Category learning is constrained by training context and prior experience

by
Casey L. Roark

A Dissertation Presented in Partial Fulfillment
of Requirements for the Degree of
Doctor of Philosophy

Graduate Supervisory Committee
Dr. Lori L. Holt, Chair
Dr. David Plaut
Dr. Julie Fiez

Department of Psychology &
Center for the Neural Basis of Cognition
Carnegie Mellon University
Abstract

Everyday behaviors like interpreting a child’s squeal as thrilled or terrified or understanding diverse acoustic signals from different talkers to each be the word “thanks” rely on categorization. Learning to treat perceptually distinct objects as functionally equivalent changes how we structure our knowledge about the world. Category learning is a major area of research that spans from cellular neuroscience to human behavioral methods to philosophy. Despite the breadth of research on category learning, much still remains unknown about how humans create and reorganize mental representations of categories. Even so, the majority of research on perceptual category learning focuses on visual categories. In this dissertation, I will focus on auditory category learning. Sound presents unique learning challenges that are important for understanding speech learning, music perception, and everyday listening.

This dissertation investigates how humans build on existing knowledge to learn new sound categories. Chapter 1 presents a theoretical framework on the interaction of sensory experience, perceptual representations, and category learning. Chapters 2 and 3 uncover how factors of the current learning context affect category learning. Chapter 4 examines how existing perceptual representations influence how learners form categories within a perceptual environment. Chapters 5 and 6 investigate how experience with statistically structured sensory information influences similarity-based perceptual representations and the effect of this experience on subsequent category learning. A neural network model presented in Chapter 7 serves as a starting point in understanding the underlying computational mechanisms that allow sensory experience to shape perceptual representations, which then influence higher-level cognition, such as the processes involved in category learning. This research advances understanding of both auditory and visual categorization by revealing how category learning is both constrained by prior experience and influenced by regularities and the context of the current learning environment.
Acknowledgements

My advisor, Lori Holt, has been instrumental in preparing this tome which reflects much of the work I have done in graduate school. I feel so genuinely lucky to know her as a person and to have her as my advisor. She has helped me develop the skills I needed to succeed in the graduate program and supported me in my endeavors outside of academia. She has always made me feel appreciated and valued in the five years that I’ve spent at CMU.

I am very thankful for the mentorship and collaboration of my committee members, Julie Fiez and David Plaut. They have supported me throughout my entire graduate school career, encouraged me when I was feeling unsure of my path, and helped me see the value in my work and abilities.

I want to thank the members of the Speech Perception and Learning Laboratory for their friendship, mentorship, and incredible assistance in managing and running my many experiments over the course of my graduate school career (which involved over 1700 subjects!!)—including Christi Gomez, Ran Liu, Matt Lehet, Charles Wu, Tim Murphy, Neeraj Sharma, Julia Zhang, Shae DiNino, Chisom Obasih, Alan Sales Barbosa, Hayoung Woo, Megan Bartoshuk, Theodore Teichman, Matthew Furfman, Kevin Bergen, Erin Smith, Joey Rohrbagh, Delaney Tyson, Krishna Rao, Uyai-Abasi Akpakpan, and many others. I’m also incredibly grateful for the newly formed Pittsburgh Cognitive Auditory Neuroscience group and all the new people I’ve been able to interact with in the past year.

I am so very lucky to have the graduate cohort that I do with members, Bo Powers, Cari Skrzynski, and Sandrine Girard. In my fellow graduate students, I have found friends, mentors, and inspiration— including Juliet Shafto, Katslyn Mascatelli, Lucy Erickson, Rony Patel, and Kevin Jarbo. I am truly grateful for every person I have known through CMU Psychology.
I am so thankful for those who remind me that there is importance in a rich and full life outside of graduate school, including dedication to the community and the next generation of scientists—Phoebe Dinh, Krista Bond, Patience Stevens, Jeanean Naqvi, Michael Granovetter, Roderick Seow, Meredith Van Vleet, Anna Fisher, and Tim Verstynen.

I am grateful for Vicki Helegson who has acted as such an advocate for me and other graduate students. I could not have navigated graduate school without guidance from Erin Donahoe, Becky Finkel, and Melissa Stupka.

I am very grateful for those who saw promise in me as an undergraduate and gave me opportunities for the experiences that have helped me see myself in research and as a PhD—Patti Adler, Tiffany Beechy, Tim Curran, Tiffany Ito, and Sarah Grover.

I am incredibly grateful for Vitaly Meursault, who has given me invaluable personal and professional support. I am happy to dedicate this dissertation to my family, who has supported and encouraged me my entire life. To my parents, Robert and Janita Roark; my sister, brother-in-law, and nephew, Samantha, Nathan, and Callen Geerdes.
Chapter 1: Introduction

Within the domains of vision and audition, existing theories have explored how we represent stimuli psychologically. These theories often make the assumption that physical stimuli are represented in multidimensional psychological space (Ashby, 1992; Goldstone, 1998). As researchers, to simplify learning challenges, we nearly always assume that the physical learning environment we present to participants and each participant’s psychological representation space are consistent (Ashby & Soto, 2015; Johannesson, 2001; Kluender, Stilp, & Kiefte, 2013). This consistency assumption makes our understanding of category learning challenges that meet this assumption possible and much easier to understand. However, this also means that even category learning challenges that do not meet this assumption are treated as if they do meet the assumption. When we do not understand the psychological representation of sensory information, it limits our understanding of which dimensions participants use during learning, whether they have preferences for certain dimensions over others, and how they are able to weight the dimensions in their category decisions during learning. What is especially worth considering is how the existing psychological representation space is changed as a consequence of learning or other experience in the world and how participants are able to cope when the physical environment does not match their existing representations.

The transformation of dimensions from physical to psychological space is not always isomorphic or faithful. Psychological space is created as a consequence of experience, which can signal to learners what is important in the world or what is irrelevant, and how the world is structured. This experience then affects what participants will be able to do when learning novel categories.
Instead of assuming that the physical world and the learner’s psychological representations are consistent or identical, it would benefit researchers to understand how those psychological representations are formed and are similar to or different from the physical environment. Further, it is necessary to understand how this transformation occurs to understand how it affects learning.

In this chapter, I present my overall framework of the connections between sensory experience, perceptual representations, and novel category learning (Figure 1). I will explore how structure in sensory experience shapes psychological, perceptual representations (Section 1). I will then investigate ways in which these transformations and existing perceptual representations can impact subsequent learning (Section 2).

Figure 1. Theoretical framework outline.
Section 1: Sensory experience and perceptual representations

Sensitivity to Statistical Regularities and Relationships between Dimensions

There is structure in the world. As Rosch (1978) wrote, “The perceived world is not an unstructured total set of equiprobable co-occurring attributes” (p. 28). In experienced environments, there are associations and regularities present in the information. This includes information about statistics of distributions of categories of experience, such as variance and other aspects of category structure. This also includes information about events generally—such as the frequency of events and the order in which events occur, which is related to causality. These regularities are often tied to the physics of the natural world, but always reflect some regularity in the world.

People are sensitive to the regularities present in the physical world, including relations among dimensions. First, I will focus on the ways in which people are sensitive to distributions and dimensions in general and then I will elaborate on how people are sensitive to the relations among dimensions. Importantly, these kinds of statistical regularities and relationships are reflected in representations of distributions and features of sensory stimuli.

