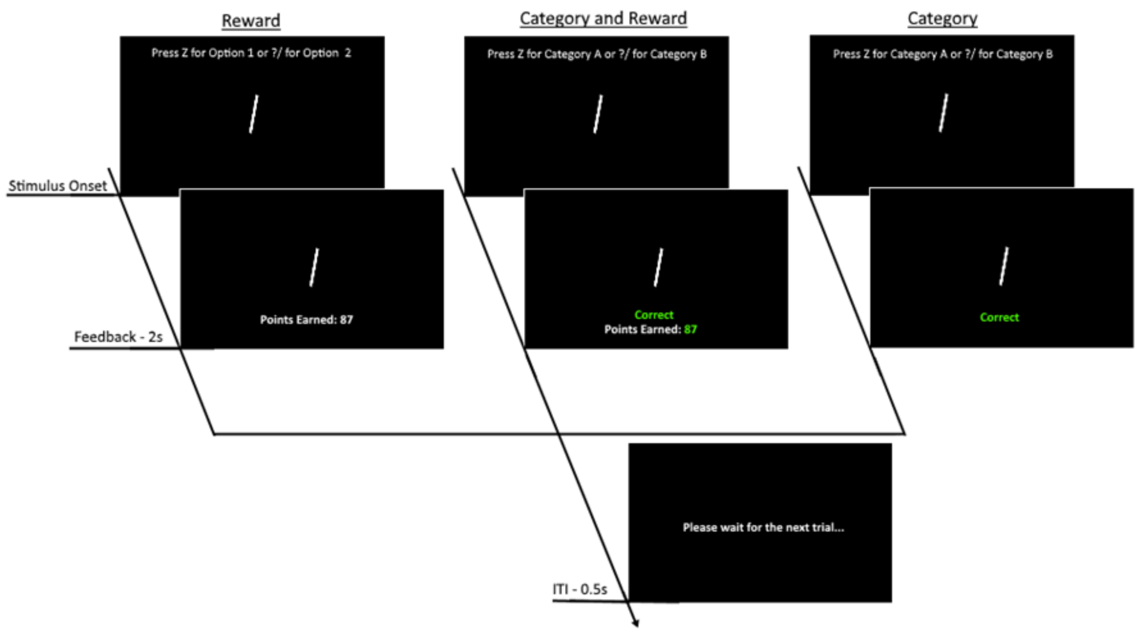

Figure 3.3 Trial diagram for each feedback type in Study 1.

Categorical feedback participants were asked to categorize each line on the screen into either Category 1 or 2 and were asked to respond with either the ‘z’ and ‘/?’ keyboard keys. Upon selection, participants were explicitly told if they were correct or not with text strings of ‘CORRECT’ and ‘INCORRECT’ colored green and red respectively before the next trial began after a 2s delay for feedback time.

Reward feedback participants were told that the line on the screen would aid them in predicting which of the two options would be the most rewarding on that trial using the same letter keys used when given categorical feedback. Upon choosing an option, the screen would display 0 or 50-100 points and continue to the next trial. Importantly, no mention of categories or classification was present in this condition. Participants given catrwd feedback were shown experiment screens identical to the categorical feedback, but reward information was also given. Like the categorical feedback participants, they were asked to categorize each line on the screen into either Category 1 or 2. Upon selection, participants would be shown both the categorical and reward feedback simultaneously.

The experiment consisted of four 100-trial blocks. A break screen separated each 100-trial block in an effort to reduce fatigue. These screens would display progress information and the number of trials that have been completed. Reward feedback participants were only told that they had completed X number of trials out of 400, and to keep trying to earn as many points as possible. In the category and catrwd feedback groups, participants were told what percentage of the previous 100 trials were correctly categorized. The primary dependent variable for both experiments was the proportion of optimal responses made across trial blocks and across all trials.

3.3. Results

3.3.1. Behavioral Results

3.3.1.1. Learning Over Time

Figure 3.4 shows the overall average proportion of optimal responses and the average proportion of optimal responses per 100-trial block for each condition. To compare the rates at which participants learned the task across conditions, we fit a mixed effects logistic regression model using R’s brms package (Bürkner, 2017). The full model predicted the average proportion of optimal responses from the three feedback types that included reward (CatRwd and Variable/Discrete Reward Feedback), the two category structures (II and CJ), the two reward structures (ES and DI), and block number (four 100-trial blocks centered). Block number was also used as a random slope and a random intercept was given for each participant.

For each of the analyses below, the parameter estimates are reported for each fixed effect predictor from the model with a 95% credible interval. This means that, given our data, the true value of the parameter is encompassed within an interval of the posterior distribution with a 0.95 probability (Nalborczyk et al., 2019). In our current analyses, a parameter value of 0 would indicate that a factor had no meaningful effect in the model. As such, we interpret any credible interval that contains 0 as evidence that the true value of the parameter has at least some probable value of not impacting the model.

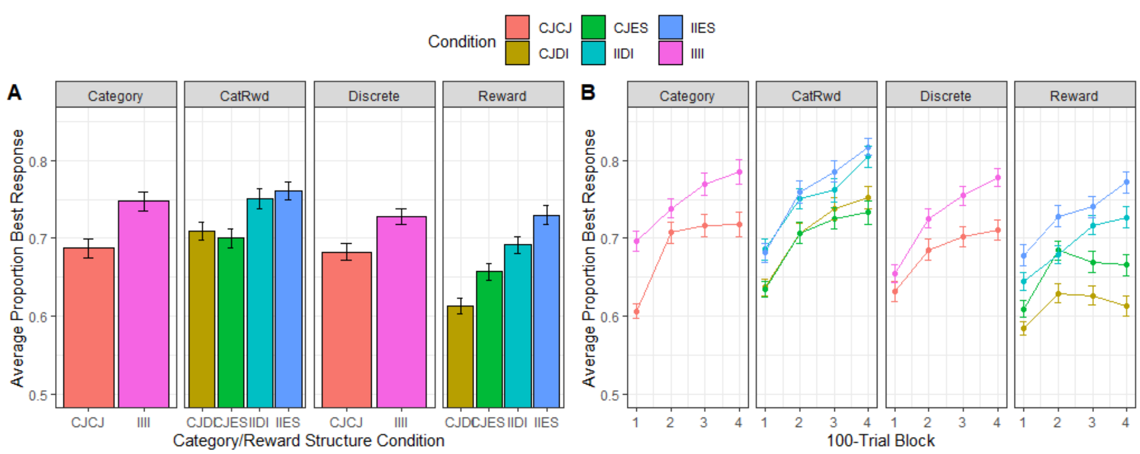

Figure 3.4 Plot of the proportion of best responses across (A) all 400 trials and (B) 100-trial blocks for each feedback type and category/reward structure condition. Error bars represent the standard deviation about the mean.

Figure 3.4 plots the average proportion of optimal choices for participants in each condition. For reference, computed proportion of optimal choices for the II category structure is ~0.7997 and ~0.7895 for the CJ structure based on the probability that an individual stimulus belongs to one category over another. Similar to our simulations, participants in the reference group (catrwd feedback) conditions made more optimal responses than participants in the variable reward feedback alone conditions (-0.09, [-0.13, -0.07]). Also, in line with the trend detailed by the simulations, we observed a main effect of category structure (0.04, [0.01, 0.07]) detailing poorer performance in conditions using the conjunctive rule category structure. While we did not specifically analyze the simulated performance by trial block, we found a main effect of block (0.04, [0.03, 0.05]) suggesting that learning did indeed occur across trial blocks.

Additionally, we observed a Feedback Type X Category Structure X Block interaction (0.02, [0.01, 0.04]), a Feedback Type X Block interaction (-0.03, [-0.04, -0.02]), and an interaction between feedback type and reward structure (0.05, [0.01, 0.09]). Plots of the interactions can be seen in Figure 3.4 below. The Feedback Type X Block interaction is due to a difference in learning slope between the catrwd and reward feedback types in which learning improves less across blocks for participants with reward feedback versus catrwd feedback (Figure 3.5a).

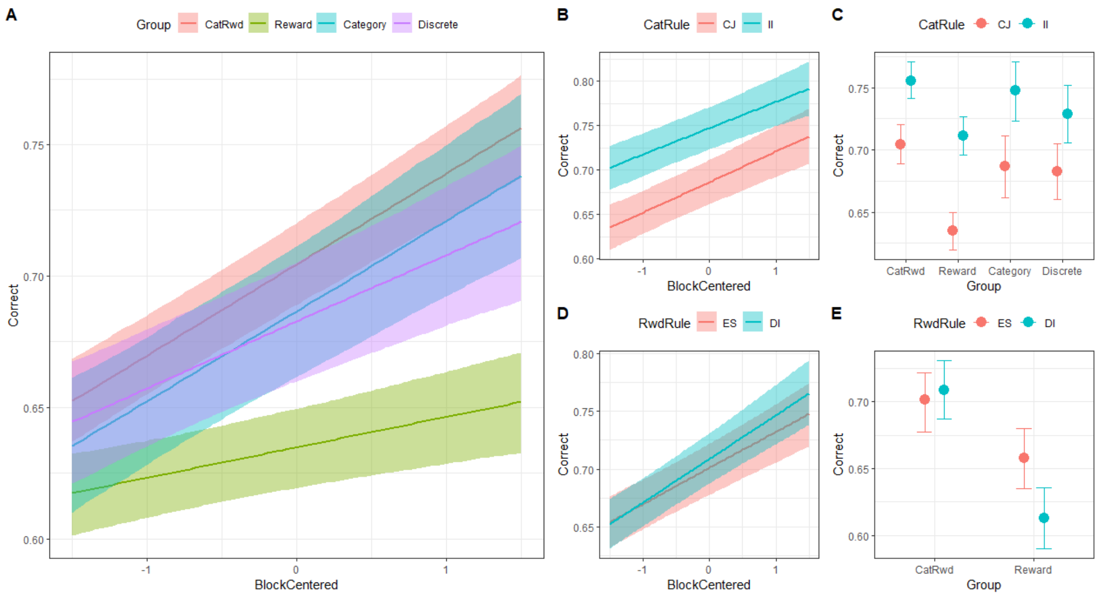

Figure 3.5 Plot detailing the differences in learnings slopes and mean proportion correct by feedback type and condition. A.) Learning slopes across time for each feedback type overall. B.) Learning slopes for category structures overall. Mean correct by category structure and feedback type. D.) Learning slopes for each reward structure overall. E.) Mean correct by reward structure and feedback type. Error bars and bands represent the 95% credible interval.

The Reward Structure X Feedback Type interaction is of particular interest. In Figure 3.5e above, there is very little difference in performance between reward structures when given catrwd feedback, but an apparent advantage for ES conditions when given reward feedback. To further explore each of these interactions, we regressed the above full model, with the feedback type parameter excluded, on the catrwd and reward feedback data independently.

For the catrwd feedback data, we found only the main effects of category structure (0.04, [0.01, 0.07]) and block (0.04, [0.03, 0.05]). The performance differences between category structures were expected based on the simulation results. For the reward feedback conditions, we found main effects of category (0.08, [0.05, 0.11]) and reward (0.04, [0.01, 0.08]) structures, and a Category Structure X Block interaction (0.02, [0.01, 0.03]). In Figure 3.5b above, we can see that halfway through the task, performance begins to decrease in the CJ conditions whereas performance continues to increase to the expected levels in the II conditions. Since this behavior is not seen in the CatRwdCJ conditions, the lack of categorical information is a possible cause for this difference. These results show that the Reward Structure X Feedback Type and three-way Feedback Type X Category Structure X Block interactions observed in the full model were due to the reward structure effect and Category X Block interaction which were present in only the reward feedback data.

To determine if catrwd and reward feedback types differed from categorical feedback alone, we ran an additional model that collapsed the catrwd and reward feedback data across reward structures. This model predicted the average proportion of optimal responses from each type of feedback (Category, CatRwd, Reward), category structure, and block number. As with the full model above, block and participant number were used as a random slope and intercept respectively. Setting categorical feedback as the 'Feedback Type' reference group, there were main effects of variable reward feedback (-0.05, [-0.08, -0.02]), and block (0.03, [0.03, 0.04]), but not catrwd feedback (0.02, [-0.01, 0.04]) or discrete reward feedback (-.005, [-.036, .029]). These differences can be visually observed in Figure 3.5c. These results suggest that categorical and catrwd feedback elicit similar performance and are somewhat surprising based on our simulation predictions. The simulations predicted that the catrwd feedback would show poorer performance than categorical feedback on the assumption that both forms of feedback were equally weighted. Further, these results provide evidence that our hypothesis, where we said that the discretized reward values and categorical information should be similar based on prediction error calculations, may be correct.

Finally, we ran two additional models to directly compare categorical feedback data to the variable reward feedback data. To do this, we split the reward feedback data by reward structure (RwdES and RwdDI) and compared each set to the total category feedback data. This model simply predicted the optimal response from both types of feedback and gave a random intercept to each participant. Similar to previous comparisons of the two feedback types, the RwdDI participants showed poorer performance overall when compared to category feedback participants (-0.07, [-0.10, -0.04]). Interestingly, however, the difference in performance between RwdES and categorical feedback participants was lower, but not significantly lower (-0.02, [-0.05, 0.00]). This suggests that when the easiest stimuli are the most rewarding, learning from continuously valued rewards is only marginally worse than learning from categorical feedback alone, and that the overall difference between feedback types stems from the poor performance in the DI conditions where the largest rewards were given to the stimuli that were the hardest to classify.

For exploratory purposes, since there was only a marginal difference in accuracy between the category feedback conditions and the RwdES conditions, we also compared the RwdES data to both the CatRwdES and CatRwdDI data using the same model above. Interestingly, the RwdES data differed from both CatRwdES (-0.04, [-0.06, -0.01]) and CatRwdDI (-0.04, [-0.06, -0.01]). Since the differences between RwdES and the CatRwd groups were more pronounced than the difference between the RwdES and the categorical feedback group, this would imply that learning performance is improved to some extent when reward information is included alongside categorical information. However, we also used the above model to compare the CatRwdES and CatRwdDI data to the category feedback data. As alluded by a main effect in a prior model, neither CatRwd dataset significantly differed from the category feedback data.

3.3.1.2. Reaction Time Analyses

We also conducted exploratory analyses on the average response times (RT) for participants in each condition. Similar to the analyses we conducted to examine learning over time, we utilized a Bayesian mixed-effects regression to determine the extent of the differences in RT between feedback types and both category and rewards structures. With categorical feedback participants as the reference group ($\overline{RT}$ = 0.643s, SD = .275s), there was no evidence of a difference in response time between categorical and catrwd feedback participants ($\overline{RT}$ = 0.756s, SD = .719s; 0.08, [-0.03, 0.20]). However, RTs did differ, on average, when comparing categorical feedback data to both variable reward feedback data ($\overline{RT}$ = 0.951s, SD = .541s; 0.266, [0.15,0.38]) and discrete reward feedback data ($\overline{RT}$ = 1.071s, SD = .539s; .459, [.322, .598]). This suggests that giving only categorical feedback led to the quickest response times, while giving only reward feedback led to the longest response times. Interestingly, even though the mean number of optimal choices did not differ between categorical and discrete reward feedback, they did differ in terms of reaction time.

We also explored how RT values changed over time. Over the course of the four 100-trial blocks, there is evidence that the overall RT values for each feedback type decreased over time as the task was learned (-.09, [-0.15, -0.03]). The interaction terms for feedback types and both category and rewards structures show no evidence of differences suggesting that the decrease in RT over time was at least partially uniform over the course of the task in each feedback type and condition. Figure 3.6 details the uniform decrease in average reaction times, and also shows the impact that categorical and reward-based information has on reaction times suggesting that more processing is needed for numerical information independent of the level of uncertainty.

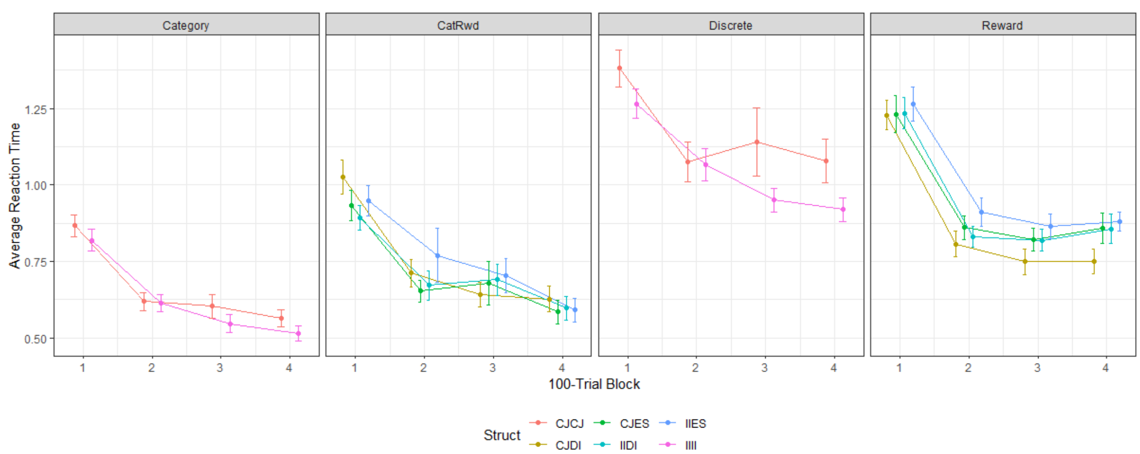

Figure 3.6 Visualization of the average reaction times across trial block by feedback type and condition. Error bars represent the standard deviation about the mean.

3.3.2. Model-Based Analyses

Next, we examined how well each of the three models described above (category-learner, reward-learner, and catrwd-learner) accounted for the data by fitting each model to each participant's data individually. We then conducted post-hoc simulations to examine how well each model could reproduce the pattern of effects found in each condition. These post-hoc simulations involved simulating the experiment with each model with the best-fitting parameters from each participant. The goal of these post-hoc simulations is to compare how well each model's simulated, or predicted, behavior aligns with the observed behavior from our participants.

As mentioned above, we assumed that categorical outcomes could be expressed as binary values of 0 and 1. To validate this, we fit all the categorical feedback participant data to a modified category-learner model that included an additional free parameter to represent incorrect categorizations. This model would fit the best value for incorrect categorizations with a value range of -1 to 0. We found that given our categorical feedback data, the best fitting value for incorrect categorizations is not likely to differ from zero (-0.041, [-0.074, -0.009], $BF_{10} = 0.639$) when conducting a Bayesian one-sample t-test with an alternative hypothesis that the best fitting value was less than 0. For our particular model and task, we conclude that binary values of 0 and 1 were appropriate values to represent categorical outcomes in the following analyses. For completeness in modeling, we also completed a model where the free parameter for incorrect feedback ranged in value from -1 to 1. Interestingly, this resulted in a mean value of 0.6558 for this parameter suggesting that participants may have been more likely to make the same response to a similar stimulus, even if that response might have been incorrect.

3.3.2.1. Comparison of Models

Each model was fit to participants' data on an individual basis by maximizing the log-likelihood of the model's prediction for the optimal response on any given trial. We then calculated the Bayesian Information Criterion (BIC), a goodness-of-fit measure (Schwarz, 1978) to compare the fits of different models. Statistics such as BIC penalize models with more free parameters. Smaller BIC values indicate a better fit of the model to the data. Importantly, the category-learner and the reward-learner are considered to be simpler models nested in the full catrwd-learner model. Our two nested models are functionally identical to the full model, but do not include the 'q' parameter. This means that while the log-likelihood of the full model cannot be worse than the nested models, the BIC values for full model may be greater, due to the parameter penalty, and thus may indicate a poorer fit.

The BIC values, along with the average best-fitting parameter estimates for each condition are shown in Table 1 below. In general, the catrwd-learner and category-learner models were fairly consistent in terms of BIC. Due to the nested nature of the models, the similar fits of the category-learner and catrwd-learner models suggest that adding reward information to the model did not provide much improvement in fit. This can also be seen by examining the q parameter, which weights categorical versus reward information. These values are less than .5, on average, for all groups, and close to 0 for every group but the RwdCJDI feedback condition. Recall that when q=0 the model relies exclusively on categorical feedback and the catrwd and category-learner models are identical.

Overall, the best fitting model for each condition was the category-learner. In every instance, aside from the RwdCJDI condition, the catrwd-learner showed the second-best fits. Interestingly, in this RwdCJDI condition where the reward-learner was the second best fit, the catrwd-learner showed a "q" parameter value greater than 0.5, which suggest for this condition alone, reward information was given more weight. For the categorical feedback participants, there was no reward information given, so the model outputs should be close to identical as the reward-learner uses reward values of 1 and 0 when fitting. Discrepancies in fits are attributed to the reward scaling function (Eq. 4). Until the full range of rewards are known, 1 and 0 in this case, the deviations between both the category- and reward-learner models may differ. Table 3.1 below details the best fitting parameters and BIC values for each model and fitted condition.