Statistical Regularities. Humans are sensitive to regularities and relationships in their environment, even from a very young age (Fiser & Aslin, 2001, 2002; Maye, Werker, & Gerken, 2002; Saffran, Johnson, Aslin, & Newport, 1999; Turk-Browne, Jungé, & Scholl, 2005). Sensitivity to these regularities and the ability to learn things about the statistical regularities in the environment has been labeled ‘statistical learning’ (Saffran, Aslin, & Newport, 1996). Even infants as young as 8 months old exhibit statistical learning behavior (Saffran et al., 1996). We know that infants and adults are sensitive to many different kinds of statistical information in the environment including category distributions (Maye et al., 2002), variance within a category
(Maye et al., 2002), and transitional probabilities between objects or nonsense words (Hunt & Aslin, 2001). Further, this occurs with non-speech sound categories (Emberson, Liu, & Zevin, 2013; Holt, 2006), speech categories (Saffran et al., 1996), visual perceptual categories (Fiser & Aslin, 2002), and even more complex concepts like causal events (Oakes & Cohen, 1990). Additionally, learning of the statistical information is constrained by lower level perceptual organization, such that the lower level perceptual information constrains the higher-order statistics learned by the listener (Emberson et al., 2013).

We can observe that humans are sensitive to these different statistical regularities because they can behave in ways that are consistent with the underlying statistics in the input. The effects on behavior lead us to believe that this is affecting underlying psychological representations. This happens in a multitude of ways that reflect different aspects of the physical signal and can be manifested differently in behavior. We will go into detail of a few different kinds of statistical regularities and relationships and the effect on representations.

A popular application of statistical learning has focused on how experience with distributional information about the frequency of occurrence of particular sounds or events can aid learning of categories (however, there are some boundaries to this application, c.f. Cristià, McGuire, Seidl, & Francis, 2011; McMurray, Aslin, & Toscano, 2009). This can occur in both infants and adults and can also occur with only passive exposure to the category dimensions (Folstein, Gauthier, & Palmeri, 2010; Goudbeek, Cutler, & Smits, 2008; Maye et al., 2002; Pierrehumbert, 2003). Eimas (1975) found that infants were able to process sounds that could be distinguished categorically based on their acoustic distributions. These two to three-month-old infants were able to differentiate these acoustic categories based on their statistical distributions (Eimas, 1975). We know that the infants must have been able to process and store the regularities
present in the sound signal to be able to differentiate the categories. Thus, humans are able to process the sounds that belong to the same category or different categories even without verbal labels.

The variance or variability of the signal is another aspect to which humans are sensitive. If participants are exposed to passive variation along a particular dimension, it can impact category learning (Antoniou & Wong, 2016; Holt & Lotto, 2006). This passive exposure to the distributional variance can point to what is most relevant or irrelevant in the signal. High variation along a dimension might be a signal that this dimension is important and so participants may be inclined to place more weight on the dimension in a multidimensional category learning task (Holt & Lotto, 2006). However, experiencing variability along a dimension may also signal that the dimension is noisy and unreliable and participants may place less weight on the dimension during category learning (Rost & McMurray, 2010).

Further, there are general aspects of the statistical information of a physical signal to which learners are sensitive. Participants can be sensitive to the probability distributions of the dimension cues that are involved in categorization and respond in a way that reflects those distributions (Clayards, Tanenhaus, Aslin, & Jacobs, 2008). Learners can approximate category distributions, even from complex, non-Gaussian distributions (Gifford, Cohen, & Stocker, 2014). Finally, learners can be sensitive to the statistical structure even within given categories, rather than between categories, indicating that the statistical nature of what is learned goes beyond just boundaries, high-level categories or stimuli with verbal labels (Gureckis & Goldstone, 2008).

I also want to note here that this kind of sensitivity to statistical distributions can occur without overt attentional mechanisms. As I have mentioned, this sensitivity is present in infants
Additionally, people can be sensitive to statistical structure present in the physical stimulus even when it is not overtly reinforced or related to the feedback given during learning (Gureckis & Goldstone, 2008).

**Relationships between Dimensions.** In addition to statistical information regarding distributions along single dimensions, people are also sensitive to information about the relations between dimensions. Because of the sheer amount of information in the physical environment, some of the dimensions of representation are redundant. This leads to dimensions being represented in an interdependent way, as opposed to independently. However, sometimes those dimensions can provide unique information and are thus represented independently. Humans are sensitive to these different kinds of relationships and thus can represent the different kinds of interactions and relationships between dimensions. These different interactions can be a consequence of either perceptual or decisional processes (Silbert, Townsend, & Lentz, 2009).

Here, I will discuss two kinds of relationships that have been described in the literature: integral and separable dimensions and correlations between dimensions.

**Integral and separable dimensions.** A particularly useful distinction between different kinds of relationships between physical dimensions is Garner's (1974) distinction between integral and separable dimensions. Integral dimensions are dimensions that cannot be easily selectively attended to individually and are more often processed through holistic or similarity processes. Separable dimensions are dimensions which can be easily selectively attended to and analyzed independently.

There are many examples of these kinds of relationships in both vision and audition, as well as other domains. Garner and Felfoldy (1970) demonstrated that value and chroma dimensions are integral when they are present in the same visual object but can be processed
separably if they are present on two different objects shown simultaneously. Additionally, the horizontal and vertical positions of a dot on a screen are processed integrally (Garner & Felfoldy, 1970). Within the auditory domain, pitch and loudness are integral (Grau & Kemler Nelson, 1988), and information about timbre is also thought to be perceived integrally with both pitch and loudness (Melara & Marks, 1990a). Humans are also able to decisionally integrate information from multiple modalities, while keeping information from the two modalities separate (Hillis, Ernst, Banks, & Landy, 2002).

Garner (1976) thought of these two kinds of relationships as two ends of a continuum, where some pairs of dimensions can be more or less integral or separable, rather than being able to classify all dimension pairs as one or the other. The continuum of separability has been found in both visual dimensions (Garner, 1978) and auditory dimensions (Smith & Kilroy, 1979). This means that while integral and separable relationships occur in these two modalities, and also between them, there are other relationships that are best described as somewhere in between on the continuum.

With complex auditory dimensions, such as in speech, researchers have shown that some acoustic dimensions (F1, F0, and closure voicing) which contribute to the voicing contrast in speech are perceptually integral (Kingston, Diehl, Kirk, & Castleman, 2008). Other dimensions contributing to higher level speech categories can be distinguished in this way as well. In Mandarin Chinese, tone and vowel information are more integral than tone and consonant information (Tong, Francis, & Gandour, 2008). Additionally, these interactions are thought to be based on prior language experience. Lin & Francis (2014) found that native Mandarin Chinese listeners process consonant and tone in an integral way, whereas native English listeners process the dimensions in a separable way. This was thought to be due to the relation between
consonants and tone in Mandarin, which is a lexically tonal language compared to English, which does not use tone to indicate different meanings. Some researchers have also shown the importance of acknowledging perceptual interaction of real-world acoustic dimensions (Neuhoff, 2004).

Typically, if stimuli contain aspects from both visual and auditory modalities, they are often perceived separably (Melara, 1989). However, this is not always the case. Visually-conveyed place of articulation and auditory voicing information are perceived integrally (Green & Kuhl, 1991). This exception to the general case is interesting because both of these dimensions are necessary for the perception of speech, especially in difficult listening environments.