Table 3.1 Model Fit and Parameter Values

Feedback Type		Structure	Model	a	s	d	s2	q	BIC
	Categorical	CJ	Category-Learner	0.3676	0.0879	0.3949	1.9780	NA	426.2529
			Reward-Learner	0.4048	0.1220	0.3621	1.9040	NA	427.8166
		II	Category-Learner	0.4435	0.0962	0.3326	2.3212	NA	376.2340
			Reward-Learner	0.4356	0.1116	0.3434	2.3618	NA	377.1694
	CatRwd	CJDI	Category-Learner	0.4479	0.0898	0.3817	2.1326	NA	402.3978
			CatRwd-Learner	0.4318	0.1007	0.2904	2.6180	0.4346	403.3981
			Reward-Learner	0.4425	0.1201	0.2378	3.1834	NA	417.3409
		CJES	Category-Learner	0.4447	0.0964	0.3500	2.3634	NA	407.3898
			CatRwd-Learner	0.4729	0.1035	0.3969	2.1078	0.1271	416.5260
			Reward-Learner	0.4555	0.1641	0.6004	1.5954	NA	426.1160
		IIDI	Category-Learner	0.4143	0.0416	0.2481	3.0687	NA	376.9999
			CatRwd-Learner	0.4137	0.0795	0.2628	3.0572	0.1182	386.0445
			Reward-Learner	0.4357	0.0999	0.4813	2.6239	NA	406.1975
		IIES	Category-Learner	0.4301	0.0781	0.2323	3.1498	NA	369.3368
			CatRwd-Learner	0.4375	0.0651	0.2149	3.2076	0.0741	373.0080
			Reward-Learner	0.4378	0.0835	0.3979	2.7167	NA	386.7135
	Re	CJDI	Category-Learner	0.5159	0.1159	0.6553	1.4119	NA	500.5736

			CatRwd-Learner	0.5526	0.1529	0.5138	1.9299	0.6234	504.4965
			Reward-Learner	0.5191	0.1221	0.5197	2.0461	NA	501.6081
		CJES	Category-Learner	0.4507	0.1212	0.5168	1.6430	NA	462.0999
			CatRwd-Learner	0.4657	0.1491	0.5375	1.6463	0.0849	467.0120
			Reward-Learner	0.4806	0.1821	0.7698	1.1953	NA	474.1996
		IIDI	Category-Learner	0.5614	0.0949	0.4880	2.0662	NA	450.2910
			CatRwd-Learner	0.5560	0.1099	0.4699	2.2264	0.1550	457.0588
			Reward-Learner	0.5323	0.1122	0.5425	2.1093	NA	462.5526
		IIES	Category-Learner	0.5044	0.0908	0.5047	2.1726	NA	421.9575
			CatRwd-Learner	0.5219	0.0950	0.5141	2.1439	0.1564	428.6039
			Reward-Learner	0.5109	0.0848	0.7040	1.7076	NA	429.8681

Note: Average best-fitting parameter estimates and BIC values for each condition. Smaller BIC values indicate a better fit of the model to the data. The best fitting model within each feedback type and condition is shaded and the BIC value is bolded. CJDI refers to a conjunctive rule with a difficult reward structure; CJES refers to a conjunctive rule with an easy reward structure; IIDI refers to an information-integration rule with a difficult reward structure, and IIES refers to an information integration rule with an easy reward structure.

3.3.2.2. Post-hoc Simulations

Using the process detailed by Ahn, Busemeyer, Wagenmakers, & Stout (2008), we ran 100 simulations using the best-fitting parameters for each participant as the input for each of the respective three main models. The data for each of the 100 simulations for each participant were aggregated by trial to produce a single averaged simulated dataset for each model/condition combination. Thus, for each condition and each model we generated the average predicted proportion of best responses on each trial, across all 100 simulations. The participant data was aggregated in the same manner which yielded the observed proportion of correct choices made on each trial, across all participants in each condition. We then used both sets of data to compute the mean square deviation values for each combination of datasets using the formula in Equation 7 below:

$$MSD = \frac{1}{n}\sum_{t=1}^{n}(\overline{D}_{exp,t} - \overline{D}_{sim,t})^2 \qquad (7)$$

where *n* is the number of trials each participant observed, and *t* represents the individual trial number. For each experiment, since there were only two alternatives, we calculated the mean square difference (MSD) of the average proportion of correct responses (D) between the experimental data and simulations. As performed in Ahn et al. (2008), we used the percentage values, instead of proportion values, when computing the MSD. We report the MSD values in their root form, as it is a more understandable metric of model performance, for each Feedback type, Model, and Structure in Table 3.2 below. The mean deviation values (MD) indicate the percentage of time the simulated data deviated from the observed data on each trial.

Table 3.2 Mean Deviations Between Actual and Simulated Data

Feedback	Structure	Category-Learner	CatRwd-Learner	Reward-Learner
Categorical	CJ	14.1796	NA	15.3241
	II	12.7686	NA	14.2685
CatRwd	CJDI	7.7376	7.7170	7.7393
	CJES	9.1892	9.1183	9.1751
	IIDI	8.2149	8.1521	7.9289
	IIES	8.4790	8.5085	8.2112
Reward	CJDI	9.1578	9.1181	9.1389
	CJES	12.3855	12.4063	12.4232
	IIDI	9.2743	9.2787	9.1626
	IIES	10.1036	10.0732	9.9518

Note: Table of mean deviation (MD) values. Lower values indicate that there were fewer deviations between both datasets on average. Shaded cells indicate which model produced the lowest deviation for a given feedback type and condition. CJDI refers to a conjunctive rule with a difficult reward structure; CJES refers to a conjunctive rule with an easy reward structure; IIDI refers to an information-integration rule with a difficult reward structure, and IIES refers to an information integration rule with an easy reward structure.

Based on Table 3.2 above, each of the models showed less than a 15% deviation in total, and about 10% deviation on average (10.224), between the post hoc simulated data and the observed behavioral data. This means that on average, over the course of the

400-trial task, a simulated participant using the same best fitting parameters as a human, would show incongruent behavior on ~40 trials. Within each feedback and structure condition, the MD values for each model are relatively similar, showing a difference in MD values of 0.2444 on average, 0.0893 median, between models with the categorical feedback conditions showing the highest deviations on average. This would equate to approximately ±1 deviation between models over the course of 400 trials. Overall, each of the models were fairly consistent within each condition in reproducing the experimental data from participants' best fitting parameters. However, the reward-learner did show the smallest deviations on average in reproducing the behavioral data. Below, in Figure 3.7, we show the post hoc simulation learning curves for each condition based on the best fitting values.

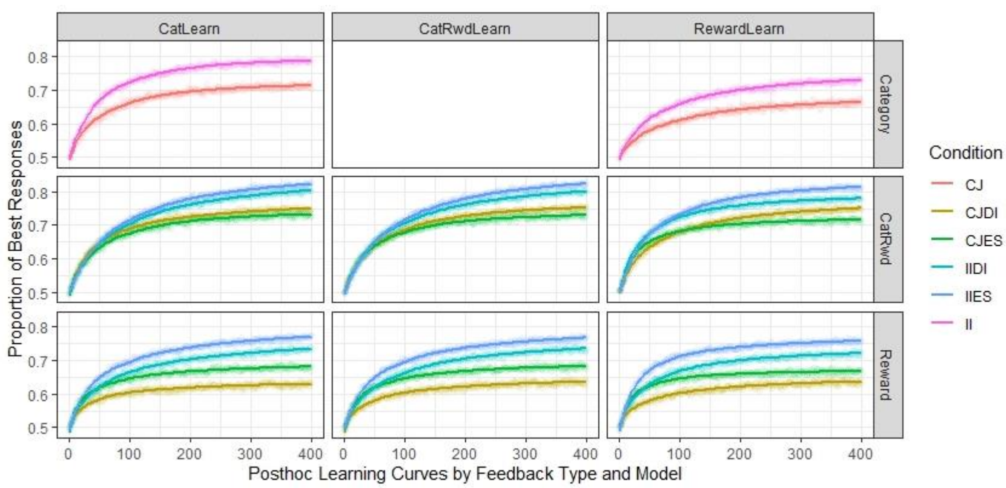

Figure 3.7 Learning curves for the post hoc predictions of each model, feedback type, and condition. The best fitting line is plotted for each model/data pair.

3.4. Study 1 Discussion

Real-world feedback often takes different forms, from feedback that is discrete and categorical to feedback that lies on a more continuous scale. In the current study, we simulated and experimentally tested each form of feedback and how they impact human category learning. Overall, based on our results, when learning to categorize novel line stimuli, it is apparent that categorical feedback produces better learning on average as compared to variable numerical feedback alone, but not discrete numerical feedback. However, the rate at which both forms of reward feedback participants learned to make optimal responses was slower, but overall performance differed. This seems to imply that the definitive information gained from categorical feedback and discrete numerical feedback better facilitates learning as compared to the nuanced, and possibly more ambiguous information gained from continuous reward feedback. These findings corroborate our model-based hypotheses and provide evidence for how different forms of feedback could potentially impact real world learning.

When given categorical and continuous rewards simultaneously, our results suggest that people tend to disregard the continuous information in favor of the categorical feedback. Participants given both categorical and continuous reward feedback learned at the same rate as participants given categorical feedback alone. Our model-fitting results indicated that most participants who were given both types of reward feedback tended to weight categorical feedback much more than continuous feedback. However, participants in the CJDI conditions gave slightly more weight to continuous reward feedback based on the value of the ‘q’ parameter for these conditions.

This is likely due to these conditions giving the largest rewards for the stimuli nearest the boundary of category membership. Since these stimuli were likely more difficult to categorize, these participants may have relied more on the numerical reward, rather than categorical, information.

3.4.1. Differences Between Feedback Types

A definitive reason accounting for the differences between categorical and numerical rewards is still unclear. Speculatively, one such reason could be the degree of variability in the two types of feedback. Categorical feedback is most often discrete, and outcomes can be understood with a degree of expectation. As an example, if you are asked to categorize an image of a dog, an outcome of correct vs. incorrect could be understood that the decision made was either right or wrong.

Comparatively, continuous numerical feedback often has more variability surrounding the initial outcomes. Receiving a value of 85 and nothing else when attempting to categorize the picture of the dog may not give much information initially. A person may have questions about the range of rewards: whether the value was high or low; or be simply unsure how to apply the knowledge they just learned. While people are indeed able to learn from continuous numerical information, based on our findings, their overall performance and rate of learning would likely be poorer as compared to if they were given categorical feedback. This might be due, in part, to the contrast in the variability, or uncertainty surrounding the expectation, of each type of outcome and feedback (e.g. Walker et al., 2019).

Further, in the modeling framework we presented, categorical feedback represented the maximum possible reward, whereas continuous feedback varied. This may have implicitly modeled a reduced level of uncertainty in the categorical condition because reward prediction errors were larger than in the continuous condition. In doing so, it may have led to larger updates of the connection weights between the hidden nodes of the model and response output nodes, and better learning of the categories. Thus, it is possible that the less variable, or more discrete, the feedback is (i.e. coin images, static values, category labels), the better the predicted performance is as compared to feedback consisting of a more variable range of values (i.e. continuous range of rewards, distribution of values, etc.). This notion is seemingly validated with the results of the discrete numerical feedback, which had no variability, showing performance similar to that of the categorical feedback participants; and it is a potential explanation of why past research has shown that there is very little difference in categorization performance when comparing cognitive and monetary feedback (e.g. Daniel & Pollmann, 2010) or when comparing groups given differing magnitudes of discrete rewards (e.g. Bellebaum et al., 2010; Miller & Estes, 1961; Peterson & Seger, 2013).

It is also possible that the process of asking categorical feedback participants to categorize object, versus asking the continuous feedback participants to predict which option would give greater rewards, partially led to the discrepancy in performance between both groups. In a blocking paradigm by Bott and Hoffman (2007), participants who were asked to predict outcomes demonstrated significantly worse performance than participants asked to learn categories, and reflective of our own findings. Further, an additional verbal account of our observed differences could be due to the option labeling we used for each option. The simple action of labeling something as a member of a category has been shown to lead to quicker classification and learning as compared to the use of generic ‘option’ labels (Lupyan, 2012; Lupyan et al., 2007).

3.4.2. Impact of Reward on Stimuli Difficulty

Prior research has shown that people show poorer performance as the observed stimuli become harder to categorize or discriminate between (Daniel et al., 2011; Krebs et al., 2012; Schevernels et al., 2014). While our study did not directly compare performance between easier and harder stimuli within participants, we did compare the effect of rewarding either type of stimuli more than the other between participants. In the ES conditions of our task, the most rewarding stimuli were the most typical of their respective category, whereas the DI conditions gave the most rewards to the stimuli at the bounds of each category. When given only reward feedback, participants in the ES conditions show far better performance than their counterparts in the DI conditions. This suggests that when learning from reward values alone, learning is best facilitated when the most typical stimuli give the greatest amount of reinforcement. An explanation for this difference, based on our model fitting, is that the increased difficulty promotes greater reliance on reward information. Interestingly however, the model also shows, via the feedback weighting parameter ‘q’ in the catrwd-learner, that poorer performance is associated with weighting reward information more heavily in the reward feedback conditions.

3.4.3. Conclusion

We have shown that giving continuous versus discrete feedback leads to important learning differences in a categorization task. We detailed a possible mechanistic account of these learning differences using connectionist learning models. Both the models and the observed data suggest that, when learning to categorize novel stimuli, giving feedback that includes categorical information will lead to significantly better performance than if given feedback consisting of only reward information. Importantly, when given both types of information simultaneously, categorical information is likely to be more heavily weighted than reward information. We detailed that these learning differences likely stem from differences in the magnitude of predictions errors associated with each form of feedback, and that the perceivably larger amount of uncertainty surrounding the reward information had an effect on how well the categories were learned. Additionally, when given reward information alone, both the modeling and behavioral data showed that the relative difficulty of categorizing the stimuli affected learning. When the most typical stimuli of a category are associated with the largest rewards, we should expect performance similar to that of categorical feedback alone. Likewise, when the least typical stimuli or those near the bounds of category membership are associated with the largest rewards, poorer performance is to be expected. Thus, the present behavioral results and theoretical account suggest that feedback can be structured in different ways to promote better learning.

4. STUDY 2

4.1. Introduction

What is more important, receiving a large reward, or choosing what feels the most familiar? In category learning, research has most often explored how similarity to learned exemplars are predictive of future categorization decisions (e.g. Davis et al., 2014; Medin & Smith, 1984; Nosofsky, 1988). In contrast, reinforcement learning tends to focus solely on the modulation of rewards signals and how it impacts decisions between alternatives (e.g. Rescorla & Wagner, 1972; Speekenbrink & Konstantinidis, 2015; Sutton & Barto, 2018). These two facets are markedly dissimilar: in category learning, the feedback is discrete, but the stimuli are variable (e.g. Ashby & Maddox, 2005); in reward learning, the feedback is variable, but the stimuli are often discrete (e.g. Daw et al., 2006). However, while both forms of feedback are effective in guiding learning in their respective paradigms and real-world scenarios, it is currently unknown how differing degrees of stimulus similarity and magnitudes of reward values affect how categories are learned.

Supposing that a task is given where a person must categorize images of dogs and cats, over time people would likely become proficient in separating both of the animals regardless of whether feedback was given or not. Stimulus similarity is assumed to be a strong factor in the determination of category membership in this regard (Pothos & Chater, 2002, 2005). However, suppose that a separate task instead includes a couple scaly reptiles in addition to the cats and dogs. With no external directions, relying on similarity alone, people would likely now group cats and dogs together in one category and place reptiles in their own category. This type of behavior is repeatedly seen in the unsupervised category learning literature, and can lead to a variety of different rules employed between subjects (Ashby et al., 1999; Clapper & Bower, 1994; Love, 2003).

Similarly, in reward learning, people are able to easily discern which, of multiple, alternatives lead to optimal rewards given sufficient time (e.g. Daw & Shohamy, 2008; Sutton & Barto, 2018). Further, research has also shown that if the contingencies surrounding the potential alternatives change over time, via variations in reward magnitude or probability, people are able to easily adapt and change their response in order to continually maximize their gains (e.g. Daw et al., 2006; Gershman, 2018b; Kool et al., 2018). Thus, if modulating the degree of similarity between stimuli can lead to differing classifications of the same stimuli, and differences in reward magnitude can prompt preferential choices among alternatives, it may also be possible that modulating the magnitude of rewards associated with a particular category or feature could have the same effect.

Much like observed in Study 1, categorical feedback is likely sufficient when attempting to learn how to categorize novel stimuli when all other factors are relatively equal (i.e. location in stimulus space, equated rewards, etc.). However, would categorical feedback still be as effective should these variables differ between categories? In a situation where the novel stimuli are closer in similarity to one category over another, it is possible that categorical and numerical feedback could elicit differing views of category membership. In this example, if Stimulus X is more similar to Category A than Category B, both categorical and numerical feedback would likely reinforce the decision made based on the similarity between the current stimulus and the exemplars of Category A. Instead, if we were to give a large reward for stimuli like Stimulus X when they are categorized as part of Category B, it is unknown if relatively large reward values would be able to ‘overrule’ stimulus similarity. While our current reasoning suggests that it is a possible outcome based on category and reward learning theory, stimulus similarity is an impactful force in human category learning (e.g. Conaway & Kurtz, 2017; Nosofsky, 1984; Tversky, 1977) which may have unforeseen results.

4.1.1. Simulated Results and Hypotheses

To explore this idea, performance on a conjunctive rule category structure task, consisting of four category clusters, was simulated using each of the three main ALCOVE variants: Category-Learner, Reward-Learner, and CatRwd-Learner. In these simulations, the agents were shown a 300-stimuli training phase comprised of equal observations of 4 category clusters which incrementally radiated from the center of the stimulus space with Category D being most central, and Category A being the most distant from the center of the stimulus space (ES and DI conditions), or category clusters which were equidistant from both the center of the stimulus space and the other three categories (EQ; Figure 4.1).

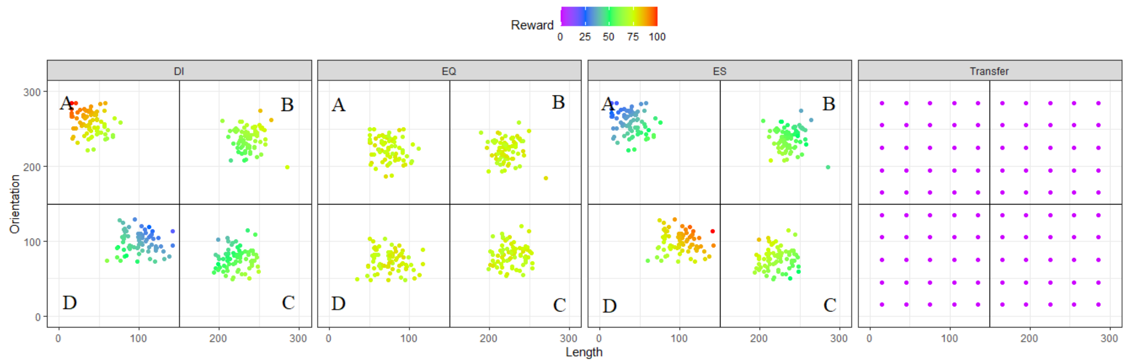

Figure 4.1 Visualization of the category and reward structures for the stimuli in each training phase condition and the stimuli for the transfer phase in Study 2.

Upon categorizing all 300 stimuli, the agents completed a test phase consisting of 100 novel stimuli which were generated from equally distributed orientation and length value combinations from across the entire stimulus space (Figure 4.1-Transfer). Thus, with category D being most central, a majority of the test stimuli should have been categorized as category D if stimulus similarity was a dominant predictor of choice. However, three different rewards structures were also utilized to determine if numerical information would have an impact on categorical decisions: the ES, DI, and EQ rewards structures. This led to a factorial combination of 9 total simulated conditions. Importantly, these reward structures are not to be confused with the ES and DI reward structures in Study 1. The ES and DI structures in the current study refer to the distance-based rewards that are associated with each condition. For the ES condition, the most central category cluster (D) is associated with the largest reward values (Figure 4.1-ES), whereas the in the DI condition the largest rewards are associated with the most distant category (A; Figure 4.1-DI). Specifics about the category cluster generation and reward calculations will be further detailed in Section 4.2 below.

During training, a correct response was determined to be the choice that correctly categorized the line stimuli into the correct category, regardless of reward structure. In the transfer phase, however, 'best' response was expected to differ by conditions depending on whether stimulus similarity or reward magnitude was a greater determinant in choice for a given participant. In the ES condition, Category D should have been the most frequently selected in the transfer phase as it was both the most similar to the majority of transfer phase stimuli and the most rewarded category cluster overall. For the DI condition, however, the best choice was likely to differ depending on whether participants were more sensitive to large reward values or stimulus similarity: Category A for reward magnitude and category D for category similarity. The mean correct responses for each conditions' training phase, and the predicted category selection proportions in the test phase, can be observed in Figure 4.2 below.

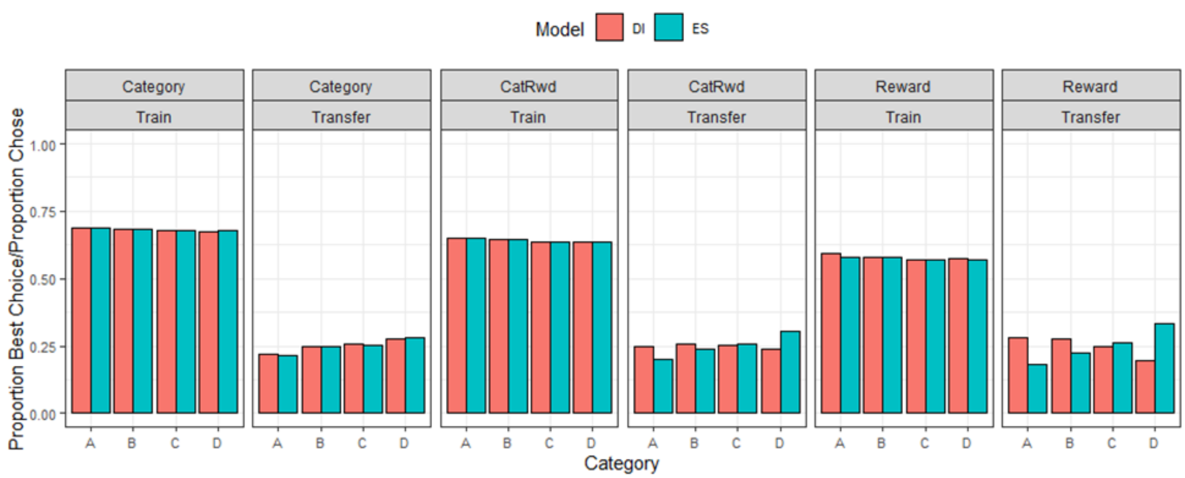

Figure 4.2 Simulated predicted best choice between models and conditions for the training phase and the proportion of category selections for the transfer phase.

While the models predict similar learning trends in the training phase as compared to Study 1, where categorical feedback promotes slightly better performance, the models detail some differences in transfer learning depending on what the optimal choice is assumed to be in the DI conditions. For categorical feedback, similarity, unsurprisingly, has the largest predicted influence on which option is determined to be most associated with the transfer stimuli as category D is chosen more often. Interestingly for the numerical feedback conditions, the results differ depending on whether the agent was in the ES or DI condition. For the ES condition, D is more likely to be chosen, whereas for the DI conditions, A is more likely to be chosen, but still less than D. However, as compared to categorical feedback, category A was selected notably more when given numerical feedback. This suggests that as categories become more dissimilar, reward values may become more influential. Additionally, seemingly reinforcing the predictions of the models in Study 1, when the most similar stimuli are the most rewarded (ES conditions), predicted performance between categorical and numerical feedback is almost indistinguishable.

Thus, given these simulated predictions, we hypothesized that attention, or sensitivity, towards stimulus similarity in the training phase would be a critical factor in predicting transfer phase choices in this paradigm. Based on the simulated category-learner, and all ES reward conditions, showing a preference for category D in the test phase despite near-identical predicted performance among all category clusters in the training phase. However, the reward-learner model suggests that reward information may overrule similarity to some extent based on the DI condition predictions where category A, the furthest and least similar, but most rewarded, category, was preferred in the test phase.

4.2. Method

4.2.1. Participants

All participants in this study were recruited from Texas A&M University and were given partial course completion credit for their participation. Each participant completed an online version of the informed consent and experimental task. In following the same study conditions simulated above, we sought to recruit 30 participants for each of the 9 conditions based on a power analysis using 80% power with a moderate effect.

In total, we recruited 31 participants for the Categorical-DI, 27 for Categorical-ES, and 29 for Categorical-EQ for the categorical feedback group; 32 for Reward-DI, 32 for Reward-ES, and 31 for Reward-EQ; and for the CatRwd feedback conditions we recruited 29 for CatRwd-DI, 28 for CatRwd-ES, and 30 for CatRwd-EQ. Overall, we recruited 269 college-attending participants. With 9 conditions and 269 participants, post hoc power calculations detail that we achieved 0.838 power to detect moderate effects ($\eta^2 = .059$).

4.2.2. Task and Structure

4.2.2.1. Experimental Task

This task consisted of two discrete phases. The first phase was similar in design to Study 1 where participants were shown a set of unique line stimuli. In this phase, participants were shown 300 unique line stimuli, with an equal number of stimuli belonging to each of the 4 category clusters, in a randomized order. These lines varied in both line length and orientation, and the participants were asked to determine which of four categories they think the line belongs to. In a second phase, participants observed 100 more unique line stimuli that were created from the combination of 10 equally distanced line lengths and orientations. This ‘test’ phase was included as a way to measure how well participants were able to generalize their learned knowledge to new, but similar stimuli.

Like Study 1, depending on the feedback type given to the participants, they received either categorical feedback (Category), continuous reward-based feedback (Reward), or a combination of both types of feedback (CatRwd) in response to their choices during the initial ‘training’ phase. However, during the test phase, no feedback of any form was given to the participants. Again, like Study 1, the terminology used in the tasks differed by feedback type. For the participants given Reward feedback, the participants were informed to use the line to aid them in predicting which option would give the largest reward value. The remaining two feedback conditions simply asked the participants to categorize the stimuli and utilize the feedback to learn.

4.2.2.2. Category and Reward Structure

As previously mentioned, the task consists of a training phase with four category clusters, and a test phase with stimuli evenly distributed across the stimulus space. A graphical depiction of this can be seen in Figure 4.1 above.

In this study, only one category structure was used with a stimulus space range of 0-300 for both line length and degrees of orientation. Each of the four training category clusters were created from repeated sampling of four bivariate normal distributions. The variance of these distributions were held constant between clusters (SD = 15), but the distribution mean points constructed a spiral outwards from the center of the stimulus space (150, 150) at a rate of ~28.28 in Euclidean distance (+20 length, +20 orientation) starting at (100, 100) for category D. From there, the mean points for the remaining category clusters radiate outwards: (220, 80) for category C, (240, 240) for category B, and (40, 260) for category A. For each category cluster, 75 stimuli were sampled for a total of 300 training stimuli. We constructed this difference in spatial similarity between category clusters to aid in our analysis of how similarity and reward impacts category learning.

The reward values for the ES and DI conditions were computed based on the Euclidian distance to the center of the stimulus space (150, 150), as compared to the GCM probabilities used for the ES and DI reward structures in Study 1. For the DI condition, the raw values from this computation were scaled to the range of 100-50, with the larger values being associated with category A, the furthest cluster from the center. For the ES condition, the rewards were inverted with the largest magnitude rewards being associated with category D, the closest cluster to the center of the stimulus space, instead. Like mentioned prior, there is no feedback in the test phase, thus no rewards were assigned to any of the test stimuli. Figure 4.1 shown prior details the reward value gradients between each reward condition.

In addition to the previously described category and reward structure, a separate category structure with uniform rewards was created to serve as a control condition for each feedback type. In this category structure condition, instead of the category clusters radiating from the center, each of the clusters are equidistant from the center, at a Euclidean distance of 99.25, and each other (107.70 distance). The reward structure for these stimuli was uniformly distributed ($U(70,80)$) for each stimuli. Given that the reward values for each of these stimuli are consistent between each category cluster, we will be able to analyze the differences between the other reward and category structure conditions as they pertain to stimuli distance, similarity, and reward value within each feedback type. Figure 4.1 above details the general structure for these equated (EQ) conditions.

4.2.3. Task Procedure

This task, much like the task in Study 1, follows a procedure similar to what is described in the General Method section. Like previously described, Study 2 consisted of both a training and test phase, each with a certain number of stimuli/trials. Upon consenting to the study, each participant was randomly assigned to one of the 9 conditions and given adequate instructions on how to progress in the experiment. Each set of instructions was tailored for each specific feedback type: mainly the inclusion or exclusion of any information relating to the process of categorization. The participants would then begin the training trials and be shown a single, unique, randomly drawn, line stimulus. The participant would see a prompt relevant to their feedback type and four buttons towards the bottom of the screen that they were told to use to make their selection. The labels of these buttons were also dependent on the feedback type the participant was assigned to as well. Upon making their choice, the participant would see the relevant feedback for a period of 2 seconds before an intertrial interval screen was shown for 0.5 seconds before repeated the trial process. A visual depiction of this process can be seen in Figure 4.3 below.

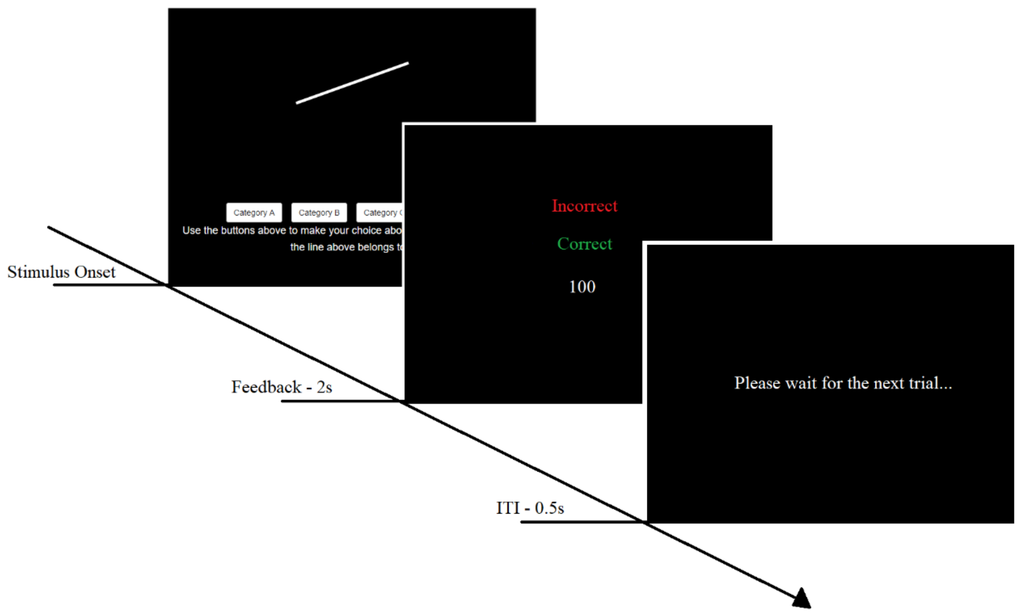

Figure 4.3 Sample trial diagram for Study 2.

The training phase of this task was split into three parts separated by two break screens. After participants successfully completed 100 trials, they were given the opportunity to take a short break before continuing. After completing two more sets of 100 trials, the participants would move on to the final test phase and be given a new screen detailing the information about the final phase. The participants were informed that they would again see new, unique stimuli, but that they would no longer receive feedback about their decisions. Upon completion of the test phase, participants were given a short debriefing about the task and the goals of the experiment and were then directed to a screen where they could claim their credit for participation.

4.3. Study 2 Results

4.3.1. Behavioral Results

4.3.1.1. Training Phase Results

4.3.1.1.1. Learning Over Time

Overall, in the training phase, each of the four individual category clusters were relatively well learned by the participants regardless of feedback type or reward condition with final block categorization rates of over ~80% (Figure 4.3b). This performance is likely due to the deterministic nature of the category clusters which has been shown to promote an ease of category learning (Liu et al., 2020). However, based on Figure 4.4a below, there appears to be distinct differences between both feedback types and reward structures conditions. For consistency with the results reported in Study 1, all of the following results will be reported in Bayesian terms. To that effect, Bayes' Factors greater than 3 in value are considered to be moderate evidence for the alternative hypothesis, and credible intervals for parameter estimates (for the forth coming Bayesian multilevel models) which do not include 0 can be interpreted as evidence that the true value of the parameter has some probable value of impacting the model (Nalborczyk et al., 2019). Both forms of results will include the credible intervals for their respective parameters (denoted by square brackets following the parameter value).

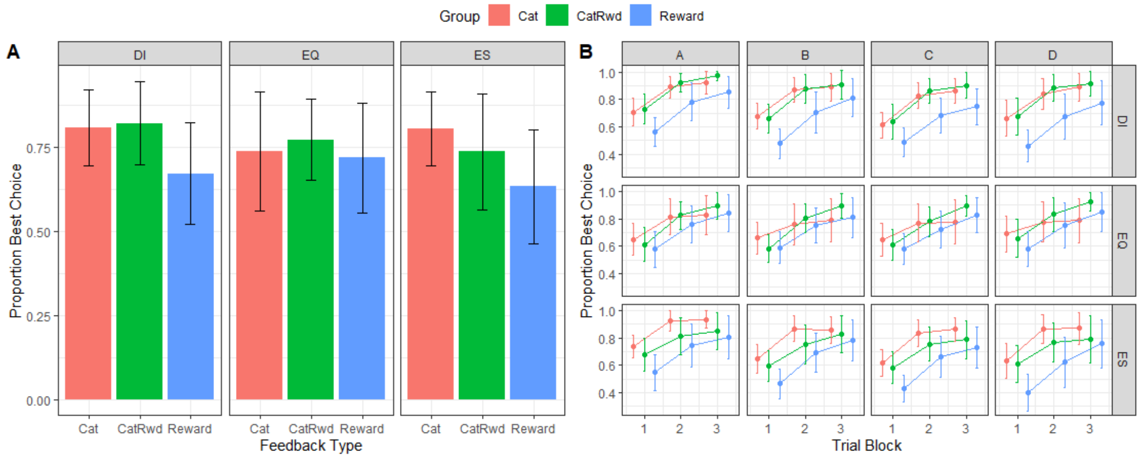

Figure 4.4 Behavioral data by feedback type and reward structure. A.) Average overall proportion of best responses/correct categorizations. B.) Average best responses across trial blocks for each category cluster. Error bars represent the standard deviation about the mean.

To compare the data, we ran multiple Bayesian multilevel models using the brms R package (Bürkner, 2017). First, to determine if either feedback type or the reward structures had an impact on training performance, a model that predicted the proportion of correct categorizations/best responses from feedback type, reward structure, and trial block using trial block as a random slope and participant number as a random intercept was used. With categorical feedback as the reference group, catrwd feedback (.022, [-.061, .110]) did not show evidence of differing in terms of the mean proportion of correct categorization/best choices in the training phase collapsing across all reward structures. Reward feedback, in contrast, did show evidence of differing when compared to categorical feedback (-.166, [-.256, -.077]). Further, trial block also had a significant impact on the model (.107, [.069, .145]) providing evidence that, overall, participants gained proficiency on the task as they progressed similar to the behavior shown in Study

1. In terms of reward structures, using the DI condition data as the reference group, neither the EQ (-.076, [-.161, .012]) or the ES (-.002, [-.089, .084]) conditions showed evidence of significantly impacting the model as compared to the DI condition. However, multiple interaction terms showed evidence of differences within each feedback type. To explore these differences, the following paragraphs will detail the output of simplified models of each feedback groups' data.

Within the categorical feedback participants, performance on the Cat-DI condition did not differ from either Cat-EQ (-.053, [-.149, .045]) or Cat-ES (.046, [-.041, .135]). As detailed by the full model, categorical feedback participants showed improved performance across trial blocks (.106, [.077, .134]), however the slope of Cat-EQ participants differed from Cat-DI (-.051, [-.093, -.007]), but not in Cat-ES participants (.007, [-.036, .050]) suggesting relatively slower learning in the Cat-DI participants. In the catrwd behavioral data, there was no evidence of a difference in performance between each of the three reward structure conditions. Though, unlike the categorical feedback participants, there was no evidence of differing slopes between reward structure conditions. For the reward feedback participants, performance in the Rwd-DI did not differ from Rwd-ES (.003, [-.103, .111]), but did differ from Rwd-EQ (.094, [.005, .183]). As with the catrwd participants, there were no differences in slope between reward structure conditions.

4.3.1.1.2. Reaction Time Analyses

In the initial analyses of reaction time, the data was heavily skewed to the right. While skewness is normal for reaction time data, since there cannot be negative reaction times, some of the reaction time values were highly improbable (trials taking longer than a few minutes) indicating that on some trials, participants may have gone off task such as to look at their phone or reply to a text message. Due to the online delivery of this task, there was a distinct lack of control as compared to in lab experimentation. As such, for the following reaction time analyses, the median reaction time values are reported, and the results should be interpreted with this in mind. In Figure 4.5 below, the reaction time, in seconds, for each feedback type and reward structure is visualized.

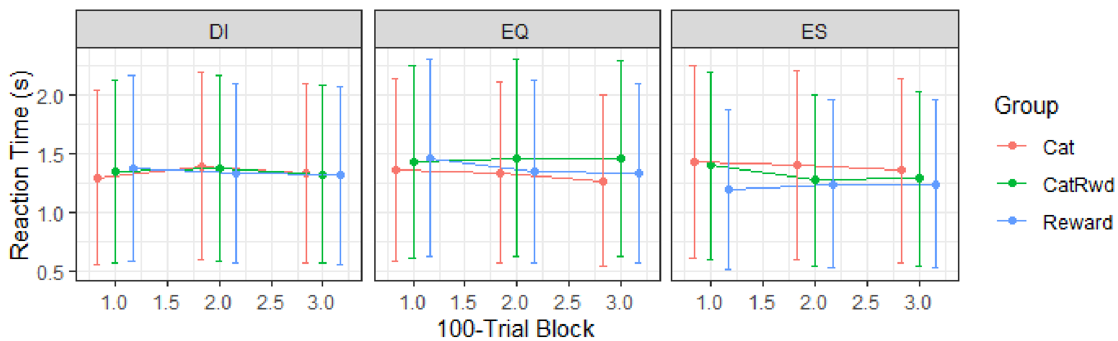

Figure 4.5 Median reaction time over trial blocks by feedback type and reward structure. Error bars represent the interquartile range.

Overall, there were no differences in the mean reaction time (RT) values between feedback types with both catrwd (-.029, [-.173, .122]) and reward (-.012, [-.116, .136]) feedback showing no difference when compared to categorical feedback. Unlike expected, based on the findings of Study 1, there was no indication of a negative slope in reaction times as the task progressed across trial blocks (.028, [-.043, .100]). The interaction terms also suggested no evidence of differences within each group of participants. As discussed prior, reaction times in a lengthy online task such as this, may not produce accurate response times.

4.3.1.2. Test Phase Results

In the test phase of this task, participants were shown 100 stimuli, comprised of an equal distribution of line length and orientation, in a random order. They were allowed to choose any of the four categories/options as their response. To determine if the overall proportion of category selections differed due to either the feedback type or reward structure, we ran a series of Category Cluster X Reward Structure Bayesian ANOVAs. In the following results, as some of the Bayes Factors are incredibly large (>100), the log version of the Bayes Factors will be reported instead to ensure that the differences in the BF magnitude is concisely conserved. When taking the natural log of the Bayes Factor, the band of values between -1 and 1 represent no evidence for either hypothesis. The values for the alternative hypothesis increase in evidence strength towards infinity, and for the null hypothesis, unlike base 10 Bayes Factors, the values will decrease towards negative infinity. (for reference, a BF_{10} of 3 is equated to a $Log(BF_{10})$ of 1.099, and a BF_{10} of 1/3 is equated to a $Log(BF_{10})$ of -1.099). Importantly, in interpreting Bayesian ANOVAs, the BFs are the result of a comparison of a model containing the factors of interest and a null model including only an intercept. In addition to reporting the outcome of the ANOVAs, we will also report the post hoc comparisons for the individual comparisons of category clusters and reward structures within each feedback type. A Cauchy (0, r = $1/\sqrt{2}$) prior was used for each post hoc test. In Figure 4.6 below, a visualization of the forthcoming comparisons of the proportions of category selections, by feedback types and reward structure, can be seen.

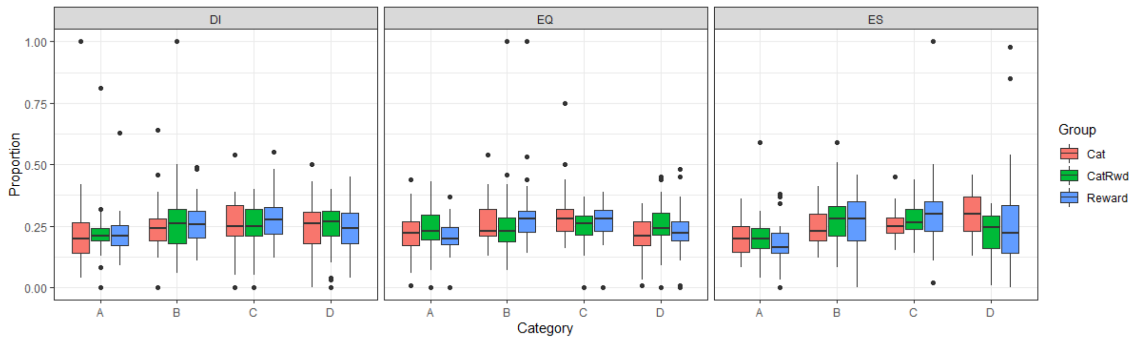

Figure 4.6 Proportion of transfer phase selections for each category cluster by feedback type and reward structure.