Several researchers have identified ways in which dimensions can have a special status in processing (Melara, Marks, & Potts, 1993; Smith & Kemler, 1978). This special status can affect how individuals are able to engage with the dimensions during certain learning challenges. Smith and Kemler (1978) posit that this can interact with experience and verbalizability: “Dimensions that we have difficulty talking about . . . are likely to have a low experiential status and be integral (p. 529). Melara et al. (1993) also call for the importance of experience: “Dimensions of experience [are] primary in perception” (p. 1102). Thus, the availability and importance of dimensions interact with an individual's experience with those dimensions.

Importantly, Garner (1976) and others have examined how the integral or separable nature of the relationship between pairs of dimensions can impact further processing. Most notably, researchers make a distinction between holistic and analytic processing (Foard & Kemler Nelson, 1984; Garner, 1974; Kemler Nelson, 1993). Holistic processing is more associated with integral dimensions and analytic processing with separable dimensions. Holistic processing requires that perceivers process the dimensions as unitary wholes, rather than
examining the dimensions independently. These decisions are based more on global similarity relations than anything about the individual dimensions (Foard & Kemler Nelson, 1984). In contrast, analytic processing allows perceivers to selectively attend to individual dimensions and ignore other dimensions. Analyzing dimensions allows for processing of objects in terms of dimensional components, rather than as a whole.

The constraints on processing based on the previously experienced relationships between the dimensions can persist even with expertise level training (Burns & Shepp, 1988). In this case, color experts and participants who were given specific training to analyze the dimensions separably were not able to process the visual dimensions of brightness and saturation separably (that is, they were not able to analyze the dimensions individually).

However, there are cases where participants are able to analyze integral dimensions and holistically process separable dimensions (Kemler Nelson, 1993). Rather than being a kind of mandatory processing as holistic or analytic when you have integral or separable dimensions, it may be the case that perceivers prefer to process these dimensions in a way that better matches their underlying representations of the dimensions. Thus, it may be difficult to separate those psychological dimensions in representational space because they are more similar in their representations, but it is not impossible at all times. Op de Beeck, Wagemans and Vogels (2003) demonstrated that discrimination performance between exemplars on a dimension improved for separable, but not integral dimensions. Thus, something about the representation of these dimensions constrains the ability to improve discrimination along the dimensions.

Dimensions may be warped in the transition from physical space to psychological space due to the nature of the neural processing mechanisms. The system may be forced to reduce redundancy and thus process information in a more efficient way (Lewicki, 2002; Schwartz &
Simoncelli, 2001; Tijsseling & Gluck, 2002). Thus, integral and separable dimensions, in experience, may force the system to reduce the redundancy in different ways. Integral dimensions require a representation that uses both dimensions, so information about the individual dimension may be lost. In contrast, separable dimensions require that information is preserved along both dimensions.

Whether dimensions can be described as integral or separable may very well depend on how they are processed in the brain. Drucker, Kerr, and Aguirre (2009) investigated how pairs of visual dimensions are represented in the brain. They found evidence for conjoint tuning of neurons involved in representing integral dimensions and independent tuning of neurons involved in processing separable dimensions. Thus, the nature of the neural processing may contribute to the integrality or separability of given dimensions. Melara and Marks (1990b) suggest that auditory dimensions can be classified as integral or separable based on whether a single processing channel is involved, or multiple processing channel are involved. Integral dimensions interact because the neurons involved in processing are overlapping and thus cannot be separated.

Studies using spatial location and frequency information of auditory tones demonstrated two stages of auditory processing—an early stage where these dimensions are treated as integral, and a later stage where they are separable (Dyson & Quinlan, 2004). Dimensions, thus, may demonstrate some differences in integral or separable status based on what individuals are being asked to do with the dimensions.

Due to the impact of the nature of the relationship between dimensions on the processing of those dimensions, we can begin to understand what effect exposure to these relationships can have on the psychological representations of these dimensions. If dimensions are perceived
separably because they do not interact in perception, we can understand how they may be stored separably in the brain. Different populations of neurons may represent these dimensions or at least the representations themselves may be orthogonal. When the relationship is integral, potentially the same groups of neurons are involved in processing the two dimensions. Thus, there is difficulty in selectively attending or analyzing those dimensions.

**Correlated and redundant dimensions.** Another kind of relationship is when the dimensions present in the signal are correlated or redundant. This may be a part of the reason why some dimensions are represented, and therefore processed, in an integral manner, but this might not always be the case. I have already briefly discussed a theoretical instance where there is redundancy present in the natural signal (Lewicki, 2002; Schwartz & Simoncelli, 2001). This redundant information can sometimes come in the form of two dimensions being highly correlated and thus the underlying representation makes use of this redundancy and represents the signal in a manner consistent with efficient coding principles. As evidenced above, this may contribute to the integrality of some physical dimensions.

Short-term passive exposure to a high correlation between two acoustic dimensions can affect the discriminability of sounds that obey the correlation and those which are orthogonal to the correlation (Stilp, Rogers, & Kluender, 2010). When participants were exposed to a high correlation between two initially independent and separable acoustic dimensions (attack/decay and spectral shape), Stilp and Kluender (2011) found that discrimination improved for stimuli that fell within the correlation and was lower for stimuli that violated the correlation. Stilp and Kluender (2012) found that the behavioral consequence of exposure to this high correlation was dependent on the nature of experience with that correlation, specifically how much exposure they had and how robust the correlation was. With more evidence and more robust evidence, there is a
stronger behavioral effect indicating that the underlying representations are encoding this information about people's experience and the relation between dimensions. This impacted their later performance with these dimensions, specifically how well they were able to discriminate stimuli that obey or disobey the correlation that was artificially created for purposes of the experiment. Sensory systems can rapidly adapt to these kinds of short-term correlations of sensory dimensions (Lu, Liu, Dutta, Fritz, & Shamma, 2019; McCollough, 1965).

While correlation and integrality may be related, they are separate constructs. Lockhead (1966) found that participants were faster during categorization when the dimensional information was redundant only when the dimensions were both correlated and integral. When the dimensions were only integral, there were no redundancy gains. This indicates that the concepts of integrality and correlation between dimensions may be related, but one does not fully predict the other.

Additionally, correlations in the current input of experience may better or worse match the correlations that the individual has previously experienced. Scharinger, Henry, and Obleser (2013) found that dimensional relationships that better matched the correlation people typically experience in speech were learned better than those who were learning a relationship that did not reflect prior experience. This may indicate that there is some relationship between prior experience and how those dimensions and relationships are encoded in representations.

Another kind of information that the physical signal carries that can be stored in psychological representations is information about the redundancy of dimensions. Biederman and Checkosky (1970) showed that the addition of redundant dimensions speeds processing of the visual dimensions of size and brightness. This redundant information provides more corroborating evidence to help the perceiver understand the physical world. The fact that having
redundant dimensions speeds processing indicates that this information is being represented in some way such that it benefits information processing.

**Behavioral Relevance**

In addition to the statistical structure present in the sensory world, the world also contains information about what is behaviorally relevant. Understanding how humans are sensitive to what is behaviorally relevant aids in our understanding in how this information is processed and stored in representations.

Being sensitive to behaviorally relevant information is an evolutionarily adaptive process. Understanding what information is relevant for behavior lowers uncertainty, helps us avoid risks, and increases the probability of future reward. This information is incredibly valuable and is the cornerstone of understanding of human experience and behavior.

An interesting perspective on how high-level information, like behavioral relevance, impacts perception is Palmeri and Gauthier's (2004) view that there is an interaction between perception and cognition. It benefits our understanding of behavioral relevance to looking at these two processes as interacting, rather than modular. We also adopt this view and the purpose of this section is to expound on the relationship between perception and cognition in the context of task relevance and the importance of previously encountered information about what is behaviorally relevant. We note that the influence of cognition on perception has been a topic of recent debate (Firestone & Scholl, 2016; Lupyan, 2015).