For the participants given categorical feedback, the simulations predicted a preference of category D selections since the lack of reward information likely had participants rely on stimulus similarity. In the categorical feedback data, there was no evidence of an overall difference in the proportions of category selections (BF_{10}= 1.423), nor between reward structures (BF_{10} = .027). Upon closer inspection, the Cat-DI participants selected each category in near equal proportions (M ~ .25; BF_{10}= .057). However, both the Cat-EQ (BF_{10}= 6.733) and Cat-ES (BF_{10}= 53.464) both show evidence of differences given the current data. In post hoc testing, should we run this same task again, choosing category A is expected to be less frequent than choosing category C overall (BF_{10}= 11.531). Thus, while no overall differences in category selections were observed in the behavioral transfer phase data, there is evidence that the least central category cluster was not a preferable choice in the transfer phase.

For reward feedback, the simulations predicted a large preference for category D in the ES condition, while category A was predicted to be preferred in the DI condition. Both being the most rewarded category in their respective conditions. Overall, the categories were selected in differing proportions (Log(BF) = 8.805), but the differences were not consistent between reward structures (Log(BF) = -3.556). For the Rwd-DI participants, there is no substantial evidence for either the alternative or null hypothesis for a difference in category selection proportions (BF_{10}= 1.425), however, analysis of both the EQ and ES reward structure conditions detailed that the proportions of category selections did differ (BF_{10}= 9.826 and BF_{10}= 3.466, respectively). In these conditions, categories B (M = .278), C (M = .286), and D (M = .246) were selected near equally often more as compared to category A (M = .190; all BF_{10} > 2.681). These results detail that there was a slight preference for the more rewarded categories in the ES and EQ conditions as predicted, but no preference for any category in the DI condition. Thus, it is possible that reward information and stimulus similarity may been contrasting information for the DI participants, but compounding for the ES participants.

For the catrwd feedback participants, the models predicted that choice behavior would likely be similar to the combination of both prior feedback types: no preference for any category in the DI condition, but a slight preference for category D in the ES condition. Overall category selections in the test phase differed were shown to differ (BF_{10} = 14.732). As predicted, the CatRwd-DI participants category selection proportions showed no evidence of differing given the data (BF_{10} = .170). Similarly, the proportions at which each category was selected in the CatRwd-EQ condition were also near equal (BF_{10} = .090). Conversely, for the CatRwd-ES participants, there was evidence of differing category selections (BF_{10} = 7.698), however, the preference was not for category D as expected. Instead, both category A (M = .210) and category D

(.225) were chosen less frequently as compared to both categories B (M = .284) and C (.281) via post hoc testing (all $BF_{10} > 3.723$). This suggest that the combination of feedback, or attention to one feedback over the other based on the findings of Study 1, may have had an impact on choice.

4.3.2. Theoretical Analyses

To compare how well each of the computational models fit the participant data, both between feedback types and between models, the Bayesian Information Criterion (BIC; Schwarz, 1978) was computed for each participants individual model fit. BIC values are more conservative, as compared to AIC (Akaike's Information Criterion; Akaike, 1974), in that the BIC calculation more heavily penalizes models with larger numbers of free parameters. Like AIC however, lower BIC values are indicative of a better fit of the data which suggest that the decision-making processes utilized by the participants are likely to be similar to that of the best fitting model.

4.3.2.1. Model Fitting

Overall, each of the base computational models (CatLearner, RewardLearner, and CatRwdLearner) fit the data better than a chance selection model (BIC ~ 834.704). Though, there is variation in how well each model fit each of the conditions as detailed in Figure 4.7 below. In comparing the BIC values between models, we use the Bayes Factor calculation method described by Nagin (1999): $BF_{ij} = \exp\,(BIC_j - BIC_i)$. Where the BIC values for groups *i* and *j* are compared. Values greater than 10 indicate strong evidence for model i, and values less than 1/10 indicate strong evidence that model *j* is a better fit.

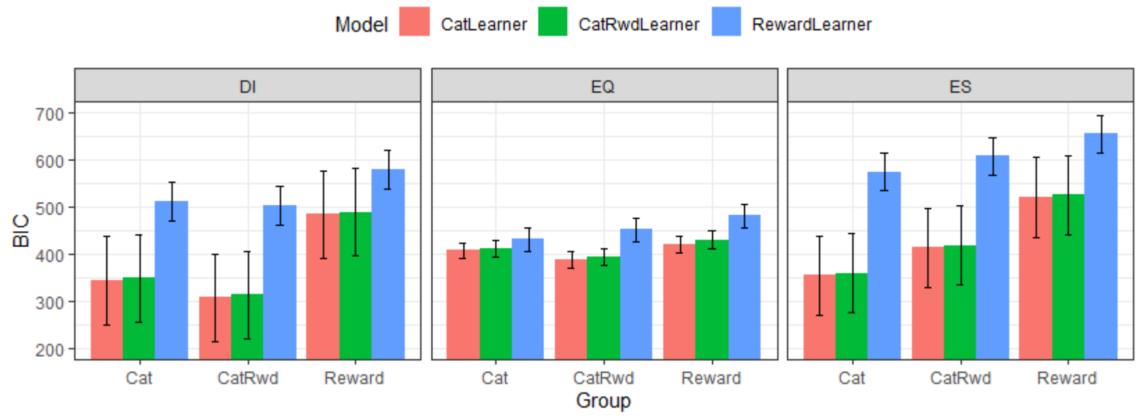

Figure 4.7 Mean BIC values for each models' fit of each feedback types' datal, split by reward structure. Error bars represent the standard deviation about the mean.

Since the behavioral data detailed no overall differences between reward structures, the following analyses focused on the overall differences in model fits between feedback types. When collapsing across reward structure conditions, the category-learner model was the best fitting model for the categorical feedback participants as compared to the reward-learner model (Log(BF) = 101.971) and the catrwd-learner model (Log(BF) = 5.851). Additionally, as detailed by Figure 4.6B, the category-learner was also the best fitting model for both catrwd (Log(BF) = -6.144) and reward (Log(BF) = -44.753) feedback participants. This suggest that stimulus similarity had a stronger effect on behavior than reward based on the theoretical distinctions of both the category- and reward-learner models: with the category-learner updating exemplar weights based on the similarity of observed stimuli and the reward-learner updating through rewarding outcomes. This was verified by looking at the q-values of the catrwd-learner (the free parameter denoting attention towards categorical information or reward feedback). Both categorical ($\bar{q} = .122$) and reward ($\bar{q} = .125$) feedback participants are assumed to give more attention to categorical, rather than numerical, information. For reference, q-values equal to 0 represents attention only towards categorical similarity whereas values of 1 are indicative of a focus only on reward information.

4.3.2.2. Post hoc Simulations

Since little differences between transfer phase category selections were observed, a potential explanation for the lack of an effect could be attributed to the scarce number of transfer trial stimuli. In the current paradigm, the participants only viewed 100 stimuli evenly spread across a stimulus space with feature values which ranged from 0-300. Thus, each transfer phase stimulus only differed in length and orientation by intervals of 15 in relation to other stimuli. It is possible that given a larger amount of transfer stimuli, the behavior predicted by the model simulations may be more apparent. Unfortunately, without the original participants, their current data and the computational models must be relied upon instead.

To ensure that the models made accurate predictions of the participants actual transfer phase category selections, the actual category proportion for each participant was compared to the to the proportions predicted by their data via their respective models (i.e. categorical feedback participants were compared to the predictions of the category-learner). Pair-wise Bayesian *t*-tests, using a Cauchy prior and an alternative hypothesis that two proportions differ, were utilized to determine if there was a difference between the actual and simulated data. The outcomes of each of these tests, for each combination of feedback type and reward structure, can be viewed in Table 4.1 below.

Table 4.1 Bayes Factors of the Simulated/Actual Proportion Differences in Study 2

Feedback	Reward Structure	Category A	Category B	Category C	Category D
Category	*DI*	.412	.192	.219	**22.275**
	EQ	.565	.207	.999	.351
	ES	.287	.211	.805	.210
Reward	*DI*	**>100**	.221	1.534	**>100**
	EQ	.449	.878	.213	.392
	ES	.651	1.280	.318	**16.530**
CatRwd	*DI*	.205	.505	.213	**12.467**
	EQ	.344	.325	.301	.224
	ES	.336	2.291	.248	**7.231**

Overall, the models seemed to accurately predict the participant data, barring five tests out of 36 where evidence suggested that the data differed. Thus, to determine the expected choice behavior to a larger range of transfer phase stimuli, 28 equidistant values of length and orientation were defined which resulted in 784 unique stimuli as compared to the 100 stimuli used in the current task. The participants' best fitting model parameters from each of the three base ALCOVE models, and exemplar node weights from the training phase of the task, were used to simulate predicted behavior on the expanded version of this task. Figure 4.8 details the aggregate predicted proportions that each category would be selected by feedback type, reward structure, and model.

However, caution should be used when interpreting the category predictions where the outcomes between actual categorizations, and the prior predicted categorizations, were incongruent.

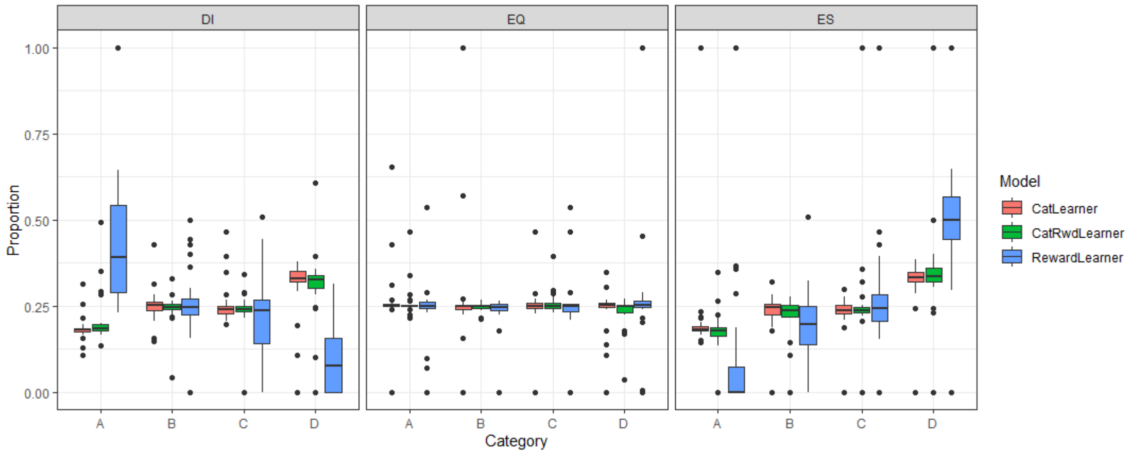

Figure 4.8 Post hoc simulated proportions that each category would be selected in the transfer phase by model and reward structure.

As can be surmised from Figure 4.8, there were indeed overall predicted differences in the proportions that each of the 784 stimuli were categorized (Log(BF) = 9.054). However, the predicted classifications are not expected to differ between feedback type (BF_{10} = .016), nor between reward structures overall (BF_{10} = .011). Though, the interaction terms Category Proportion X Reward Structure (Log(BF) = 50.478) and Category Proportion X Reward Structure X Feedback Type (Log(BF) = 82.541) suggest evidence of classification proportion differences between and within feedback conditions.

For the simulated categorical feedback data, the previously described differences are only expected to occur within the Cat-DI (Log(BF) = 25.559) and Cat-ES (Log(BF)

= 3.783) conditions. Consistent with our prior hypotheses, both predict that category D will have a larger probability of selection as compared to the other categories (all $BF_{10} >$ 9.747). Similarly, in the catrwd feedback participants, the differences in classification predictions are anticipated for the CatRwd-DI (Log(BF) = 9.350) and CatRwd-ES (Log(BF) = 9.772) conditions with category D being selected only slightly more than remaining categories (all Log(BF) > 3.338). Interestingly, for the reward feedback participants, while we find that the categorization proportions are predicted to differ in both the Rwd-ES (Log(BF) = 25.981) and Rwd-DI conditions (Log(BF) = 29.953) in terms of the overall preferred category. Consistent with our original hypotheses again, category D is expected to be preferred in the ES condition (all Log(BF) > 6.408) and category A in the DI conditions (all Log(BF) > 6.164). Additionally, across all feedback types, there were no differences in the proportion of category selections (BF_{10} = .019) or between feedback types (BF_{10} = .036) in the EQ reward structures which is expected as these conditions held both similarity and reward constant.

These predictions suggest, that given enough stimuli, the behavioral differences predicted by the initial simulations where the categorical-based feedback would result in a preference for category D, and reward feedback would bias choice towards either categories A or D depending on the reward structure, would likely occur.

4.4. Study 2 Discussion

In real-world learning, differing magnitudes of reward values often lead to representations of expected values that guide learning towards the maximization of outcomes (e.g. Rescorla & Wagner, 1972; Speekenbrink & Konstantinidis, 2015; Sutton

& Barto, 2018). Conversely, for category learning, stimulus similarity is an effective predictor of category learning performance (e.g. Medin & Smith, 1984; Nosofsky, 1988a; Pothos & Chater, 2002). In the current study, we created a paradigm which attempted to set reward magnitude against stimulus similarity by having multiple category clusters, which differed in degrees of total similarity to each of the test phase stimuli and had differing degrees of reward magnitudes for each cluster depending on the reward structure condition. It was predicted that reward feedback would lead to a differences in preferred category selections in the transfer phase of this task, with the DI conditions showing a preference for category A and the ES condition showing a preference for category D. Conversely, for categorical feedback, no differences were expected between reward structures, and only category D was expected to be preferred due to stimulus similarity.

From the behavioral results, people are indeed able to learn to classify each of the four categories well above chance during training, however, in line with the results of Study 1, numerical feedback resulted in markedly poorer performance as compared to both categorical and catrwd feedback. As detailed in Study 1, a potential cause for this difference may be the increased amount of variability in the outcomes of the numerical feedback versus the more discrete outcomes of categorical feedback. This idea is further compounded by the model fits of the catrwd-learner data showing a large preference towards categorical, rather than numerical, information resulting in categorization performance near identical to that of the categorical feedback participants. Additionally, these differences between feedback types were found to persist consistently across all trial blocks with numerical feedback reaching asymptote at a relatively suboptimal level. Interestingly, while this study used numerical information in an opposing respect, the current results are reflective of a study which assigned zero and nonzero costs to incorrect categorization attempts: participant who were not penalized for incorrect categorizations showed improved categorization performance relative to the participants who were charged a nonzero amount for incorrect choices (Maddox & Bohil, 2000). Interestingly, referring back to the Rwd-EQ training results, this was the only condition where the reward magnitude was the only major difference between category clusters. Participants in this condition were observed to exhibit the best training performance as compared to both the Rwd-ES and Rwd-DI conditions where both similarity and reward magnitude differed. It is possible that with the added difference in similarity, slightly more attention was given to the reward information, and thus resulted in poorer performance. Looking again to the q-values for each condition, it is plausible: Rwd-EQ ($\bar{q} = .064$), Rwd-ES ($\bar{q} = .142$), and Rwd-DI ($\bar{q} = .170$).

If stimulus similarity is indeed an influential factor in the determination of category membership (Barsalou, 1985; Homa & Cultice, 1984; Shepard, 1987), it is expected that category D would be selected in greater proportion during the transfer phase over the remaining categories since this category was the most central in the stimulus space, and thus most similar to a majority of transfer phase stimuli. In the current results, this was indeed the case as there was evidence of a preference for category D in the behavioral results of the participants who received categorical feedback information in the ES and DI reward structure conditions. For categorical feedback, these conditions were equated since no reward information was delivered. However, for the participants who received numerical feedback, categories C and D were the most preferred proportionally in the ES condition where the largest rewards were given for these two categories during training, but no overall differences in category selections were found in the DI condition when the least similar stimuli were the most rewarded. It is the singular condition outside of the EQ rewards structure conditions where we find this lack of preference. However, despite the previously superior performance by the categorical feedback participants, the proportions that each category that was selected during the transfer phase were near identical, thus providing evidence of consistent generalization across feedback types.

A potential explanation for the current behavioral results could lie in classification automaticity, a process where stimuli are classified at a quicker rate with lessened cognitive impact as the task is learned (i.e autopilot; Palmeri, 1997). In the case of the ES and EQ conditions, positive feedback, regardless of the form of the feedback, may have been enough to associate certain lengths and orientations with a particular category. Thus, when asked to classify the transfer trials, the impact of the numerical information may have been lessened. Similarly, from the reward learning perspective, the participants may have simply generalized the learned rewards to the most similar stimuli (Kahnt et al., 2012; Wimmer et al., 2012). It is possible that the act of receiving a reward was sufficient in defining the association between stimulus and reward. However, both explanations account for the ES condition, but fail to account for the DI condition for numerical feedback.

For the ES reward structure, the results interestingly seem to conform to the findings of Schleglemitch and von Helverson (2020) where, despite the difference degrees of reward values given for stimulus categorization, discrete and variable numerical feedback were shown to have no influence on the transfer phase categorization performance. As detailed in Study 1, discrete numerical feedback results in similar categorization behavior when compared to categorical feedback. Conversely, for the DI condition, outside of the post hoc simulations, there is no indication of a clear preference for category A despite it being the most rewarded. In this condition alone, the behavioral data does differ from that of the categorical and catrwd feedback participants. This finding is in contrast to most expectations of reward learning which detail that choice is biased towards reward maximization in most instances and should have resulted in a preference for category A, and it is also incongruent with the expectations of stimulus similarity as the behavior differs from that of the categorical participants who preferred category D.

Thus, we find an interesting detraction from the expectations of both theories, suggesting that numerical information may have overridden the impact of stimulus similarity, but not enough to result in a clear preference in categorical choice. Further, we find that when stimulus similarity and reward magnitude are not in opposition, transfer phase category preferences between feedback types are near equated and show a defined preference for the most similar, and most rewarded, categories.

5. STUDY 3

5.1. Introduction

While Study 2 addressed the effect of variable reward magnitudes in category learning, Study 3 aims to explore the impact of another facet of reward learning on category learning: reward frequency. With most learning being the result of repetitious observations of choice and outcome pairings (e.g. Bornstein et al., 2017), one might ask how well people are able to learn from sparse knowledge or infrequent observations. In some cases, such as touching a hot stove, it may only take one event to understand that an oven can inflict pain if someone is not careful, whereas in situations such as determining which stores have the best prices, or trying to study vocabulary for an exam, knowledge is likely acquired over a certain time period.

Likewise, in most category learning paradigms, opposing categories are most often equally observed, even in paradigms utilizing transfer phases (e.g. Seger et al., 2015). According to the previously described models of category learning, as stimuli are observed along with their outcomes, the model exemplars eventually become associated with one category over another as a function of their weights that are incrementally adjusted over time (Kruschke, 1992; Nosofsky et al., 1994). Thus, effective predictions as to what category a novel stimulus belongs to can be made given sufficient knowledge. However, in instances where the frequency of category observations differ, categorization behavior is expected to change (Estes, 1986; Nosofsky, 1988b).

These differences that observational frequency, or base rates, have on human decision making have been investigated for a few decades now (e.g. Don & Livesey, 2017; Estes, 1976; Kalish, 2001; Maddox, 2002; Medin & Edelson, 1988; Tversky & Kahneman, 1981). In the purview of category learning, altering the base rates of certain categories can lead to an inverse base-rate effect where people tend to categorize novel stimuli into low-frequency categories at a rate much higher than the underlying base-rate would predict (Kruschke, 1996). Taking an example from Winman et al. (2005), in a cue-association task, if participants are asked to determine which symptoms predict certain diseases, and some diseases are rarer than others, when a common disease predictor is compared against a rare disease predictor, people tend to choose the rarer disease. Thus, despite the intuition that the common diseases are more probable than rare ones, people tend to neglect base-rate information in favor of the information gained from specific, rather than aggregate, cases (Don et al., 2021).

This line of research sparked controversy in the field with multitudes of researchers attempting to determine a possible mechanism to explain the effect (cf. Koehler, 1996). In category learning paradigms specifically, some believed that it is due to differentiation in the learned associations between stimuli/category pairs (e.g. Gluck & Bower, 1988), some argued about whether it was the result of rule-based processing (Lamberts & Kent, 2007; Winman et al., 2005), whereas others argued as to whether or not people were sensitive to base-rate information (e.g. Bar-Hillel & Fischhoff, 1981; Kahneman & Tversky, 1973). Kruschke (1996) attempted to reconcile some of the prominent theories at the time with two theses: first, people learn differing frequency categories at different rates, so what people learn about the rare categories is dependent on what they've already learned about the common categories; second, common and rare categories are encoded differently, thus biasing decision-making by modulating base-rates can lead to results which appear to be inconsistent. Essentially, what is known about an infrequent category is likely based off of what is known about the more frequently observed categories, and this can lead to counterintuitive categorization behavior.

In reward learning, we see almost the opposite effect. Herrnstein's *Matching Law* (Herrnstein, 1974), which states that choice behavior among alternatives is reflective of the proportion of reinforcement each alternative has received, provides a valid explanation of the observable behavior for a variety of reward learning paradigms (Davison & McCarthy, 2016; for a review). For example, in multi-alternative bandit tasks, people will consistently explore, and exploit, given alternatives in an effort to maximize the overall number of rewarding outcomes or alternative with the maximum reward (Daw et al., 2006; Racey et al., 2011; Speekenbrink & Konstantinidis, 2015). In these tasks, behavior often conforms to the tendency of melioration—preferring the better alternative overall, all things equal, but when the alternatives' ratios differ, alternatives which maximize the rate of reinforcement will be preferred instead (Herrnstein & Prelec, 1991; Sims et al., 2013; Vaughan, 1981). Therefore, if a stimulus is observed, or reinforced, more frequently, it is more likely to be chosen in the future. Further, in a study by Don et al. (2019), people were observed to forego alternatives with larger potential payments for alternatives that were more frequently rewarding. Concisely, choosing the option that had the largest cumulative reward over the option with the largest average reward.

The question then becomes, what is the difference between category and reward learning? In both types of task, there is evidence that when all things are considered equal, either the most similar category or most rewarding alternatives will be preferred. However, there are differences between each form of learning which can be simplified to differences in task framing (e.g. Radulescu et al., 2019). When learning categories, there is a degree of similarity between stimuli that can be relied upon to guide learning, whereas in reward learning, the stimuli are fairly discrete or non-existent. In both type of tasks, the categories and options are often discrete (i.e. 'Options [1,2,3,4]' or 'Category [A,B,C,D]'), but the stimuli are variable in different respects. Thus, a present question is how participants would respond when given different forms of feedback on a category learning task with differing degrees of stimulus similarity and differing rates of category observation frequency. Would the inverse base-rate effect be observed, or will people show a preference for the most frequently observed categories? Further, would categorical or numerical feedback bias decision-making towards either of these effects? In the following section, a series of computational models were used to simulate participant behavior and gain insight into what would be expected from real participants behavior.

5.1.1. Simulated Results and Hypotheses

As completed in prior studies, the three types of feedback were employed to determine the impact of categorical and numerical information, stimulus similarity, and frequency on category learning. In the following simulations, the agents were shown a 300-stimuli training phase where the 300 stimuli were unequally allocated to 4 category clusters which either radiated from the center of the stimulus space (SpreadA and SpreadD conditions) or were equidistant from both the other categories and the stimulus space (EQ). The category structure was near identical to the category structures presented in Study 2; only differing in the frequency that each of the four category clusters were observed. Additionally, like in Study 2, each of the category clusters were distinct with a high degree of intra-category similarity to emulate the aspects of a reward learning task. Figure 5.1 below details the category clusters for each condition as well as the design of the transfer phase stimuli, which remained identical to the design detailed in Study 2.

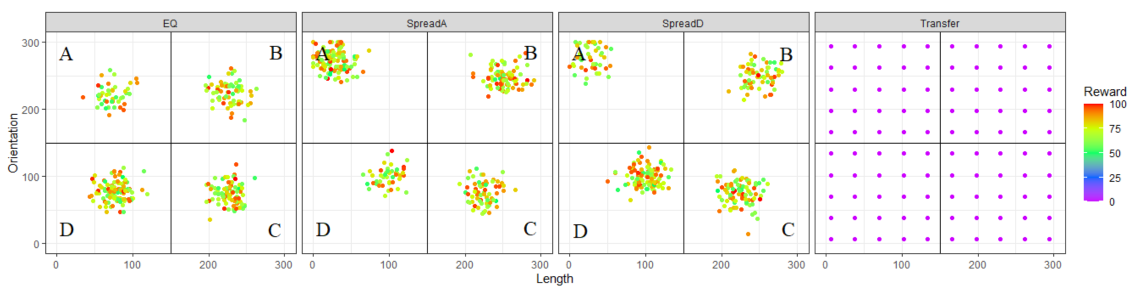

Figure 5.1 Category structures and individual stimuli observed in each condition in Study 3, including the transfer phase stimuli.

Upon categorizing all training phase stimuli, each agent, regardless of condition, then categorized the same 100 novel stimuli. These stimuli were created from equally distributed values of length and orientation combinations from across the stimulus space (Figure 5.1-Transfer). In this paradigm one category cluster was observed most frequently with all other categories being observed at an increasingly infrequent rate. In the SpreadA condition, A was the most frequent; in the SpreadD and EQ condition, category D was the most frequent. In each condition, the most frequently observed category was seen almost twice as many times as the least frequent category. . Thus, if categorization behavior in this task were to follow the inverse base-rate effect, the least frequent category was expected to be selected most often in the transfer phase. However, should the behavior observed in the reward learning literature present itself, the more frequent categories should have accounted for the majority of test phase selections in the numerical reward feedback conditions. Though, as observed in the prior studies, stimulus similarity may have had a unique impact on the results.

Importantly, for each condition in training, the ‘best’ choice was defined as choosing the correct category in the training phase. In the transfer phase, however, the best choice differed by condition. In the following simulations, the best choice for the SpreadD condition was defined as choosing category D as it was both the most frequent category, and the closest to the center of the stimulus space, during training. In the SpreadA condition, the best choice was defined as choosing category A since it was the most frequently observed in training. In the EQ condition, since each category was equidistant to the center of the stimulus space, category D selections were considered to be the best choice as category D was also the most frequently observed category in the EQ condition. In addition to the models used in both Studies 1 and 2, the ALCOVE-Decay model variants were also employed in the following simulations. Figure 5.2 below details the predicted proportions of best choices by task phase, model, model variant, and condition.

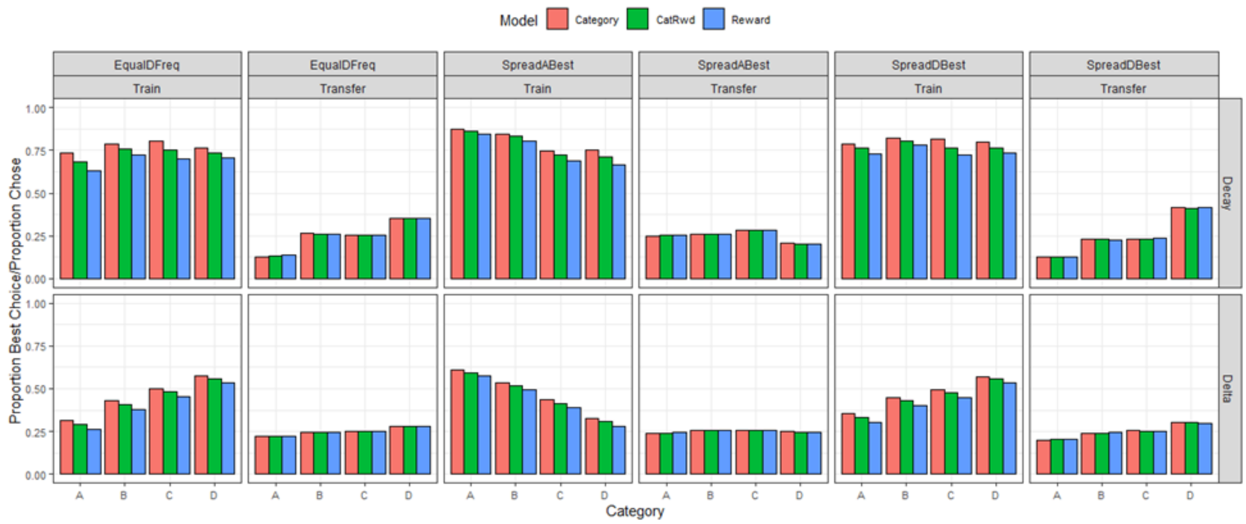

Figure 5.2 Plot of predicted correct categorizations in the training phase, and predicted proportion of best choices in the transfer phase, for category clusters A through D, by each of the three conditions. The top row details the ALCOVE-Decay model predictions, and the second row details the predictions of the base ALCOVE models.

The models again predicted slightly better performance for agents that were given categorical feedback in each condition during training. This was likely due to the inclusion of uniformly distributed rewards that are essentially uninformative. However, the models did predict some frequency-based differences in the transfer phase. Interestingly, in the base ALCOVE models, each model almost uniformly predicted an equal number of category selections in the transfer phase between conditions. In the decay variants of ALCOVE, there was a noticeable preference for the more frequent categories in SpreadD and EQ, but near equal category selections in SpreadA. This suggests that the feedback type and the frequency that each category was observed were a strong determinant in choice. This is especially true when the category clusters were spread throughout the psychological space and the closest category cluster to the center of the stimulus space was the most frequently observed as shown the model predictions for the SpreadD and EQ conditions where category D was selected well over chance (25% for four options).

Interestingly, similar to Study 2 where performance and similarity seemed to have an interactive effect, it seemed that similarity also had a substantial impact based on the model predictions for the SpreadA and SpreadD conditions. In the SpreadD condition, where the most frequent category is the most similar to the majority of the transfer phase stimuli, frequency and similarity may have a compounding effect. However, in the SpreadA condition, where the most frequent category was the least similar, the proportion that each category was selected in transfer was predicted to be near chance between feedback types. Additionally, in line with the results of Don et al. (2019), for the conditions with differing frequencies of category observations, the decay model predicted a greater preference for the more frequently seen categories overall as compared to the agents that utilized a delta-based updating method within the same conditions.

Thus, given these simulated predictions, it was hypothesized that the frequency that each category cluster was observed in training would be predictive of the subsequent transfer phase stimuli. Based on the decay variants of ALCOVE, in the SpreadD and EQ conditions, the participants are expected to choose category D in a greater proportion as compared to the remaining three categories. However, in the SpreadA condition, where the most frequent category was also the most distant from the center of the stimulus space, the participants are expected to show no differences in the proportion that each category is selected during the test phase which is likely due to the influence of stimulus similarity. Should participants rather follow delta-based learning processes, minimal differences between conditions and test phase category selections are expected instead.

5.2. Method

5.2.1. Participants

All participants in this study were recruited from Texas A&M University and were given partial course completion credit for their participation. Each participant completed an online version of the informed consent and experimental task. In following the same study conditions simulated above, 30 participants were recruited for each of the 9 conditions based on a sample size calculation using 80% power with a moderate effect.

In total, we recruited 29 for the Categorical-SpreadA, 33 for Categorical-SpreadD, and 32 for Categorical-EQ for the categorical feedback group; 30 for Reward-SpreadA, 31 for Reward-SpreadD, and 29 for Reward-EQ; and for the CatRwd feedback conditions we recruited 31 for CatRwd-SpreadA, 27 for CatRwd-SpreadD, and 29 for CatRwd-EQ. Overall, we recruited 271 participants from an online college-aged sample. Given the number of participants we recruited, using a moderate effect size ($\eta^2 = .059$) and alpha error value of 0.05, we achieved post hoc power of 0.842.

5.2.2. Task and Structure

5.2.2.1. Experimental Task

This task consisted of two distinct training and test phases, similar in structure to the paradigm described in Study 2. In the training phase, participants were shown 300 unique line stimuli unequally split amongst four category clusters. In the subsequent test phase, participants observed 100 more unique line stimuli that were created from the combination of 10 equally distanced line lengths and orientations. The test phase was designed to measure how well participants were able to generalize their learned knowledge to new, but similar stimuli, from across the stimulus space.

Depending on the feedback type each participant was assigned to, they received either categorical feedback (Category), variable reward-based feedback (Reward), or a combination of both types of feedback (CatRwd) in response to their choices during the initial training phase. During the test phase however, no feedback of any form was given to the participants. Similar to the prior studies, the terminology, or framing, in each feedback condition differed by feedback type. For the participants given Reward feedback, the participants were informed to use the line to aid them in predicting which option would give the largest reward value. The remaining two feedback conditions simply asked the participants to categorize the stimuli and utilize the feedback to learn.

5.2.2.2. Category Structures and Rewards

As briefly mentioned above with the simulations, each condition of this task consisted of four training category clusters and a test phase of evenly distributed stimuli. Following an identical construction of the structures used in Study 2, the SpreadA and SpreadD category clusters radiated from the center of the stimulus space with category D being the closest to the center to category A being farthest. In the EQ condition, identical to Study 2 as well, each of the four category clusters were equidistant from the center of the stimulus space and each of the other clusters (Figure 5.1). Regardless of the condition, each of the four category clusters were drawn from a bivariate normal distribution with a standard deviation of 15. For the stimuli frequency manipulation, the most frequent category cluster, either category A or D depending on the condition, was created from 105 independent samples/observations with each subsequent cluster shaving a reduced frequency of 20 observations (a difference of 80 observations between the most and least frequent clusters). The rewards in each category cluster were uniformly distributed, $U(50,100)$, amongst all training stimuli. Figure 5.1 above details the differences in category structure, observation frequency, and uniform reward for each of the three structure conditions.

5.2.3. Task Procedure

The task procedure in this study was identical to the procedure detailed in the general method section and Study 2. Upon giving consent to taking part in this study online, each participant was randomly assigned to one of the 9 conditions and given adequate instructions on how to progress in the experiment. Each participant repeatedly viewed a single, unique, line stimulus from random category on each trial. On each trial, concurrent with the line stimuli, a prompt relevant to each feedback type directed the participants to use the four buttons at the bottom of the screen to make their selection. The labels of these buttons were also dependent on the feedback type the participant was assigned to as well (Option X or Category X). Participants would make their decision about which category/option they thought matched the line stimulus on screen, and then receive relevant feedback for a period of 2 seconds before a 0.5s intertrial interval screen that served as a transition between each individual trial as shown in Figure 5.3.

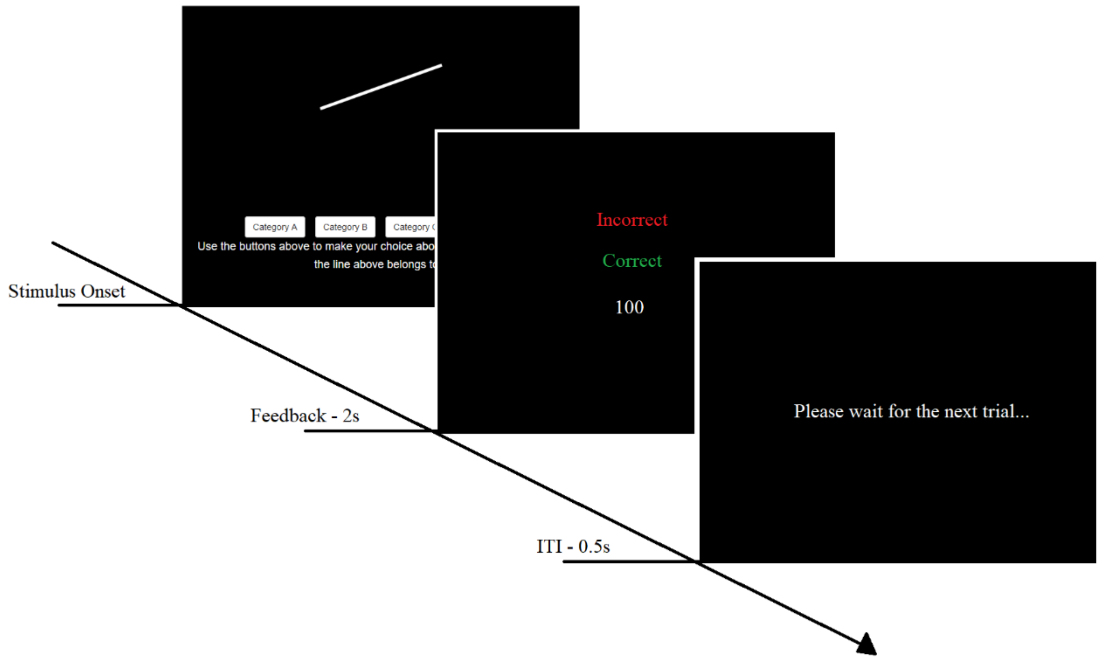

Figure 5.3 Trial diagram for Study 3. Participants make their selections, then receive one form of feedback before observing an intertrial interval screen before the next stimulus onset.

The training phase of this task was divided into 100-trial sections separated by a break screen. Upon completion of the training phase, participants were shown a screen detailing the instructions for the test phase. Each participant was informed that they would again see new, unique, stimuli, but that they would no longer receive feedback about their decisions. After completing all 400 trials, participants were debriefed, which included information about the task and the goals of the experiment. They were then directed to a screen where they could claim their credit for online participation.

5.3. Results

5.3.1. Behavioral Results

5.3.1.1. Training Phase Results

5.3.1.1.1. Learning Over Time

In the training phase, each of the four category clusters were observed to have been individually learned by the participants to some degree based on Figure 5.3B below showing final trial block proportions of best choice. However, based on both plots in Figure 5.4, there are noticeable differences in the performance of the participants who received reward feedback as compared to the participants who received some form of categorical feedback. For consistency with the results reported prior in Studies 1 and 2, all of the following results will be reported in Bayesian terms.

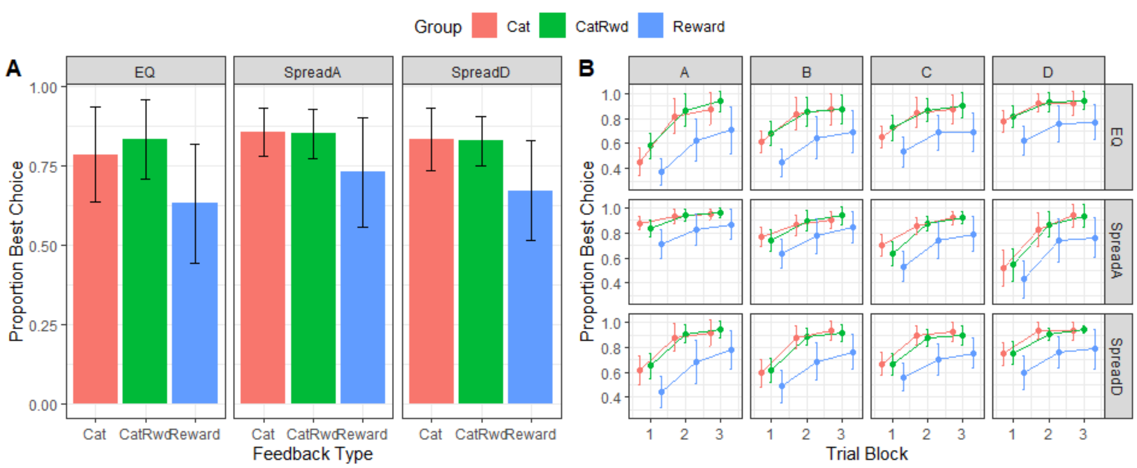

Figure 5.4 Mean behavioral data by feedback type and category structure during the training phase. A.) Mean overall correct categorization/best responses during the training phase. B.) Correct categorizations/best responses across trial blocks. Error bars represent the standard deviation of the mean.

In a Bayesian Multilevel model with feedback type, category structure, and trial block as factors, and participant number and trial block and the random intercept and slope respectively, we determined whether overall learning differed between feedback types and category structures over time. With categorical feedback as the reference group, and collapsing across category structures, catrwd feedback shows evidence of improved learning (.100, [.028, .169]) over categorical feedback, whereas reward feedback is observed to be poorer than both categorical and catrwd feedback (-.124, [-.191, -.057]). In comparison to the EQ category condition, collapsing across feedback types, both the SpreadA (.094, [.029, .159]) and SpreadD (.105, [.040, .168]) category structures showed moderately better performance in the categorization of training trial stimuli. Across all feedback types and reward structures, there is evidence that the participants did indeed learn to correctly categorize the stimuli. Overall, the rate at which participants learned was fairly consistent between conditions (-.003, [-.055, .047]).