We will first explain task demands and other contexts that constrain perceptual representations. We will then explain how different aspects of prior experience with dimensions or stimuli impacts future behavior with regard to those dimensions.
**Task demands.** The learner’s goals, task, and motivation can affect what they attend to in the environment (Love, Medin, & Gureckis, 2004). When we interact with information in a learning task, we are doing it for a purpose. Rosch’s (1978) concept of ‘cognitive economy’ suggests that our goal is to extract the relevant information using as few cognitive resources as possible. Thus, the task we use to train learners can change the actual goal of the learner. The task used to query learners’ knowledge can impact how we understand what they know. Additionally, different training tasks can direct attention to different parts of stimuli or categories and these different targets of attention impact perceptual representations of the dimensions or categories.

One classic example of how different training tasks can lead to different representations is the comparison of inference training and classification training. Classification training requires participants predict labels for a category exemplar. For instance, if the participant was learning different categories of bugs, they might be asked, “Is this a beetle or a ladybug?” Inference training requires participants predict missing features of a category exemplar. For instance, if a participant is learning the different categories of bugs, they might be asked, “If this was a beetle, what would its antenna look like?”

Inference and classification training lead to different outcomes in terms of representation in memory (Markman & Ross, 2003). Classification training leads to representations that represent structure in terms of the rules defining the relevant feature (what really defines a beetle versus a lady bug?). Inference training leads to representations that are structured in terms of the similarity between exemplars (what do beetles generally look like?).

Classification training also leads to a narrower attentional scope (Hoffman & Rehder, 2010). A narrow scope refers to the attention placed on individual features rather than on global
similarity (Treisman, 2006). This narrow scope allows participants to identify diagnostic dimensions and succeed in a category learning task. However, this narrow scope keeps participants from being flexible enough to redirect their attention if another aspect of the signal becomes relevant. Inference training, in contrast, induces a broader attentional scope (Hoffman & Rehder, 2010). This broad scope allows for more flexible allocation of attention if some previously irrelevant dimension becomes relevant. Additionally, this broad scope allows participants to represent more complex aspects of the stimuli, such as correlations between features or understanding of category typical features (Hoffman & Rehder, 2010; Jee & Wiley, 2014). Research has shown that classification allows for representation of the distribution of atypical and typical features (Jee & Wiley, 2014).

This kind of research has also been extended into real-world classroom environments (Sakamoto & Love, 2010). These two formally equivalent tasks have vastly different learning outcomes. Inference training led to fewer errors and better retention of category exemplars compared to classification training for fifth grade students learning categories of shark breeds (Sakamoto & Love, 2010). Thus, understanding how training task impacts memory and psychological representations for category members, dimensions, and entire categories is important to understanding real-world learning challenges.

Attentional scope also can change with development (Deng & Sloutsky, 2015). Children apply a broader attentional scope, even when learning categories that can be distinguished by a single diagnostic features (Deng & Sloutsky, 2016). Children have better memory for features that are non-diagnostic than adults because they have a broader attentional scope (Deng & Sloutsky, 2016).
Beyond differences induced by inference and classification training, other aspects of the task can also influence how people represent the dimensions and stimuli involved in learning. For instance, in tasks that make participants focus more on the categorical nature of the stimuli (such as grouping exemplars into different categories) highlights features that distinguish the categories. In contrast, moving participants’ focus to the individual nature of stimuli and having them differentiate between stimuli, even within categories, highlights many features of the stimuli. Category learning increases attention to the relevant dimension, whereas naming objects from the category boosts attention to multiple dimensions, including the dimension that was irrelevant for categorization (Van Gulick & Gauthier, 2014).

This difference in training affects representation of the dimensions in a way that affects behavior and neural representations. For instance, training participants on individuation or categorization changed how trained objects were represented in visual cortex (Wong, Palmeri, Rogers, Gore, & Gauthier, 2009; Wong, Folstein, & Gauthier, 2012). Learning to group some objects into one category focuses attention on what makes those objects similar and what makes them different from other categories (Goldstone, 1998).

People are sensitive to the informativeness of a dimension in a particular learning context. Learners are able to extract which dimension is relevant for categorization and use that dimension, rather than other irrelevant dimensions (Holt & Lotto, 2006; Idemaru, Holt, & Seltman, 2012). Usually, this process is guided with feedback, but can also be learned in an unsupervised manner (Ashby, Queller, & Berretty, 1999; Goudbeek et al., 2008; Toscano & McMurray, 2010). While learners are able to focus on the most informative dimension in their decisions, it is sometimes still difficult to disengage completely from an irrelevant dimension (Holt & Lotto, 2006).
More general differences in context can also impact how people represent information and this impacts their behavior. Heuer and Schubö (2016) demonstrated that which action people were preparing to take (grasping or pointing at a visual object) can affect which visual dimension they relied on. When people were preparing to grasp the object, they paid more attention to size, whereas when people were preparing to point to the object, they paid more attention to color (Heuer & Schubö, 2016). The context of information affects behavior. When different information is relevant in different contexts, the information can be represented based on how it was relevant in the given context. If attention is directed more to one dimension than another during learning, that dimension may be weighted more in representation. In sum, different task demands differently affect memory and representation.

**Prior experience.** In addition to task demands creating constraints on representation, prior experience with dimensions or objects or categories in a given context can change how they are represented and thus how people interact with them in the future, even in different contexts.

General prior knowledge about existing categories can influence how people learn new categories. Kaplan and Murphy (2000) demonstrated that even minimal prior knowledge about higher level categories (e.g. birds have wings) affects how new categories are formed. My interest is primarily focused on perceptual categories rather than conceptual categories, which may be more affected by prior knowledge. However, participants bring many different kinds of experience into an experiment, much of which we, as experimenters, do not expect them to have and do not probe which prior categories they have. Understanding that general prior knowledge can affect learning of new categories is critical to uncovering what about existing experience affects representations.
Once in an experimental context, even passive exposure to stimuli or to variability along dimensions can influence category learning behavior (Holt & Lotto, 2006; Schyns, Goldstone, & Thibaut, 1998). Simply being exposed to the categories or dimensions can have profound influence on how people approach category learning problems. For instance prior exposure can highlight individual dimensions or the physical space of the categories. Holt and Lotto (2006) found that after passive exposure to variability along a dimension, participants placed more weight on that dimension in a category learning task. If participants have more meaningful experience with the stimuli or dimensions than simple passive exposure, it can color their representations potentially even more than passive exposure.

The specific goals of learners and their past experience affect what they pay attention to during categorization. Lynch, Coley, and Medin (2000) demonstrated that experts and novices differed in how they categorized taxonomic tree categories. Novices based their responses on familiarity with the exemplars, whereas experts used the optimal dimensions that maximally differentiated the tree categories (height and weediness). Besides using the optimal dimensions, experts are also thought to have more access to features than novices (Tanaka & Taylor, 1991). Thus, the more you interact with something and know the categories, the better able you may be to access the dimensions distinguishing the categories.

I will address several different kinds of information that can possibly make up a learner’s experience with stimuli or dimensions, so much so that it can be fused in their representations of those stimuli or dimensions. These are cue validity, reliability, relevance, salience, reinforcement history, verbalizability, and language experience.