With the differences between feedback type and category structure identified, the following analysis examined the possible impact that category frequency may have had on how well each category was learned over time. From the trendlines detailed in Figure 5.5b, there is visual evidence that the more frequently that a category was observed in training, the quicker that category learned based on the proportion of correct categorizations within each trial block. Statistically, this was the case when collapsing across each feedback type. When comparing the most frequent category, category A, in the SpreadA condition there is evidence that it was learned at a faster rate overall than category A in the SpreadD (.067, [.042, .093]) and EQ (.087, [.062, .113]) conditions where category A was the least frequent. The inverse of this relationship was also observed for category D which was the least frequent for the SpreadA condition, but was the most frequent for the SpreadD (.100, [.059 .142]) and EQ (.086, [.044, .128]) conditions and thus learned at an accelerated rate. However, as detailed by the overall proportion of correct categorization, the reward feedback participants showed consistently poorer learning over time (-.126, [-.149, -.103]), but closely followed the same learning curve as the other feedback types (-.001, [-.015, .014]). Though, as detailed by Figure 5.5b, each category, in each condition was learned to a relatively similar asymptote regardless of the frequency manipulation.

5.3.1.1.2. Reaction Time Analyses

Much like Study 2, there were many aberrant reaction time values amongst the participants; likely due to the online format of this study having less environmental control. To attempt to account for this, we again report all of the following analyses using the median RT values for each participant. Figure 5.5 below details the median RT values for each feedback type/category structure condition by 100-trial block.

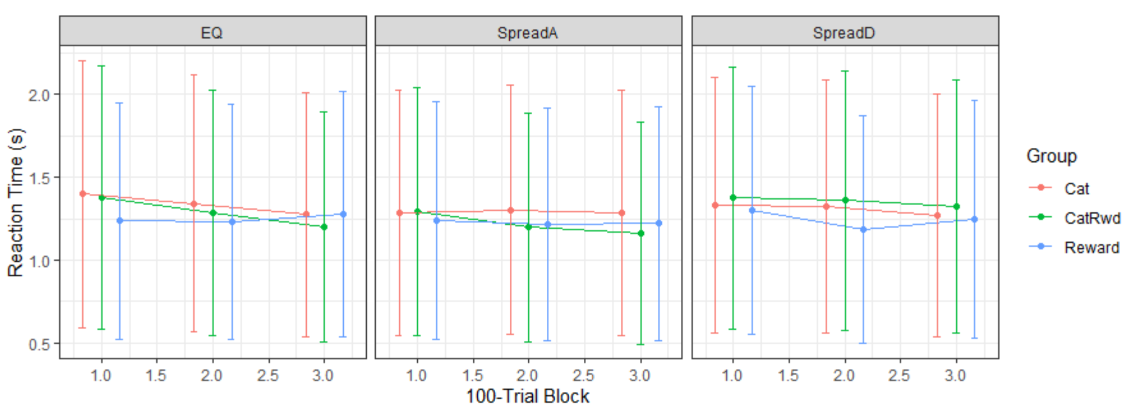

Figure 5.5 Median reaction times across trial blocks for each feedback type and category structure. Error bars represent the interquartile range.

As can be derived from the figure above, there was no evidence that the overall reaction time decreased over time (-.043, [-.130, .047]). Additionally, there were no differences between categorical feedback reaction times and either reward (-.060, [-.205, .085]) and catrwd (.023, [-.123, .171]) feedback, nor differences between reward and catrwd feedback reaction times (-.004, [-.144, .138]). Further, neither SpreadA (.053, [-.086, .195]) or SpreadD (-.010, [-.139, .126]) differed in median reaction times as compared to the EQ category structure condition. Reaction times also did not differ between SpreadA and SpreadD category structures (-.068, [-.210, .069]).

5.3.1.2. Test Phase Results

While the training phase results focused more on how well each individual category cluster was learned, the test phase results focused on the proportions at which each of the 100 test phase stimuli were classified into each of the four previously learned categories. In Figure 5.6 below, the proportion that each of the test phase stimuli were categorized into each of the categories is detailed, grouping by either feedback type or category structure. For the following results, the log of the Bayes Factors for each analysis will be reported along with any relevant parameter values.

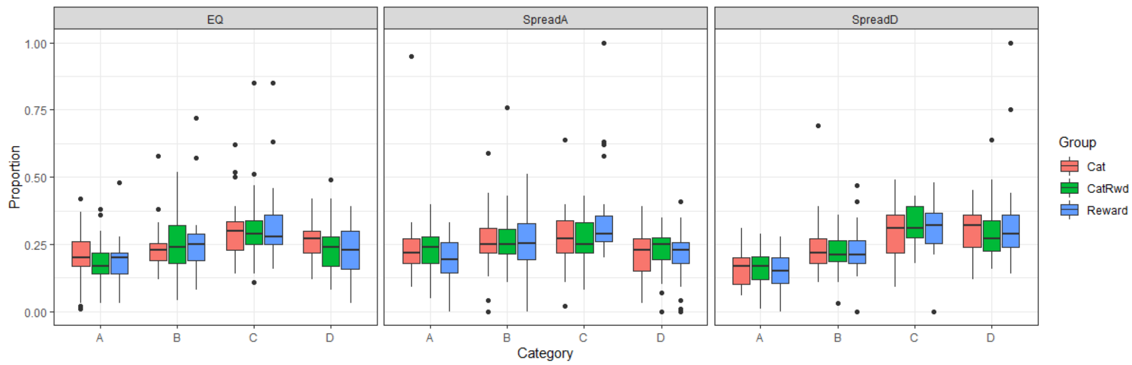

Figure 5.6 Mean proportion of category selections in the transfer phase by feedback type and category structure.

To determine if feedback type had an impact on the proportion that each category was selected, the data was collapsed across category structures. A Bayesian ANOVA detailed that category selections did not differ between feedback types overall (Log(BF) = -4.524), but the individual proportions for categories A-D did differ from each other (Log(BF) = 65.956). Similarly, to determine if the category structures prompted differences in behavior, the data was collapsed across feedback types and showed no differences in the proportions that each category was selected between category structures (Log(BF) = -4.566). These findings were consistent with the initial model simulations for both the delta- and decay-based ALCOVE variants which detailed no major differences between aggregate category selections based on either feedback type or category structure overall.

However, both model variants did predict an increased proportion of category D selections in the SpreadD and EQ conditions where category D was the most frequently observed. Though, the predicted magnitude of the preference for category D differed between variants. In the SpreadD condition, collapsing across feedback type, category D

(~.308) was selected more frequently as compared to least frequent categories A (~.158; Log(BF) = 37.253) and B (~.227; Log(BF) = 9.393), but not category C which was the second most frequent category (~.307; Log(BF) = -1.822). This was partially reflective of the model predictions.

Interestingly, the proportions that each category was selected in the transfer phase of the EQ condition deviated from the model predictions. While the least frequent category, category A, was selected the least often (~.190) as predicted, the most frequently observed category, D, accounted for only 24.4% of the total category selections when collapsing across feedback types. In this condition, categories B (~.252) and C (.313) were chosen the most frequently, but only category C was chosen more frequently than category D (Log(BF) = 6.072). However, category D was still selected more frequently than category A (Log(BF) = 5.507).

The SpreadA model predictions, for both model variants, detailed that the transfer stimuli would be categorized at a near equal rate, but the least frequent category, D, would also be selected slightly less frequently than the other three categories. The behavioral data reflected these predictions to an extent. In this condition, the most frequent category, category A (~.224), did not account for the majority of the transfer phase selections. Categories B (~.265) and C (~.295) were selected at a slightly higher rate, and the least frequent category, D, was the least frequently selected as predicted. Overall, only category C, the second least frequent category, differed from the most frequent category: A (Log(BF) = 5.614). Whereas the most frequent, and least frequent categories, A and D respectively, were selected at an equal rate (Log(BF) = -1.630).

Overall, aside from the EQ condition, the behavioral data moderately reached the expectations given by the model simulations. The selection behavior in the SpreadD condition details that the frequency of category observations did indeed influence the proportion at which the transfer phase stimuli were attributed to categories C and D. However, in the EQ and SpreadA conditions, it is possible that stimulus similarity may have had a more significant impact on behavior than the frequency at which the stimuli were observed.

5.3.2. Theoretical Analyses

5.3.2.1. Model Fitting

As mentioned in the introduction, prior reward and category studies have shown how frequency of stimulus or reward observations impacts learning. More recently, studies have shown that differing reward frequencies can result in contrasting predictions in computational models which focus on either the average or cumulative reward (Don et al., 2019). The following results, we will detail how well the base ALCOVE models, and their decay model variants, individually fit the participant data, as well as determine if there are any differences in fit between both sets of models. Figure 5.7 below details the average BIC values for each task condition, feedback type, and model variant.

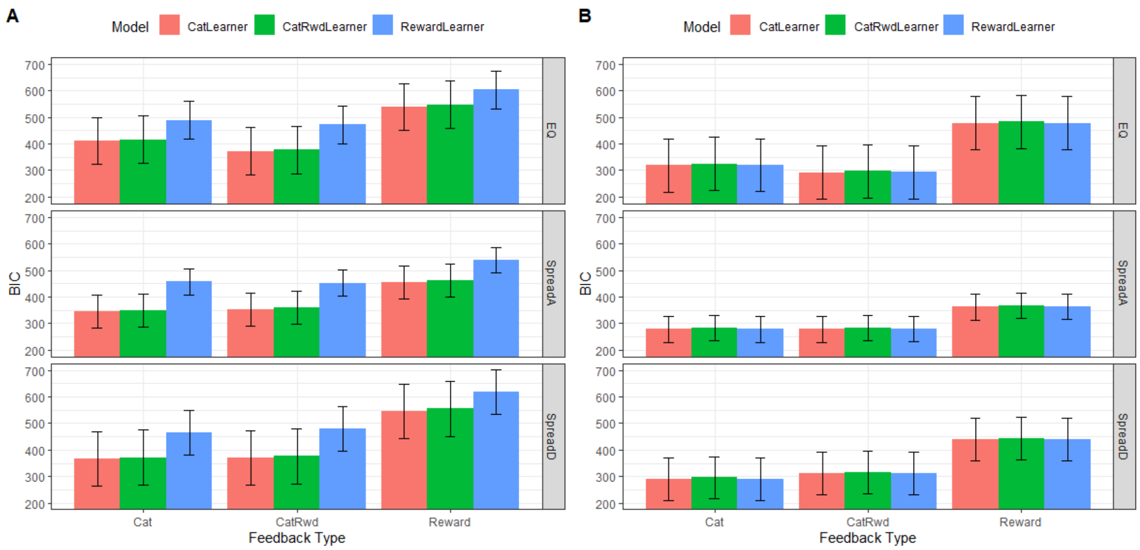

Figure 5.7 Mean BIC values for each models' fits of the participants' data. A.) BIC values for the delta-based ALCOVE models. B.) BIC values for the decay-based ALCOVE models. Error bars represent the standard deviation of the mean.

For Delta-rule ALCOVE variants, the average BIC values differed between the three models (Log(BF) = 14.989), and only moderately differed between category structure conditions (Log(BF) = 1.133). Additionally, the same trends were observed in Studies 1 and 2 where the reward-learner is the poorest fit to the data overall (rwd v. cat: Log(BF) = -90.483; reward v. catrwd: Log(BF) = -84.336). Following suit with the prior studies, and despite the closeness in fit shown in Figure 5.7, the category-learner still provides a better fit to the participant data as compared to the catrwd-learner (Log(BF) = 6.148).

In analyzing the fits for the decay model ALCOVE variants, there was again sufficient evidence that the models differed in their fits of the participant data (Log(BF) = 45.217) and in the average fits between category structure conditions (Log(BF) = 3.343). Congruent with the results of the prior two studies and models, the category-learner provided superior fits of the participant data as compared to the catrwd-learner (Figure 5.8; Log(BF) = 4.950). Interestingly, in these data, the overall fits of the reward-learner and the category-learner were near equated (Log(BF) = -.587) which suggested that when observational frequency differs, both categorical and numerical feedback are equally, potentially, effective feedback types in guiding category learning. Additionally, the fitted 'q' values of the catrwd-learner, which represented the attention given to either categorical or numerical feedback, were notably higher for reward ($\bar{q}$ = .399) and catrwd ($\bar{q}$ = .365) feedback in this paradigm as compared to the q-values reported for the reward ($\bar{q}$ = .255) and catrwd ($\bar{q}$ = .189) feedback groups in Study 1. This suggest that when the frequency of category observations is modulated, both categorical and numerical information is more heavily weighted.

Finally, based on Figure 5.8 above, the fits of the decay model variants of ALCOVE surpassed those of the delta-based (or base model) variants of ALCOVE. When the average BIC values for each feedback type were compared between model variants, collapsed across category structures, the decay-based variant fits were found to far exceed the fits of the delta-based ALCOVE variants: categorical feedback (Log(BF) = 110.633); reward feedback (Log(BF) = 112.970); catrwd feedback (Log(BF) = 106.034). For reference, as it pertains to these models, exemplar node weights in the base model of ALCOVE were incremented or decremented via a learning rate modulated prediction error, whereas the weights in the decay variant models were decayed at a constant rate throughout the task with the only reinforcement coming from the raw reward value of the chosen option. Thus, with the best fits of the data being provided by the ALCOVE-Decay models, it is possible that some form of category representation decay occurs during the category learning process.

5.3.2.2. Post hoc Simulation

Following suit with Study 2, post hoc simulations were conducted to explore the predicted behavior of participants should they have seen a larger amount of test phase stimuli. First, identical to Study 2, the models were assessed as to whether they would accurately predict participants transfer trial behavior based on their best fitting model parameters and exemplar node weights. The simulations utilized the same 100 stimuli the participants observed in the current paradigm and calculated the probability that each stimulus would be classified as a member of each category cluster. Finally, a pairwise Bayesian t-test determined whether the actual transfer phase data differed from the newly simulated data. In Table 5.1 below, the Bayes Factors for each comparison is reported for each feedback type/category structure combination. Additionally, while the ALCOVE-Decay model provided the best fit of the data, the results for both ALCOVE mode variants are reported for completeness.

Table 5.1 Bayes Factors of the Simulated/Actual Proportion Differences in Study 3

		Delta - Decay							
Feedback	**Structure**	**Category A**		**Category B**		**Category C**		**Category D**	
Category	*SpreadA*	.218	.260	.211	.211	.423	.213	1.592	.661
Delta	*SpreadD*	.358	2.258	.258	.211	2.220	**8.121**	**3.134**	**6.271**
	EQ	.270	.200	.203	.241	.856	.701	.398	.238
Reward	*SpreadA*	.867	.381	.236	.271	**16.742**	.833	.645	.241

	SpreadD	.890	.201	.253	.317	.232	.325	.218	.706
	EQ	.242	.206	.289	.297	**4.909**	.231	1.114	.202
CatRwd	*SpreadA*	.218	.221	.458	.504	.210	.220	3.339	.733
	SpreadD	.224	1.251	.259	.257	**32.741**	**237.442**	**5.287**	**54.032**
	EQ	.659	.410	.347	.222	.684	**4.753**	.523	.372

Note: For these analyses, a BF of 3 indicates that the simulated data does differ from the actual data, whereas a BF of 1/3 indicates that the models sufficiently produced the observed data. Values above in bold show the conditions and categories which were incorrectly modeled.

Interestingly, based on Table 5.1, the test comparing the actual data to the ALCOVE-Decay variants showed an equal number of divergent actual/simulated category proportion predictions (5/36) as compared to the base ALCOVE models (5/36). This suggests that both models were sufficiently able recreate the category selection proportions in the transfer phase behavioral data for most conditions. Since the decay variants provided the best fit of the training data, the following results will detail the predictions for the expanded range of stimuli utilizing the ALCOVE-Decay model parameters. However, as some of the models overpredicted the behavioral data for some categories, the following results should be interpreted with caution.

Identical to Study 2, the combination of equal intervals of length and orientation, ranging in value from 15 to 285, resulted in 784 test phase simulation stimuli. For reference, the participants in the current paradigm saw only 100 stimuli derived from the same range of feature values. For these simulations, the best fitting parameters and final exemplar weights, for each participant, were used to predict the classification probability for each of the new stimuli. The aggregated simulated results can be seen in Figure 5.8 below.

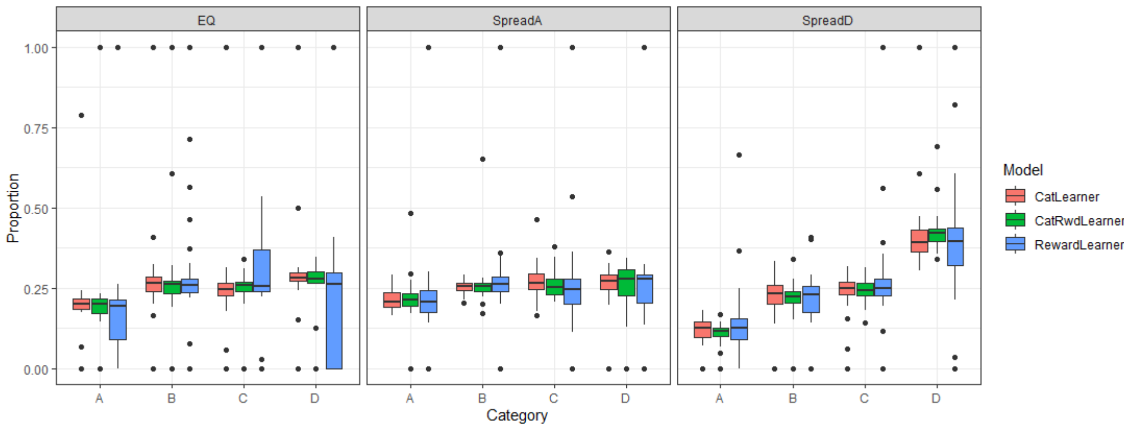

Figure 5.8 Post hoc predicted mean category selection proportions for the transfer phase data by model and category structure using the ALCOVE-Decay model variants.

To determine the effect that the expanded range of transfer phase stimuli would have potentially had on the participants, the transfer phase analyses in section 5.3.1.2 were recompleted using the simulated data. For the SpreadD condition, where category D was the most frequent and central to the stimulus space, the choice behavior was more indicative of the initial simulations. Collapsing across each feedback type, there was a distinct difference in the proportion that each category was selected (Log(BF) = 81.444), and category D (~.417) was selected in a much greater proportion as compared to the other three categories (all Log(BF) > 18.690). As compared to the behavioral results, the preference for the more frequently presented categories during training is more pronounced.

For the EQ condition data, where category D was the most frequent and all category clusters were equidistant, the proportions that each category is predicted to selected, given a greater range of stimuli did not change much as compared to the behavioral data. There was still evidence of a slight difference in the proportion that each category was selected (BF_{10}= 3.142), however it was only category A (~.194), the least frequently observed category, that was chosen to a lesser degree. Categories B-D are predicted to be selected in equivalent proportions (~.271) and distinct from category A (all BF_{10}> 4.504).

The predictions for the SpreadA condition, where category A was the most frequent and most distant from the center of the stimulus space, did change slightly from the results of the behavioral data. While overall there were no predicted differences in the proportion that each category was selected (BF_{10} = .656), categories B-D are predicted to be selected near equally (~.261), but only category B, the second most frequent category, was predicted to differ from category A (BF_{10} = 4.909). This differed from the behavioral results in that category A is expected to have the smallest proportion of transfer phase selections as opposed to category D. Thus, given a greater range of stimuli, it is possible that stimulus similarity may have larger effect.

5.4. Study 3 Discussion

In this category learning study, we utilized categorical, numerical, and hybrid feedback to determine if the different forms of information had an impact of how categories are learned. In addition, to prompt different behavior responses, we modulated the underlying category structures of each study condition and also manipulated the frequency at which stimuli from each category were observed by the participants. This resulted in each condition having one category that was the most frequently observed and one category that was the least frequently observed, and categories that were the most similar to the majority of the transfer phase stimuli. In both category and reward learning research, augmenting the rate at which stimuli are seen, or the probability that a certain stimulus results in a reward, has been shown to facilitate differences in choice behavior (e.g. Don et al., 2019; Nosofsky, 1988b). As such it was hypothesized that if participants were more sensitive to the frequency that the categories were observed during training, they would be more likely to attribute the majority of the transfer phase stimuli to that category. However, if stimulus similarity was more predictive of transfer phase choice, the most central category cluster should receive the largest proportion of responses. However, based on the model simulations, outside of the training phase, feedback type was not expected to influence category learning behavior.

Much like Studies 1 and 2, participants in each condition learned to correctly categorize the stimuli by the end of the training phase. Though, the trend detailed in the prior two studies where reward feedback produced the poorest training performance, was again observed in the current study. Additionally, while the paradigm of this study was supervised in terms of feedback delivery, the overall categorization performance in the training phase was interestingly congruent with that of semisupervised category learning research, where categories are either infrequently labeled or categorization results in sparse feedback (e.g. Vandist et al., 2009). In these paradigms, participants typically learn the categories well, but the rate at which they are learned, and the overall accuracy is thought to be dependent on the relative frequency that the feedback is received (McDonnell et al., 2012; Vandist et al., 2019).

Further, when looking at the rate at which each individual category was learned, there were distinct learning differences in the categories that were most frequently observed as compared to the remaining categories. Generally, the less frequent categories were learned at an overall slower rate over the first two blocks, as compared to the most frequently seen categories. However, by the end of the training phase, each category reached a comparable asymptote. These results seemed to parallel the findings of the category learning base rate research from which this study drew partial inspiration (Kruschke, 1996; Maddox, 2002; Maddox & Bohil, 1998; Nosofsky, 1988b). In these same studies, transfer phase choices followed the inverse base-rate effect where the least frequent/rare category was the most often selected. Thus, for the EQ condition, where the category clusters were equally similar to all transfer phase stimuli, this effect was expected to be observed in the behavioral data. However, this was not the case as the least frequent category was selected fewer times than the other three categories. Interestingly though, the most frequent category was selected relatively few times as well.

For the SpreadA and SpreadD, where similarity was expected to have an impact, the category selections in the SpreadD condition were shown to follow the base rates of the underlying categories regardless of feedback type. However, for the SpreadA condition, where the most frequent categories were the least similar to the majority of the transfer phase stimuli, a situation similar to the Rwd-DI condition in Study 2 was found where frequency may have overridden stimulus similarity to an extent, but ultimately resulted in no differences in category selections amongst each feedback type. Thus, given the current data, when stimulus similarity and frequency are corresponding, the generalization of categorical knowledge is likely to be reflective of the underlying base-rates, and the inverse base-rate effect may not be applicable. However, when similarity and frequency are conflicting, generalization may be more difficult and result in decision-making outcomes that deviate from the base-rates.

Based on recent work by Don et al. (2019), which detailed that people tend to forego larger rewards for smaller, more frequent, rewards, this task was designed to test whether more frequent categories would outweigh stimulus similarity. As already discussed, this is not the case. However, in their paper, they additionally reported that when frequency is a factor in decision making, behavior is likely better fit by the Decay model (Erev & Roth, 1998; Yechiam & Busemeyer, 2005) as opposed to the classical Delta rule models (Rescorla & Wagner, 1972; Widrow & Hoff, 1960). In the reward learning literature, the delta model encapsulates a wide variety of learning by assuming that choice is based on the maximization of rewards on any given trial, whereas the decay model assumes that behavior is defined by the maximization of the cumulative reward. Interestingly, the participant data was found to be best fitted by the ALCOVE-Decay model to a much greater degree than the delta-based ALCOVE models.

This is novel as no models of category learning, per current knowledge, explicitly assume that representations of category membership decay over time. For example, models such as COVIS (Ashby et al., 2011), ATRIUM (Erickson & Kruschke, 1998), and EXIT (Kruschke, 2001) tend to focus on explaining category learning behavior through prototype, exemplar, or connectionist constructs, while assuming that the relevant prototypes, exemplars, or nodes are updated through some form of attention weighted error. Having a decay-based model fit the participant data the best implies that for categories that are sparsely observed, people will likely have issues correctly classifying future instantiations given a long enough time period as the representation of that specific category will have decayed over time. A possible real-life scenario could be similar to finding out that domestic longhair cats are different from Maine Coons. If only one instance of a Maine Coon is observed as compared to the more common longhair cat, a person may be less likely to correctly categorize Maine Coons observed in the near future. However, this brings up an interesting point made by Kruschke (1996) that for less frequent categories, it may not necessarily be what you know about category Y, rather what it known about category X and how members of category Y are distinguished from the more commonly seen stimuli. Thus, based on this and the current results, it is also possible that the Maine Coon would be similar enough to longhair cats, that the distinctive features observed from a few instances would be sufficient for future classifications.

Further, category learning (e.g. Palmeri, 1997) and reward learning (e.g. Wimmer et al., 2018) research report that in paradigms with multiple sessions, delayed by either weeks or days, task knowledge typically decays between sessions, but that it is also rapidly relearned and results in an overall net increase in proficiency. Thus, while the current results provide evidence that category representations may decay over time when the frequency of category observations differs, there is also the possibility that categorical knowledge may decay over time regardless of observational frequency. If this is the case, it is possible that both similarity and frequency may dictate the rate at which categorical knowledge decays over time. However, only further experimentation will provide a definitive answer.

6. CONCLUSION

6.1. General Summary

Each study in this dissertation, has sought to determine the potential impact that categorical and numerical reinforcers has on category learning. A review of the category and reward learning literature detailed striking parallels in both the behavior displayed when attempting to learn to categorize novel stimuli or determining which alternatives lead to the largest rewards, and in the theoretical mechanism by which learning is thought occur. While there have been previous attempts to apply numerical feedback to category learning paradigms (e.g. Abohamza et al., 2019; Daniel & Pollmann, 2010; Daniel & Pollmann, 2014; Montague et al., 2006), each utilize *discrete* reward values as feedback. For reference, this type of reward feedback could be simple values of 0 and 1, red or green colored indicators, or other types of invariable numerical feedback. In most of these studies, reward feedback is reported to be either comparable to categorical feedback or result in improved performance. In Study 1, corroborating evidence detailed that *discrete* numerical rewards did result in comparable performance as compared to categorical feedback, however, when *variable* numerical rewards were introduced, categorization was consistently poorer. Interestingly, however, participants who received categorical feedback were consistently quicker in determining category membership as compared to numerical feedback participants. In Studies 2 and 3, the overall reward magnitude or observation frequency of category clusters were modulated, and the same trend where numerical information results in poorer categorization accuracy was still observed.

While the first study focused solely on comparing the categorization accuracy between feedback types, studies 2 and 3 focused more on how facets of reward learning impact the generalization of categorical knowledge. Determining how people generalize learned knowledge is a staple in both the category and reward learning literatures: with classifying novel stimuli into previously learned categories (Seger & Peterson, 2013) and the associative learning observed in behavioral conditioning (Kakade & Dayan, 2002; Myers et al., 2003, 2017; Suri & Schultz, 1998). As such, in the current studies, both reward frequency and magnitude were modulated to determine if the more variable reward information would affect category generalization to a differing degree than categorical information.

As an example, if a person were asked to categorize circles and squares, would seeing, or being paid more, for circular stimuli result in more circle classifications of stimuli that lie in between circles and squares? Theories of reward learning would suggest that this would be the case (Rescorla & Wagner, 1972; Sutton & Barto, 2018), however, based on the current results, there seems to be no major differences between numerical and categorical feedback in most conditions. Our results are supported by the theories of category learning as they pertain to stimulus similarity (e.g. Nosofsky, 1988b; Pothos & Chater, 2005; A Tversky, 1977), and further backed up by the model fits of each study which detailed that the category-learner was the best fitting model on average. However, this is not to say that numerical information was completely disregarded. In the DI condition of Study 2 and the SpreadA condition in Study 3, the conditions where the most distant category cluster was the most rewarded or frequent, respectively, the data showed evidence of a potential conflict between stimulus similarity and reward/frequency. Further, the q-parameter values extracted from the catrwd-learner models did show that participants, in general, were not wholly focused on the categorical information—only mostly.

Finally, analysis of the model fits provided novel evidence that category learning may be subject to decay processes. In Study 3, where frequency was manipulated, the participant data was best fit by the decay variants of the ALCOVE models. Succinctly, when categories were observed at differing frequencies, the category representations for the less frequently observed categories decayed relatively more over time. However, as noted in Study 3, there was no evidence on whether or not a decay process would occur outside of frequency manipulated paradigms. To assess this, the Study 2 data was fit to decay model. This produced evidence that the ALCOVE-Decay models also provided the best fit to the data when only reward magnitude was manipulated (Figure 6.1). While this provided additional evidence that decay processes impact the category learning process, further experimentation is required.

Figure 6.1 Plot of the mean BIC values in the Study 2 behavioral data as fit both the ALCOVE and ALCOVE-Decay models. Errors bars represented by SD.

6.2. Limitations and Future Directions

In Studies 2 and 3, one of the major limitations was the online format of the category learning task. While the design of the online tasks attempted to control for external variables as much as possible, they were still unable to account for the expected reaction time slopes that are normally, and consistently, observed in category learning research (e.g. Ashby et al., 2003; Ashby & Maddox, 2005). However, the general learning trends between feedback types observed in Study 1 were again observed in Studies 2 and 3 suggesting the possibility that the online format may have only impacted reaction time. Though there is a non-zero probability that behavior may differ were these studies recompleted in a lab-based setting. Though, these two studies did address a possible limitation in Study 1 where non-deterministic category structures were used. In Study 1, there was not a distinct boundary between the two main categories, and this was discussed as a potential reason for the poorer performance observed in the reward feedback participants data. Based on the training data observed in Studies 2 and 3, it is likely that the difference between feedback types may occur regardless of the category/reward structure.

The EQ conditions in both studies 2 and 3, both used uniformly randomized rewards to serve as a control condition for each study. While these conditions served the desired purpose, it may have been prudent to have additional EQ conditions which varied in reward magnitude. For Study 2, this would have held similarity constant between categories, so it would have been easier to determine the base impact of reward magnitude on transfer phase categorizations. Similarly for Study 3, if conditions with more variable rewards were included, it is possible that a multiplicative, or additive, effect between reward and stimulus frequency may have been observed. Thus, to get a more complete picture of the impact that reward magnitude and frequency has on category learning, additional data should be collected.

Finally, as mentioned in the post hoc simulation sections for studies 2 and 3, a relatively sparse transfer phase was used to observe generalization behavior. In the post hoc simulations, the number of transfer stimuli was expanded, and resulted in predicted behavior which confirmed the original simulations. However, a few of the comparisons between the actual and predicted behavior showed evidence of a difference suggesting that the model deviated from the actual behavior. To concretely determine if the manipulations employed by these studies do indeed result in behavioral differences, future research would need to include a more expansive transfer phase. Additionally, based on the asymptotic training behavior in each study, it is possible that the length of the training phase allowed participants to learn to a criterion. This may have resulted in the participants viewing numerical outcomes as dichotomous (i.e. reward received/not received) after sufficient learning. An alternative design would include a truncated training phase with an expanded transfer phase.

6.3. Conclusions

Overall, the results indicate that people are able to utilize both numerical and categorical reinforcers to successfully learn how to categorize novel visual stimuli. While the data suggested that numerical feedback leads to poorer overall performance as compared to categorical feedback during initial learning, ceteris paribus, these differences seemed to disappear when participants were asked to generalize their knowledge to novel stimuli given the current paradigms. While this may imply that there were no overall differences in the impact of categorical and numerical information in the generalization of categorical knowledge, it is also possible that participants may have simply sufficiently learned each category resulting in both the observed lack of differences between feedback types, and an explanation for the model fitted preference for categorical information across each study and condition. Finally, the implementation of the decay-based variants of the ALCOVE models resulted in novel evidence that categorical representations may decay over time, which provides yet another link between the reward and category learning literatures.

REFERENCES

Abohamza, E., Weickert, T., Ali, M., & Moustafa, A. A. (2019). Reward and punishment learning in schizophrenia and bipolar disorder. *Behavioural Brain Research, July,* 112298. https://doi.org/10.1016/j.bbr.2019.112298

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). *A Primer on Foraging and the Explore / Exploit Trade-Off for Psychiatry Research*. *42*(10), 1931–1939. https://doi.org/10.1038/npp.2017.108

Ahn, W.-Y., Busemeyer, J., Wagenmakers, E.-J., & Stout, J. (2008). Comparison of Decision Learning Models Using the Generalization Criterion Method. *Cognitive Science: A Multidisciplinary Journal*, *32*(8), 1376–1402. https://doi.org/10.1080/03640210802352992

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Anderson, B. A., & Halpern, M. (2017). On the value-dependence of value-driven attentional capture. *Attention, Perception, and Psychophysics*, *79*(4), 1001–1011. https://doi.org/10.3758/s13414-017-1289-6

Anderson, B. A., Laurent, P. A., & Yantis, S. (2012). Generalization of value-based attentional priority. *Visual Cognition*, *20*(6), 647–658. https://doi.org/10.1080/13506285.2012.679711

Apitz, T., & Bunzeck, N. (2012). Reward modulates the neural dynamics of early visual category processing. *NeuroImage*, *63*(3), 1614–1622. https://doi.org/10.1016/j.neuroimage.2012.08.046

Aristotle. (1975). *Aristotle's Categories* (Ackrill (ed.)). Oxford University Press.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295X.105.3.442

Ashby, F. G., & Gott, R. E. (1988). Decision Rules in the Perception and Categorization of Multidimensional Stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53. https://doi.org/10.1037/0278-7393.14.1.33

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New*

York Academy of Sciences, *1224*(1), 147–161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., Noble, S., Waldron, E. M., Filoteo, J. V., & Ell, S. W. (2003). Category learning deficits in Parkinson’s disease. *Neuropsychology*, *17*(1), 115–124. https://doi.org/10.1037/0894-4105.17.1.115

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. *Formal Approaches in Categorization*, 65–88.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception and Psychophysics*, *61*(6), 1178–1199. https://doi.org/10.3758/BF03207622

Ashby, F. G., Smith, J. D., & Rosedahl, L. A. (2019). Dissociations between rule-based and information-integration categorization are not caused by differences in task difficulty. *Memory & Cognition*. https://doi.org/10.3758/s13421-019-00988-4

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419. https://doi.org/10.1016/S0028-3908(98)00033-1

Bar-Hillel, M., & Fischhoff, B. (1981). *When do base rates affect predictions?*

Barsalou, Lawrence, W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227. https://doi.org/10.1037/0278-7393.11.1-4.629

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629

Bayer, H. M., & Glimcher, P. W. (2005). *a Quantitative Reward Prediction Error Signal*. *47*(3), 129–141. https://doi.org/10.1016/j.neuron.2005.05.020

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1–3), 7–15. http://www.ncbi.nlm.nih.gov/pubmed/8039375

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bellebaum, C., Polezzi, D., & Daum, I. (2010). It is less than you expected: The feedback-related negativity reflects violations of reward magnitude expectations.

Neuropsychologia, *48*(11), 3343–3350. https://doi.org/10.1016/j.neuropsychologia.2010.07.023

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Berridge, K. C. (2000). Reward learning: Reinforcement, incentives, and expectations. *Psychology of Learning and Motivation*, *40*, 223–278. https://doi.org/10.1016/S0079-7421(00)80022-5

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*(3), 309–369. https://doi.org/10.1016/S0165-0173(98)00019-8

Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-dynamic programming*. Athena Scientific.

Bódi, N., Kéri, S., Nagy, H., Moustafa, A., Myers, C. E., Daw, N., Dibó, G., Takáts, A., Bereczki, D., & Gluck, M. A. (2009). Reward-learning and the novelty-seeking personality: A between-and within-subjects study of the effects of dopamine agonists on young parkinsons patients. *Brain*, *132*(9), 2385–2395. https://doi.org/10.1093/brain/awp094

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*(May 2015), 1–9. https://doi.org/10.1038/ncomms15958

Bott, L., Hoffman, A. B., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General*, *136*(4), 685–699. https://doi.org/10.1037/0096-3445.136.4.685

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software; Vol 1, Issue 1 (2017)* . https://www.jstatsoft.org/v080/i01

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, *58*(5), 313.

Capaldi, E. J. (1966). Partial reinforcement: A hypothesis of sequential effects. *Psychological Review*, *73*(5), 459.

Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. (2016). A Comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human Brain Mapping*, *37*(10), 3557–3574.

https://doi.org/10.1002/hbm.23259

Clapper, J. P., & Bower, G. H. (1994). Category Invention in Unsupervised Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 443–460. https://doi.org/10.1037/0278-7393.20.2.443

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Conaway, N., & Kurtz, K. J. (2017). Similar to the category, but not the exemplars: A study of generalization. *Psychonomic Bulletin and Review*, *24*(4), 1312–1323. https://doi.org/10.3758/s13423-016-1208-1

Constantino, S. M., & Daw, N. D. (2015). *Learning the opportunity cost of time in a patch-foraging task*. *April*, 837–853. https://doi.org/10.3758/s13415-015-0350-y

Cornwall, A. C., Byrne, K. A., & Worthy, D. A. (2018). Gender differences in preference for reward frequency versus reward magnitude in decision-making under uncertainty. *Personality and Individual Differences*, *135*. https://doi.org/10.1016/j.paid.2018.06.031

Daniel, R., & Pollmann, S. (2010). Comparing the Neural Basis of Monetary Reward and Cognitive Feedback during Information-Integration Category Learning. *Journal of Neuroscience*, *30*(1), 47–55. https://doi.org/10.1523/JNEUROSCI.2205-09.2010

Daniel, Reka, & Pollmann, S. (2014). A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiology of Learning and Memory*, *114*, 90–100. https://doi.org/10.1016/j.nlm.2014.05.002

Daniel, Reka, Wagner, G., Koch, K., Reichenbach, J. R., Sauer, H., & Schlösser, R. G. M. (2011). Assessing the neural basis of uncertainty in perceptual category learning through varying levels of distortion. *Journal of Cognitive Neuroscience*, *23*(7), 1781–1793. https://doi.org/10.1162/jocn.2010.21541

Davis, T., Xue, G., Love, B. C., Preston, A. R., & Poldrack, R. A. (2014). Global neural pattern similarity as a common basis for categorization and recognition memory. *Journal of Neuroscience*, *34*(22), 7472–7484. https://doi.org/10.1523/JNEUROSCI.3376-13.2014

Davison, M., & McCarthy, D. (2016). *The matching law: A research review*.

Daw, N D, O'Doherty, J. P., Dayan, P., Dolan, R. J., & Seymour, B. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

https://doi.org/10.1038/nature04766

Daw, Nathaniel D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*(2), 199–204. https://doi.org/10.1016/j.conb.2006.03.006

Daw, Nathaniel D., & Shohamy, D. (2008). The cognitive neuroscience of motivation and learning. *Social Cognition*, *26*(5), 593–620. https://doi.org/10.1521/soco.2008.26.5.593

Daw, Nathaniel D, Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Daw, Nathaniel D, O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational Neuroscience Series.

Dayan, P., & Balleine, B. W. (2002). Reward, Motivation, and Reinforcement Learning. *Neuron*, *36*(2), 285–298. https://doi.org/10.1016/S0896-6273(02)00963-7

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly. *Current Opinion in Neurobiology*, *18*(2), 185–196. https://doi.org/10.1016/j.conb.2008.08.003

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. https://doi.org/10.3758/s13428-015-0567-2

Don, H.J., Otto, A. R., Cornwall, A. C., Davis, T., & Worthy, D. A. (2019). Learning reward frequency over reward probability: A tale of two learning rules. *Cognition*, *193*. https://doi.org/10.1016/j.cognition.2019.104042

Don, Hilary J., & Livesey, E. J. (2017). Effects of outcome and trial frequency on the inverse base-rate effect. *Memory and Cognition*, *45*(3), 493–507. https://doi.org/10.3758/s13421-016-0667-y

Don, Hilary J., Worthy, D. A., & Livesey, E. J. (2021). Hearing hooves, thinking zebras: A review of the inverse base-rate effect. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-020-01870-0

Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: A case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 933–948. https://doi.org/10.1037/xlm0000083

Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(4), 840–859. https://doi.org/10.1037/a0027867

Ell, S. W., Marchant, N. L., & Ivry, R. B. (2006). Focal putamen lesions impair learning in rule-based, but not information-integration categorization tasks. *Neuropsychologia*, *44*(10), 1737–1751. https://doi.org/10.1016/j.neuropsychologia.2006.03.018

Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H. J., Ornelas, S., Koay, S. A., Witten, I. B., Thiberge, S. Y., Daw, N. D., & David, W. (2019). Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature*, *570*, 509–513. https://doi.org/10.1038/s41586-019-1261-9

Erev, I., & Barron, G. (2005). On Adaptation, Maximization, and Reinforcement Learning Among Cognitive Strategies. *Psychological Review*, *112*(4), 912–931. https://doi.org/10.1037/0033-295X.112.4.912

Erev, I., & Roth, A. E. (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, *88*(4), 848–881. https://doi.org/10.1086/250095

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology. General*, *127*(2), 107–140. https://doi.org/10.1037//0096-3445.127.2.107

Estes, W.K. (1986). Array models for category learning. *Cognitive Psychology*, *18*(4), 500–549. https://doi.org/10.1016/0010-0285(86)90008-3

Estes, William K. (1976). The cognitive side of probability learning. *Psychological Review*, *83*(1), 37.

Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the Frontal Lobes. *Psychological Science*, *21*(3), 415–423. https://doi.org/10.1177/0956797610362646

Fisher, S. (1916). The Process of Generalization Abstraction: and Its Productt, the General Concept. *Psychological Review*, *21*(2).