**Cue Validity and Reliability.** Cue validity refers to how well a given cue or dimension predicts an expected response. People are sensitive to the relative validity of cues (Le Pelley,
Beesley, & Griffiths, 2014). When cues, or dimensions, are more valid, they will be given more weights in perceptual judgements (Himmelfarb, 1970; Holt & Lotto, 2006).

Another aspect of the incoming signal that can be represented based on its statistics is the reliability of information, and specifically of a dimension in a given context. Toscano and McMurray (2010) demonstrated that human cue weighting is not always statistically optimal. By this they mean that sometimes people will use cues, or dimensions, sub-optimally in a given context. They will not use the most relevant dimension or potentially, they will place more weight than is optimal on less informative dimensions. When weights on these cues are not statistically optimal, the reliability of the cue in the context can explain why it is not optimal. Cue reliability can be understood as how reliable the predictions that are based on that cue are for a given behavior in a given context, giving the lowest variance for a given perceptual estimation (Ernst & Banks, 2002; Fetsch, Pouget, DeAngelis, & Angelaki, 2013; Jacobs, 1999) and the reliability of cues is something that can be learned (Atkins, Fiser, & Jacobs, 2001). When learners do not weight cues optimally based on their information alone (i.e. relevance of the cue in the context), their weighting can be explained based on how reliable that cue has been (Toscano & McMurray, 2010).

**Relevance.** Relevance refers to the diagnosticity of a dimension in a category learning task or whether a dimension was even relevant to behavior during any learning task. By relevance, I refer to anything that aids and drives behavior, rather than being ignored, or more generally, something that is tied to the feedback a learner is given.

People are sensitive to what is relevant in tasks and with feedback and learn to direct their attention to the relevant cue, in a process known as ‘diagnosticity-driven learning’ (Schyns et al., 1998). Learning which dimensions are relevant makes those dimensions more salient in working
memory (Wagar & Dixon, 2005). Even when dimensions become irrelevant, previously relevant dimensions are easier to switch to than previously irrelevant dimensions (Kruschke, 1996; Roberts, Robbins, & Everitt, 1988). In new category learning environments, people will also direct their attention to dimensions that were previously predictive in other environments (Kersten, Goldstone, & Schaffert, 1998).

With category level feedback, people are able to shift their attention from one cue to another depending on whether the cue is relevant (Francis, Baldwin, & Nusbaum, 2000). Participants become more sensitized to relevant features and less sensitized to irrelevant features after category learning (van der Linden, Wegman, & Fernandez, 2013). For these visual categories, category learning also led to dimension-based changes to representations in both frontal and occipitotemporal cortices, with both areas increasing in sensitivity for the informative features after training (van der Linden et al., 2013).

Ahissar and Hochstein (2004) argued that top-down, task-relevant information about what is relevant drives visual perceptual learning such that task-relevant information is boosted and irrelevant information is pruned. When something is task and behaviorally relevant, it comes to the forefront of attention and thus is what learning is centered on. This centering of attention on what is relevant means that that is what is primarily reinforced during learning and thus more strongly associated with the behavior and outcome in perceptual representations.

Rodrigues and Murre (2007) argue that the dimensions relevant for learning change in their representations across the course of learning. New dimensions can be uncovered by learners based on what is relevant during learning and can continue to change in a flexible manner based on experience (Schyns & Rodet, 1997).
While learners are sensitive to what is relevant and what is relevant to behavior does get a boost during learning, this does not necessarily mean that people ignore what is irrelevant. Huettel and Lockhead (1999) demonstrated that participants were slower in their responses when there was large variation in a dimension irrelevant for categorization. These participants were affected by the irrelevant dimension and thus must have been storing and processing it in some way.

*Salience.* Salience refers to how a dimension pops out against other dimensions or the propensity of learners to use a given dimension over others during learning. Learners are sensitive to the salience of cues (Le Pelley et al., 2014).

Some research has shown that learners process which dimensions are relevant for categorization, regardless of their salience (Guest & Lamberts, 2010). While salience does have an impact on learning—particularly on what is brought to the forefront of attention initially, participants are still drawn toward what is relevant in the signal for that task.

Importantly, salience is something that can be learned and can change over the course of learning (Livingston & Andrews, 1995; Livingston, Andrews, & Harnad, 1998).

*Reinforcement history.* Reinforcement history refers to the prior experience learners have with given dimensions, categories, and stimuli regarding reward or punishment. The dimensions or information that lead to reward or lead to the highest reward are often the relevant dimensions. Many times, this is how relevance is determined—by what leads to a rewarding outcome.

Recent research has shown that the neural response to stimuli is to reduce dimensionality to the dimensions in the signal that are relevant to predicting future reward (Niv et al., 2015). The information that is strongest in representation is what leads to the highest reward.
Whether or not stimuli are treated as functionally equivalent can be partly determined by their reinforcement histories (Honey & Hall, 1989). If two stimuli both lead to the same rewarding outcomes, they may be more likely to be perceived as functionally and perceptually equivalent than two stimuli which led to different outcomes.

Researchers have found that stimuli that have been previously associated with reward remain salient even when they are no longer relevant to the task and need to be ignored (Anderson, Laurent, & Yantis, 2011a, 2011b). Reward and reinforcement histories have a strong influence on the salience and availability of individual dimensions and stimuli. This can affect perception and learning, even when what was rewarded is irrelevant.

**Verbalizability.** The verbalizability of a dimension refers to the ability of a participant to explain or use words or label what they are perceiving or experiencing. If a label exists for a dimension or category, it affects how it is learned and how well people are able to use the dimension or learn the category.

The words we are able to use to label dimensions affect our space of object representation and how we approach different tasks. The ability to interact with new category learning spaces is a consequence of our prior experience and which distinctions we find important (i.e. relevant) and which we can find words to describe (Austerweil & Griffiths, 2001; Perry & Lupyan, 2014).

Wisniewski and Medin (1994) demonstrated that when learners were given meaningful labels to different categories, they interpreted the stimuli based in their own intuitive theories. These theories create a space of prior expectations through which participants can look for abstract features that support their theories rather than just searching through the perceptual features that distinguish the categories. Thus, labels and prior expectations shape our hypothesis space and how we approach category learning problems. Similarly, Fotiadis and Protopapas
(2014) found that the ability to name a cue or dimension affects the hypothesis testing underlying category learning and aids category learning that requires selective attention to a single dimension.

The importance of verbal labels is a pattern that is also evident when comparing the category learning of young children. Smith, Gasser, and Sandhofer (1997) posit that as language develops, children learn the labels for dimensions and learn how to describe those dimensions verbally. As they learn these labels, they can then employ selective attention to those dimensions. Thus, the label on the dimension and the ability to verbalize it changes how children are able to interact with those dimensions. This is likely also a process that affects learning in adults for much more complex categories, for which there are no clear verbalizable dimensions, like with speech categories.

When experimenters gave adult learners explicit dimension primes during a Mandarin tone category learning task, they found that telling participants to pay attention to the relevant dimension improved their learning (Chandrasekaran, Yi, Smayda, & Maddox, 2016). Even so, participants were not able to fully disengage from the irrelevant dimension in this case. When giving participants a verbal label that describes the relevant dimension, it remains difficult to direct attention to that dimension in the context of speech category learning.

**Language experience.** Language experience refers to the experience learners have with language-specific regularities and categories that people encounter based on which languages they speak or understand.

Languages have specific regularities associated with different sound and phonetic categories that people experience in them (Hillenbrand, Getty, Clark, & Wheeler, 1995; Maye et al., 2002; McMurray & Aslin, 2005). Sensitivity to acoustic features that are not reinforced by
native language experience are lost over language development (Pisoni, Aslin, Percy, & Hennessy, 1982; Werker & Tees, 1984). Before the age of 10 months, infants are able to distinguish the sounds of the world’s languages (Werker & Tees, 1984). After 12 months, this general sensitivity diminishes but infants improve on their ability to distinguish native-language contrasts (Kuhl et al., 2006; Werker & Tees, 1984). Thus, after one year old, humans are much more sensitive to the language that they continue to experience than languages for which they have minimal or no experience.

Native language experience affects how listeners are able to use dimensions during novel category learning. For example, native language experience affects how participant use duration and spectral cues in processing speech and this evident in pre-attentive measures like the MMN (Lipski, Escudero, & Benders, 2012). Due to the regularities present in Japanese, native Japanese listeners use a suboptimal acoustic cue (F2 onset frequency) when distinguishing English /r/-/l/ categories. Specifically, this difficulty is driven by the existence of categories in Japanese that make use of that dimension and not the F3 onset frequency dimension, which would be ideal for distinguishing the English categories (Lotto, Sato, & Diehl, 2004).

Further, some theories posit that how well a non-native speech category will be learned depends on how similar or different it is to the existing native-language categories (Best, McRoberts, & Goodell, 2001). If the non-native contrast is phonologically equivalent to existing categories, then learning should be easy, whereas if the non-native contrast conflicts with existing categories, learning should be very difficult.

Even non-speech categories are learned in a way that reflects representations built up based on native language experience (Xu, Gandour, & Francis, 2006). Thus, experience with
dimensions in an individual’s language experience may distort their perception of those dimensions even outside of a speech context.

Language experience also affects whether or not dimensions will be perceived separably or integrally (Lin & Francis, 2014). Thus, this is a learned distinction that is affected by experience and thus experience tags dimensions and is able to fuse them, depending on the nature of that experience. For instance, native Mandarin Chinese listeners process consonant and tone dimensions integrally, whereas native English listeners process them separably (Lin & Francis, 2014).

The weighting of acoustic cues in speech categories changes with language experience in development (Nittroer, Manning, & Meyer, 1993). As children get older, they begin to weight dimensions for native speech category distinctions more optimally (Nittroer, 1996). With their more limited language experience, children give equal weight to acoustic dimensions in many different contexts, whereas adults will change their weights based on whatever is relevant in the given context (Nittroer & Miller, 1997). With more language experience, children gain and are able to use this flexibility in a more adult-like manner.

Additionally, learners are sensitive to correlations between dimensions that reflect the correlations they have experienced in their language experience. Scharinger et al. (2013) showed that learning was better for non-speech categories that had correlations between dimensions that best reflected the correlations between dimensions of speech categories. The encoded correlations of experience thus affected novel category learning.

Further, other non-language acoustic experience that learners have can affect their performance in learning tasks. Music experience modulates lower level auditory abilities that
may affect higher-level learning, such as category learning (Strait, Kraus, Parbery-Clark, & Ashley, 2010).

**Effects on psychological representations.** There are multiple potential effects on psychological representations induced by these differences in task demands and prior experience. Due to the task or the specific prior experience, the system is processing those dimensions, categories, or stimuli as more relevant to behavior than they had been. We have also summarized how the scope of attention toward dimensions can also be modified. These effects have been examined both on individuals’ behavior and on neural representations. We will discuss these in turn now. Additionally, specific effects of categorical perception can be induced based on training and experience. We will summarize this at the end of this section.

**Behavioral effects.** One of the most typical effects seen as a consequence of training or prior experience is that the representations of dimensions are warped in some way, relative to pre-training or pre-exposure levels. This has been discussed as a shrinking or stretching of psychological dimensions through processes of acquired distinctiveness and acquired similarity (Goldstone, 1998; Lawrence, 1949, 1950).

Acquired distinctiveness refers to the effect induced by category learning that increases discriminability along the category-relevant dimension (Goldstone, 1998; Lawrence, 1949, 1950). This leads to the entire dimension or part of a dimension being stretched in psychological representational space. This affects discrimination performance and also future learning.

Acquired similarity or equivalence refers to the effect induced by category learning that decreases discriminability along the category-irrelevant dimension (Goldstone, 1998; Lawrence, 1949, 1950). This leads to the entire dimension or part of the dimension shrinking in psychological representational space. This worsens discrimination performance but allows same-
category members to be perceived as more similar to one another. This forms the basis of within-category discrimination being reduced after learning, which is a hallmark of categorization. These acquired equivalence effects are thought to be stable, long-term effects of learning that directly modify cortical representations (Folstein, Palmeri, Van Gulick, & Gauthier, 2015). However, evidence for acquired equivalence has been mixed. Folstein, Palmeri, and Gauthier (2014) were unable to find acquired equivalence after category training with separable visual dimensions.

Acquired distinctiveness has been found with both visual and auditory dimensions (Folstein, Palmeri, & Gauthier, 2014; Goldstone, 1998; Goldstone, 1994; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999; Hockema, Blair, & Goldstone, 2005; van der Linden et al., 2013; Van Gulick & Gauthier, 2014). This can occur with separable or integral dimensions and can even alter the classification of the dimensional relationship. For instance, Goldstone (1994) found that integral dimensions can gain acquired distinctiveness with training and become differentiated or non-integral. Dimensions may, in fact, gain these statuses based on training or other experience (Goldstone & Styvers, 2001; Soto & Ashby, 2015). Additionally, this acquired distinctiveness can have an effect on the entire dimension that is relevant or only part of the dimension by creating a local sensitization (Goldstone, 1994; Van Gulick & Gauthier, 2014).

Another specific kind of warping on dimension levels has been found to occur around phoneme category prototypes (Kuhl, 1991). Native language experience shapes representational space of the sound categories. This concept has been labeled ‘perceptual warping’ and relates to Kuhl’s (1991) perceptual magnet theory. This theory poses that category prototypes (the best exemplar of a category) are the area of perceptual space where perceptual distance between exemplars is smallest (i.e. shrinks) and areas of worst instances are stretched (Iverson & Kuhl, ...
The prototype acts as a magnet that pulls nearby exemplars toward it and shrinks perceptual distances between them.

The perceptual magnet and its effect on perceptual space is thought to be shaped by early language experience that creates this warping within the representational space around the experienced sound categories (Kuhl & Iverson, 1995). This has been shown to affect not only the native language categories, but the dimensions that are available in the signal to distinguish those categories (Pajak & Levy, 2014). Listeners demonstrated a general sensitivity along dimensions that were relevant for distinguishing native language sound categories (Pajak & Levy, 2014).

This warping of dimensional space can be instantiated through connectionist networks that show that the warping may occur to reduce the amount of redundant information in the signal and process only what is behaviorally relevant and informative (Tijsseling & Gluck, 2002). In some ways, this warping can be thought of as an outcome due to the perceptual system constraints, perhaps instantiating a kind of efficient coding. Importantly, the consequence of these processes is that psychological representations are warped in some way that can then affect future behavior, such as category learning.

**Brain-based effects.** Concurrently with, and most likely contributing to, the behavioral effects, this kind of learning and modulation of the dimensions also cause neural changes.

As we discussed the specific task that is used can change the scope of attention and thus what is represented in memory (Hoffman & Rehder, 2010). Different training regimens that have different task demands can also affect representation of objects and features in visual cortex (Wong et al., 2009; Wong et al., 2012).