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review*, *113*(2), 300–326. https://doi.org/10.1037/0033-295X.113.2.300

Fryer, R., & Jackson, M. O. (2008). A Categorical Model of Cognition and Biased Decision Making. *The B.E. Journal of Theoretical Economics*, *8*(1). https://doi.org/10.2202/1935-1704.1357

Gershman, S. J. (2018a). Deconstructing the human algorithms for exploration. *Cognition*, *173*(December 2017), 34–42. https://doi.org/10.1016/j.cognition.2017.12.014

Gershman, S. J. (2018b). Uncertainty and exploration. *Decision*. https://doi.org/10.1037/dec0000101

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*, 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Gershman, S. J., & Uchida, N. (2019). Believing in dopamine. *Nature Reviews Neuroscience, 20*(11), 703–714. https://doi.org/10.1038/s41583-019-0220-7

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108*(Supplement 3), 15647–15654.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*(3), 227–247. https://doi.org/10.1037/0096-3445.117.3.227

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning ? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Haruno, M., & Kawato, M. (2006). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *Journal of Neurophysiology*, *95*(2), 948–959. https://doi.org/10.1152/jn.00382.2005

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In *ALI ASADZADEH (2017)the role of tourism on the environment and its governing law.Electic journal of biology 13.* (Issue 3, pp. 163–199). https://doi.org/10.1016/S0079-7421(00)80034-1

Herrnstein, R. J., & Hineline, P. N. (1966). Negative Reinforcement As Shock-Frequency Reduction 1 . *Journal of the Experimental Analysis of Behavior*, *9*(4), 421–430. https://doi.org/10.1901/jeab.1966.9-421

Herrnstein, R J. (1970). *On the law of effect*. *2*(2), 243–266.

Herrnstein, R J. (1974). Formal Properties Of The Matching Law. *Journal of the Experimental Analysis of Behavior*, *21*(1), 159–164. https://doi.org/10.1901/jeab.1974.21-159

Herrnstein, Richard J, & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives*, *5*(3), 137–156.

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141. https://doi.org/10.1016/S0042-6989(97)00043-6

Hickey, C., Kaiser, D., & Peelen, M. V. (2015). Reward guides attention to object categories in real-world scenes. *Journal of Experimental Psychology: General*, *144*(2), 264–273. https://doi.org/10.1037/a0038627

Homa, D., & Cultice, J. C. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 83–94. https://doi.org/10.1037//0278-7393.10.1.83

Homa, D., Cross, J., Cornell, D., Goldman, D., & Shwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, *101*(1), 116–122. https://doi.org/10.1037/h0035772

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs*, *28*(1), 1–86.

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

Kahnt, T., Park, S. Q., Burke, C. J., & Tobler, P. N. (2012). How glitter relates to gold: Similarity-dependent reward prediction errors in the human striatum. *Journal of Neuroscience*, *32*(46), 16521–16529. https://doi.org/10.1523/JNEUROSCI.2383-12.2012

Kakade, S., & Dayan, P. (2002). Dopamine: Generalization and bonuses. *Neural Networks*, *15*(4–6), 549–559. https://doi.org/10.1016/S0893-6080(02)00048-5

Kalish, M. L. (2001). An inverse base rate effect with continuously valued stimuli. *Memory and Cognition*, *29*(4), 587–597. https://doi.org/10.3758/BF03200460

Kant, I. (1988). *Critique of Pure Reason* (P. Guyer & A. W. Wood (eds.)). Cambridge University Press.

Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). *The nature of belief-directed exploratory choice in human decision-making*. *2*(January), 1–12. https://doi.org/10.3389/fpsyg.2011.00398

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). *Anticipation of Increasing Monetary Reward Selectively Recruits Nucleus Accumbens*. *21*, 1–5.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(1), 1–53. https://doi.org/10.1017/s0140525x00041157

Kool, W., Cushman, F. A., & Gershman, S. J. (2018). Competition and cooperation between multiple reinforcement learning systems. *Goal-Directed Decision Making: Computations and Neural Circuits*, 153–178. https://doi.org/10.1016/B978-0-12-812098-9.00007-3

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychological Science*, *28*(9), 1321–1333. https://doi.org/10.1177/0956797617708288

Krebs, R. M., Boehler, C. N., Roberts, K. C., Song, A. W., & Woldorff, M. G. (2012). The Involvement of the Dopaminergic Midbrain and Cortico-Striatal-Thalamic Circuits in the Integration of Reward Prospect and Attentional Task Demands. *Cerebral Cortex*, *22*(3), 607–615. https://doi.org/10.1093/cercor/bhr134

Krueger, P. M., Wilson, R. C., & Cohen, J. D. (2017). Strategies for exploration in the domain of losses. *Judgment and Decision Making*, *12*(2), 104–117.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. https://doi.org/10.1037/0033-295X.99.1.22

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(1), 3–26. https://doi.org/10.1037/0278-7393.22.1.3

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*(6), 812–863. https://doi.org/10.1006/jmps.2000.1354

Kruschke, J. K. (2012). Models of attentional learning. *Formal Approaches in Categorization*, 120–152. https://doi.org/10.1017/cbo9780511921322.006

Lake, B. M., & McClelland, J. L. (2011). Estimating the strength of unlabeled information during semi-supervised learning. *Proceedings of the 33rd Annual Conference of the*

Cognitive Science Society.

Lamberts, K., & Kent, C. (2007). No evidence for rule-based processing in the inverse base-rate effect. *Memory and Cognition*, *35*(8), 2097–2105. https://doi.org/10.3758/BF03192941

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Liu, Z., Zhang, Y., Ma, D., Xu, Q., & Seger, C. A. (2020). Differing effects of gain and loss feedback on rule-based and information-integration category learning. *Psychonomic Bulletin and Review*, 274–282. https://doi.org/10.3758/s13423-020-01816-6

Lohse, K. R., Miller, M. W., Daou, M., Valerius, W., & Jones, M. (2020). Dissociating the contributions of reward-prediction errors to trial-level adaptation and long-term learning. *Biological Psychology*, *149*(January 2019), 107775. https://doi.org/10.1016/j.biopsycho.2019.107775

Love, B. C. (2003). The multifaceted nature of unsupervised category learning. *Psychonomic Bulletin and Review*, *10*(1), 190–197. https://doi.org/10.3758/BF03196484

Luce, R. D. (1977). *The Choice Axiom after Twenty Years*. 215–233.

Ludvig, E. A., Balci, F., & Spetch, M. L. (2011). Reward magnitude and timing in pigeons. *Behavioural Processes*, *86*(3), 359–363. https://doi.org/10.1016/j.beproc.2011.01.003

Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, *3*(MAR), 1–13. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Maddox, W. T., & Bohil, C. J. (2000). Costs and benefits in perceptual categorization. *Memory and Cognition*, *28*(4), 597–615. https://doi.org/10.3758/BF03201250

Maddox, W. Todd. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. *Journal of the Experimental Analysis of Behavior*, *78*(3), 567–595. https://doi.org/10.1901/jeab.2002.78-567

Maddox, W. Todd, Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory and Cognition*, *32*(4), 582–591. https://doi.org/10.3758/BF03195849

Maddox, W. Todd, & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(6), 1459–1482. https://doi.org/10.1037/0278-7393.24.6.1459

Maddox, W. Todd, Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition*, *108*(2), 578–589. https://doi.org/10.1016/j.cognition.2008.03.010

Maddox, W Todd, Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 650–662. https://doi.org/10.1037/0278-7393.29.4.650

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*(4), 592–613. https://doi.org/10.1037/0033-2909.129.4.592

Markman, E. M. (1989). Categorization and naming in children: Problems of induction. In *Categorization and naming in children: Problems of induction.* (p. 250). The MIT Press.

McDonnell, J. V., Jew, C. A., & Gureckis, T. M. (2012). Sparse category labels obstruct generalization of category membership. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*, 749–754.

Meadows, C. C., Gable, P. A., Lohse, K. R., & Miller, M. W. (2016). The effects of reward magnitude on reward processing: An averaged and single trial event-related potential study. *Biological Psychology*, *118*, 154–160. https://doi.org/10.1016/j.biopsycho.2016.06.002

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066X.44.12.1469

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, *117*(1), 68.

Medin, D. L., Ross, B. H., & Markman, A. B. (2004). *Cognitive Psychology 4th ed.*

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Reviews of Psychology*, *35*, 113–138. https://doi.org/10.1017/CBO9780511801938.011

Medin, D., & Schaffer, M. (1978). Context Theory of Classification Learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Mervis, C B, & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, *32*(1), 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Mervis, Carolyn B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*(3), 283–284. https://doi.org/10.3758/BF03337190

Miller, L. B., & Estes, B. W. (1961). Monetary reward and motivation in discrimination learning. *Journal of Experimental Psychology*, *61*(6), 501–504. https://doi.org/10.1037/h0040489

Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of COVIS. *European Journal of Neuroscience*, *34*(8), 1326–1336. https://doi.org/10.1111/j.1460-9568.2011.07847.x

Montague, P. R., King-Casas, B., & Cohen, J. D. (2006). Imaging Valuation Models in Human Choice. *Annual Review of Neuroscience*, *29*(1), 417–448. https://doi.org/10.1146/annurev.neuro.29.051605.112903

Mory, E. H. (2004). Feedback research revisited. In *Handbook of research on educational communications and technology* (2nd ed., pp. 745–783). Lawrence Erlbaum Associates Publishers.

Moustafa, A. A., & Gluck, M. A. (2011). A neurocomputational model of dopamine and prefrontal-striatal interactions during multicue category learning by Parkinson patients. *Journal of Cognitive Neuroscience*, *23*(1), 151–167. https://doi.org/10.1162/jocn.2010.21420

Moustafa, A. A., Gluck, M. A., Herzallah, M. M., & Myers, C. E. (2015). The influence of trial order on learning from reward vs. punishment in a probabilistic categorization task: Experimental and computational analyses. *Frontiers in Behavioral Neuroscience*, *9*(JULY), 1–15. https://doi.org/10.3389/fnbeh.2015.00153

Myers, C. E., Rego, J., Haber, P., Morley, K., Beck, K. D., Hogarth, L., & Moustafa, A. A. (2017). Learning and generalization from reward and punishment in opioid addiction. *Behavioural Brain Research*, *317*, 122–131. https://doi.org/10.1016/j.bbr.2016.09.033

Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., Golomb, J., Schnirman, G., & Schwartz, R. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience*, *15*(2), 185–193. https://doi.org/10.1162/089892903321208123

Nagin, D. S. (1999). Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychological Methods*, *4*(2), 139.

Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & Bürkner, P. C. (2019). An introduction to bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154. https://doi.org/10.1016/j.jmp.2008.12.005

Nomura, E. M., Maddox, W. T., & Reber, P. J. (2007). Mathematical models of visual category learning enhance fMRI data analysis. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, *29*(Nosofsky), 539–544. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.160.6451&rep=rep1&type=pdf

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 10, Issue 1, pp. 104–114). https://doi.org/10.1037/0278-7393.10.1.104

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988a). Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (1988b). Similarity, Frequency, and Category Representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54–65. https://doi.org/10.1037/0278-7393.14.1.54

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., Mckinley, S. C., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*(3), 352–369. https://doi.org/10.3758/BF03200862

Nosofsky, R. M., Sanders, C. A., Zhu, X., & McDaniel, M. A. (2019). Model-guided search for optimal natural-science-category training exemplars: A work in progress. *Psychonomic Bulletin and Review*, *26*(1), 48–76. https://doi.org/10.3758/s13423-

018-1508-8

O'Brien, E. L., & Hess, T. M. (2020). Differential focus on probability and losses between young and older adults in risky decision-making. *Aging, Neuropsychology, and Cognition*, *27*(4), 532–552. https://doi.org/10.1080/13825585.2019.1642442

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*, 35–53. https://doi.org/10.1196/annals.1390.022

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(2), 324–354. https://doi.org/10.1037/0278-7393.23.2.324

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When Does Feedback Facilitate Learning of Words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Peterson, E. J., & Seger, C. A. (2013). Many hats: Intratrial and reward level-dependent BOLD activity in the striatum and premotor cortex. *Journal of Neurophysiology*, *110*(7), 1689–1702. https://doi.org/10.1152/jn.00164.2012

Petter, E. A., Gershman, S. J., & Meck, W. H. (2018). Integrating Models of Interval Timing and Reinforcement Learning. *Trends in Cognitive Sciences*, *22*(10), 911–922. https://doi.org/10.1016/j.tics.2018.08.004

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. In *Cognitive Science* (Vol. 26, Issue 3). https://doi.org/10.1016/S0364-0213(02)00064-2

Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(4), 733–752. https://doi.org/10.1080/02724980443000322

Racey, D., Young, M. E., Garlick, D., Pham, J. N. M., & Blaisdell, A. P. (2011). Pigeon and human performance in a multi-armed bandit task in response to changes in variable interval schedules. *Learning and Behavior*, *39*(3), 245–258. https://doi.org/10.3758/s13420-011-0025-7

Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic Reinforcement Learning: The Role of Structure and Attention. *Trends in Cognitive Sciences*, *23*(4), 278–292. https://doi.org/10.1016/j.tics.2019.01.010

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, *2*, 64–99.

Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making*, *28*(6), 850–865. https://doi.org/10.1177/0272989X08327066

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychological Science in the Public Interest, Supplement*, *7*(1), 1–44. https://doi.org/10.1111/j.1529-1006.2006.00026.x

Rolls, E. T., Mccabe, C., & Redoute, J. (2008). *Expected Value , Reward Outcome , and Temporal Difference Error Representations in a Probabilistic Decision Task*. *March*. https://doi.org/10.1093/cercor/bhm097

Rosch, E. (1978). Principles of Categorization. In *Cognition and Categorization* (pp. 27–48).

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rose, J. E., & Behm, F. M. (2004). Extinguishing the rewarding value of smoke cues: Pharmacological and behavioral treatments. *Nicotine and Tobacco Research*, *6*(3), 523–532. https://doi.org/10.1080/14622200410001696501

Ross, B. H. (2000). The effects of category use on learned categories. *Memory and Cognition*, *28*(1), 51–63. https://doi.org/10.3758/BF03211576

Salatas, H., & Bourne, L. E. (1974). Learning conceptual rules: III. Processes contributing to rule difficulty. *Memory & Cognition*, *2*(3), 549–553. https://doi.org/10.3758/BF03196919

Schevernels, H., Krebs, R. M., Santens, P., Woldorff, M. G., & Boehler, C. N. (2014). Task preparation processes related to reward prediction precede those related to task-difficulty expectation. *NeuroImage*, *84*(1), 639–647. https://doi.org/10.1016/j.neuroimage.2013.09.039

Schlegelmilch, R., & von Helversen, B. (2020). The Influence of Reward Magnitude on

Stimulus Memory and Stimulus Generalization in Categorization Decisions. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000747

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, *57*, 87–115. https://doi.org/10.1146/annurev.psych.56.091103.070229

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*(1), 23–32. https://doi.org/10.18910/29195

Schultz, W. (2017). Reward prediction error. *Current Biology*, *27*(10), R369–R371. https://doi.org/10.1016/j.cub.2017.02.064

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *13*(3), 900–913. https://doi.org/8441015

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Seger, C. A., Braunlich, K., Wehe, H. S., & Liu, Z. (2015). Generalization in category learning: The roles of representational and decisional uncertainty. *Journal of Neuroscience*, *35*(23), 8802–8812. https://doi.org/10.1523/JNEUROSCI.0654-15.2015

Seger, C. A., & Peterson, E. J. (2013). Categorization=decision making+generalization. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1187–1200. https://doi.org/10.1016/j.neubiorev.2013.03.015

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 219–236. https://doi.org/10.1016/j.neubiorev.2007.07.008

Sims, C. R., Neth, H., Jacobs, R. A., & Gray, W. D. (2013). Melioration as rational choice: Sequential decision making in uncertain environments. *Psychological Review*, *120*(1), 139–154. https://doi.org/10.1037/a0030850

Skinner, B. F. (1938). *The behavior of organisms; an experimental analysis*. Appleton-Century Company, Incorporated.

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Harvard University Press. https://doi.org/10.4159/harvard.9780674866270

Smith, J. D. (2014). Prototypes, exemplars, and the natural history of categorization. *Psychonomic Bulletin & Review*, *21*(2), 312–331. https://doi.org/10.3758/s13423-013-0506-0

Sohoglu, E., & Davis, M. H. (2016). *Perceptual learning of degraded speech by minimizing prediction error*. https://doi.org/10.1073/pnas.1523266113

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*(2), 351–367.

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors , dopamine neurons and learning. *Nature Publishing Group*, *16*(7). https://doi.org/10.1038/nn.3413

Suri, R. E., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Experimental Brain Research*, *121*(3), 350–354. https://doi.org/10.1007/s002210050467

Sutton, R., & Barto, A. (2018). Reinforcement Learning: An Introduction. In *MIT Press*. https://doi.org/10.1179/175622708X282910

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*(4), i–109. https://doi.org/10.1037/h0092987

Thorndike, E. L. (1927). The Law of Effect. *The American Journal of Psychology*, *39*(1), 212–222.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 27–50. https://doi.org/10.4324/9780203789155-11

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Tversky, Amos, & Kahneman, D. (1981). Evidential Impact of Base Rates. *Judgment Under Uncertainty: Heuristics and Biases*, *197*, 153–160. https://doi.org/10.1002/car.2380010211

Van Den Bos, R., Houx, B. B., & Spruijt, B. M. (2006). The effect of reward magnitude differences on choosing disadvantageous decks in the Iowa Gambling Task. *Biological Psychology*, *71*(2), 155–161. https://doi.org/10.1016/j.biopsycho.2005.05.003

Vandist, K., De Schryver, M., & Rosseel, Y. (2009). Semisupervised category learning: The impact of feedback in learning the information-integration task. *Attention, Perception, & Psychophysics*, *71*(2), 328–341. https://doi.org/10.3758/APP.71.2.328

Vandist, K., Storms, G., & Bussche, E. Van Den. (2019). *Semisupervised category learning facilitates the development of automaticity*. 137–157.

Varshney, L. R., Rhim, J. B., Varshney, K. R., & Goyal, V. K. (2011). Categorical decision making by people, committees, and crowds. *2011 Information Theory and Applications Workshop, ITA 2011 - Conference Proceedings*, 425–434. https://doi.org/10.1109/ITA.2011.5743615

Vaughan, W. (1981). Melioration, Matching, and Maximization. *Journal of the Experimental Analysis of Behavior*, *36*(2), 141–149. https://doi.org/10.1901/jeab.1981.36-141

Vong, W. K., Navarro, D. J., & Perfors, A. (2016). The helpfulness of category labels in semi-supervised learning depends on category structure. *Psychonomic Bulletin and Review*, *23*(1), 230–238. https://doi.org/10.3758/s13423-015-0857-9

Vygotskiĭ, L. S. (2012). *Thought and language*. MIT press.

Walker, A. R., Luque, D., Le Pelley, M. E., & Beesley, T. (2019). The role of uncertainty in attentional and choice exploration. *Psychonomic Bulletin and Review*, *26*(6), 1911–1916. https://doi.org/10.3758/s13423-019-01653-2

Weinstein, L. (1971). Effects of an increment in monetary incentive magnitude on instrumental responding and repeated increases in reward magnitude in humans. *Psychonomic Science*, *25*(4), 235–237. https://doi.org/10.3758/BF03329104

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*. Stanford Univ Ca Stanford Electronics Labs.

Williams, R. J. (1992). *Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning*. *256*, 229–256.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology. General*, *143*(6), 2074–2081. https://doi.org/10.1037/a0038199

Wimmer, G. E., Daw, N. D., & Shohamy, D. (2012). Generalization of value in reinforcement learning by humans. *European Journal of Neuroscience*, *35*(7), 1092–1104. https://doi.org/10.1111/j.1460-9568.2012.08017.x

Wimmer, G. E., Li, J. K., Gorgolewski, K. J., & Poldrack, R. A. (2018). Reward learning over weeks versus minutes increases the neural representation of value in the human brain. *Journal of Neuroscience*, *38*(35), 7649–7666. https://doi.org/10.1523/JNEUROSCI.0075-18.2018

Winman, A., Wennerholm, P., Juslin, P., & Shanks, D. R. (2005). Evidence for rule-based processes in the inverse base-rate effect. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(5), 789–815. https://doi.org/10.1080/02724980443000331

Wise, R. A. (1989). The brain and reward. In *The neuropharmacological basis of reward.* (pp. 377–424). Clarendon Press/Oxford University Press.

Worthy, D. A., Markman, A. B., & Todd Maddox, W. (2013). Feedback and stimulus-offset timing effects in perceptual category learning. *Brain and Cognition*, *81*(2), 283–293. https://doi.org/10.1016/j.bandc.2012.11.006

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin and Review*, *12*(3), 387–402. https://doi.org/10.3758/BF03193783

Zaki, S. R., & Kleinschmidt, D. F. (2014). Procedural memory effects in categorization: Evidence for multiple systems or task complexity? *Memory and Cognition*, *42*(3), 508–524. https://doi.org/10.3758/s13421-013-0375-9

Zeithamova, D., & Maddox, W. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 731–741. https://doi.org/10.1037/a0015005