Accompanying the acquired distinctiveness effect that is caused by category learning, representations in visual cortex (anterior fusiform and extrastriate occipital areas) are sensitive to
what is relevant (Folstein, Palmeri, & Gauthier, 2013). Representations in these areas becomes more sensitive to small variations along category-relevant dimensions based on what the relevant dimension is (Folstein et al., 2013).

Category learning is also known to cause drastic changes in representation in inferior temporal (IT) cortex when learning visual object categories (Op de Beeck & Baker, 2010). This can be related to the dimension that is relevant during the task. The representations of the relevant dimension become expanded in neural representations (Op de Beeck & Baker, 2010). The selectivity of shape-selective neurons for both relevant and irrelevant dimensions increased after category training for visual categories in IT (De Baene, Ons, Wagemans, & Vogels, 2008). While sensitivity for both irrelevant and relevant dimensions increased in IT representations, indicating that exposure to dimensions is relevant, this increase was larger for the relevant dimension, indicating that behavioral relevance specific to the dimensions is encoded in the representations of those dimensions. Representation of features in IT is thought to be shaped by categorization experience and specifically what is diagnostic gets enhanced in representation relative to non-diagnostic dimensions (Sigala & Logothetis, 2002). Category training can also lead to the sharpening of representation for object features in lateral occipital cortex (Jiang et al., 2007).

Representations of frequency in primary auditory cortex have also been shown to be able to increase in sensitivity with more training experience (Recanzone, Schreiner, & Merzenich, 1993). They found that cortical representations and the sharpness of tuning increased with training and that this correlated with frequency discrimination performance. In other words, as frequency discrimination increased (a kind of acquired distinctiveness), representations grew in primary auditory cortex. Both primary and secondary auditory cortex can demonstrate shifts in
receptive fields based on feedback experienced during learning (Weinberger, 1993). When dimensions are relevant for behavior and thus are attended to, there are also sharpening of neural responses in primary auditory cortex (Connell, Barczak, Schroeder, & Lakatos, 2014).

The perceptual magnet effect and perceptual warping also have a neurally plausible instantiation (Guenther & Gjaja, 1996). In this model, there is an auditory map that is governed by the principles of population vectors, wherein single neurons have preferred frequencies for which they will fire more strongly and frequently and average together to encode the exact frequency. They found that in this model, the auditory neural map was warped in a way that mirrors the perceptual warping and perceptual magnet effects. There were peaks in the distribution map of cell firing preferences around the prototypes (which act as the perceptual magnet) and then responses of areas in the map that are close to the prototype are more similar to one another than those further away (Guenther & Gjaja, 1996).

**Categorical perception.** Categorical perception can be thought of as a special kind of acquired distinctiveness that occurs at a category boundary (Goldstone, 1998), where everything within the category gains an acquired equivalence, but at the boundary, there is a sharp distinction or distinctiveness between the categories.

This was thought to be a special component of speech, being defined as “a mode by which stimuli are responded to, and *can only be responded to, in absolute terms*” (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970, p. 234, italics in original). Categorical perception has since been found in other instances and modalities, though speech does remain an incredibly interesting example of categorical perception.
Acquired distinctiveness adds to the categorical perception nature of speech categories. Liberman, Harris, Hoffman, & Griffith (1957) demonstrated that discrimination of speech sounds is better at the categorical boundaries than in the middle of the category distributions.

This can also affect how participants are able to learn new categories even in adulthood. Myers & Swan (2012) demonstrated that when learning the dental/retroflex speech contrast, native English speakers learn to become sensitive to between-category differences and can gain a sense of categorical perception.

Categorical perception can be independent of speech and has been shown to be determined by the acoustics, rather than the phonetics, of the sound (i.e. The same thing was found with non-speech correlates of the sounds as with speech; Eimas, 1975). While categorical perception has been discussed by some as being a critical feature of speech categories, there is also evidence that even speech categories contain more fine-grained within-category structure. That is, not all members of a speech category are perceived to be equally good exemplars of that category (Iverson & Kuhl, 1996; Kuhl & Iverson, 1995)

Conclusion

The perceptual processes through which a signal travels from the physical world to ultimately be represented in human psychological representations are sometimes relatively straightforward. For example, at some levels of processing, mental representations can reflect a specific color, as in the cones in the retina, or the orientation of a line in the retinal ganglion cells, or frequency on the basilar membrane. However, the transformation or processing of this information is often not this straightforward. Aspects of the physical signal are sensitive to what is being selectively attended, the statistical regularities of the signal, the learner’s past experience and what is behaviorally relevant.
Section 2: Impact on Category Learning

In Section 1, I discussed how internal, psychological representations reflect the physical world, but also that representations are impacted by statistical regularities of the signal, the behavioral relevance of information, and prior experience of the learner. Through these means, the physical world becomes transformed into psychological representations that are warped, containing different information than simply what was present in the physical signal.

In this Section, I will review research that demonstrates how these transformations of the physical signal to psychological space subsequently affect behavior in new learning environments. Behavior is dependent on how similar the new environments are to experienced environments (and thus psychological representations). I will examine the implications of having these mismatched environments and ways in which prior experience and transformations can affect our ability to interact with dimensions for new learning.

Environments that have contributed to shaping psychological representations can either be consistent or inconsistent with newly encountered learning environments. These new environments can be experimentally defined with assumptions of researchers or these new learning environments can exist in natural environments. Either way, it is not always the case that newly encountered environments will contain the same regularities, relationships, or relevant information that previously encountered environments have. When these environments contain different information, existing psychological representations are warped in a way that is inconsistent with new learning environments.

This inconsistency does not necessarily make learning impossible. However, it may make the learning challenge more demanding and much more difficult. I will first consider some
different examples of consistent and inconsistent environments to better define this distinction, which is not typically considered in existing theories of category learning.

Consistent correspondences between new physical environments and existing psychological representations appear often in the literature. This is typically the assumed relationship between physical and psychological representations—a consistent, compatible, one-to-one, orthogonal correspondence (Johannesson, 2001; for exceptions c.f. Cheng & Pachella, 1984; Goldstone & Styvers, 2001; McKinley & Nosofsky, 1996). However, this assumption is usually not made outright or obvious. The underlying assumption is that whatever we, as researchers, are asking participants to do in a task, they will be able to adapt and learn because it is not incompatible with their representations in any major way. This includes learning to selectively attend to separable dimensions in a category learning task. For example, previous experience that tells learners that line width and line orientation vary independently and are thus represented separably will make it easier to selectively attend to these dimensions in new learning tasks. Representations reflect that independent processing of the dimensions. In the new environment, when learners are asked to selectively attend to these dimensions, it is compatible with their previous experience. The previous experience fits their expectations and they do not need to further transform the signal to selectively attend to the separable dimensions. Thus, learning is fast and accuracy is high (Ashby & Maddox, 2005, 2011).

Inconsistent correspondences between new physical environments and existing psychological representations also occur, but arguably less frequently. Because researchers often make the assumption that there is a one-to-one correspondence between the physical world and psychological representations, it is often not even considered that there may be a mismatch here. However, when learning proves to be extremely difficult or impossible in the course of an
experiment, the cause may be that the physical learning environment is inconsistent and thus incompatible with existing representations. A clear example of an inconsistency between a physical learning environment and trained psychological representations comes from the non-native speech category learning literature. When learning English /r/-/l/ sound categories, native Japanese speakers show profound difficulty (Yamada & Tohkura, 1992). The distinction between these two English speech categories can be nearly perfectly described with the third formant (F3) onset frequency cue (Lotto et al., 2004). A secondary cue, the second formant (F2) onset frequency cue can also be used to partially dissociate the two categories; however, it is not fully informative (Lotto et al., 2004). In Japanese, these two categories do not exist. Rather, there is a single native Japanese category called the rhotic flap that overlaps the English /r/-/l/ category space along the F2 onset and F3 onset frequency dimensions (Lotto et al., 2004). Native Japanese listeners fail to discriminate between the English /r/-/l/ sounds because they are using the suboptimal F2 onset frequency cue to distinguish the categories (Iverson, Kuhl, Akahane-Yamada, & Diesch, 2003). We will return to this example shortly.

Another example of this mismatch between existing representations and new learning environments is the case where a new task requires selectively attending to dimensions that are represented integrally. Integral dimensions are thought to possibly be able to be selectively attended to with greater cognitive effort (Smith & Kemler, 1978). Thus, learning is not always completely impossible when this mismatch occurs, but it might be extraordinarily more difficult.

Some of the most interesting and profound learning challenges may occur because of the inconsistency or lack of a correspondence between the physical world and psychological representations. It is important to consider what the underlying representations are in order to understand how we might be manipulate them to help speed up learning when this does occur.
Without the full understanding that this difficulty exists, we might dismiss a category learning problem as impossible without understanding or probing the reason why. However, if we consider that there may be some incompatibility with the underlying representations due to previous experience with other environments, we may be able to gain more insight into how these problems arise and how to overcome them.

Now that I have explained why this issue is important and better defined the situations in which this incompatibility might occur, I will explain some ways in which having warped psychological space may influence learners’ ability to perceive and use dimensions during different learning tasks.

First, I will consider the effects of this warping on the perception or simple availability of dimensions. This can happen in three ways: affecting the salience or bias for particular dimensions, affecting the verbalizability of dimensions, and the presence of latent dimensions in the signal.

**Effects on perception of dimensions**

*Bias or preference for dimensions.* Changes in the psychological representation of dimensions that occur due to prior experience with those dimensions can affect the use of dimensions: specifically, participants initial preference for dimensions in category learning tasks.

More salient dimensions are more readily available in a learning task. Learners are more likely to access and use dimensions if they are salient. Critically, salience can change with learning and thus affect learners’ ability to use those dimensions in the future.

General demands of a task can increase the salience of individual dimensions (Lawrence, 1949). This trained salience, caused by previous experience with cues, can extend to a new environment with the same cues (Lawrence, 1950). Specific kinds of interactions with
dimensions during learning can also impact the salience of those dimensions. For instance, reporting the value of a feature during learning and thus accessing it in short-term memory increases the ability to use that feature (Fan & Turk-Browne, 2013). When the feature is accessed in short term memory, the representation in long-term memory is biased toward that reported dimension and away from any unreported features (Fan & Turk-Browne, 2013).

In the category learning context, previously relevant dimensions can be more salient than previously irrelevant dimensions or dimensions which were not present in a previous learning environment (Kersten et al., 1998; Livingston & Andrews, 1995; Wagar & Dixon, 2005). This can be thought of as affecting the hypothesis space that participants try to access during learning (Livingston & Andrews, 1995) and more specifically affects object representation in working memory (Wagar & Dixon, 2005).

Other prior experience with dimensions or features can also impact the salience of those dimensions or other cues. For instance, when salient stimuli have been previously associated with reward, there is a magnification of salience of the that stimulus (Anderson et al., 2011a).

A more specific kind of experience can also impact what is available and more salient in new learning environments. Native language experience can change which dimensions are more salient or more likely to be used in non-native speech category learning (Aliaga-Garcia, 2010; Lipski et al., 2012). For instance, Lipski et al. (2012) demonstrated that Spanish listeners are more likely to use duration cues more than spectral cues when distinguishing Dutch vowels, whereas Dutch listeners use spectral cues more. This difference can be attributed to the differences in the regularities of these two languages and thus the experience of the listeners in their native language.
These native-language specific differences can affect how listeners use cues during learning. Furthermore, within-subject variability about which dimensions are preferred during category learning can be high (Christensen & Humes, 1996). One important aspect that may influence which dimensions learners have a preference for could be their prior experience with those dimensions.

**Verbalizability of dimensions.** The ability to verbally identify, or verbalize, dimensions also may be changed with experience. The psychological representation of some dimensions may correspond with the ability to verbalize the dimensions. However, this is not always the case.

If dimensions are more verbalizable or more easily described by a verbal label, they are more available to learners in a task. With more verbalizable dimensions, participants can expand their hypothesis space of possible relations between the categories and may more easily find rules to describe the categorical distinction. Verbalizability is also something that can change with learning and affect the ability of learners to use those dimensions during subsequent learning.

Ashby and colleagues distinguish between kinds of categories that require a verbalizable rule and those which cannot be described verbally (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005). For a rule to be able to be described verbally, Ashby and Maddox (2005) describe three requirements: that there is a semantic label for each dimension, that selective attention to the dimension is possible, and that the rule itself can be described verbally. Thus, the verbalizability of a dimension affects how it is able to be used by participants.

Dimensions that have linguistic labels or which are easy to describe verbally may be more accessible or intuitive for learners. Pothos & Close (2008) demonstrated that classification rules which are more intuitive will be the ones that learners prefer to use. If more verbalizable
rules are more intuitive, this could be an explanation of why the presence of linguistic labels or verbalizable rules aids learning.

Learners can also benefit from learning verbal labels for dimensions during learning. Smith et al. (1997) posit that psychological dimensions are a product of young children learning to describe dimensions verbally. Thus, what is represented psychologically may be directly tied to what has been experienced and given explicit verbal labels.

Verbal, linguistic labels may also interfere with lower level perceptual processing. Melara and Marks (1990c) demonstrated that verbal labels such as “high” and “low” led to Garner interference with perceptual dimensions which could either be consistent or inconsistent with those labels—i.e. auditory pitch and position on a screen. Thus, labels are not a part of an independent process that has no effect on perceptual processing. Lower level processing can be impacted by linguistic labels and their availability.

Additionally, giving learners the explicit labels during category learning can direct their attention toward those dimensions and thus improve category learning (Chandrasekaran et al., 2016; Perry & Lupyan, 2014). However, it is not always possible to direct attention to perceptual dimensions with verbal labels or explicit instructions (Francis et al., 2000). This may especially be true for acoustic dimensions and integral dimensions, which may be more difficult to describe with simple verbal labels.

Learning artificial verbal labels for dimensions also impacts learning because it impacts the hypothesis space of learners (Fotiadis & Protopapas, 2014). That is, when categories being learned required explicit hypothesis testing or verbalizable rules, having verbal labels to describe the dimensions that need to be selectively attended to can improve learning.
**Latent dimensions.** The presence of latent dimensions in the signal is something that the learner might have access to, but the experimenter is not aware of. Thus, we make a lot of assumptions about what people are able to do with the dimensions and stimuli we use in our experiments. However, participants may be using dimensions in the signal that experimenters are not aware of. Thus, the presence of latent dimensions is something that the learner may be bringing to the experiment or new learning challenge as a consequence of their previous experience that experimenters are not aware of or are not accounting for. This affects our ability as researchers to make conclusions about what participants are doing. Without considering the information that may be present in the signal that can be accessed by the participant, we cannot fully understand the scope of the learning challenge.

Researchers have spent a lot of time trying to identify the underlying dimensions of natural categories. Sometimes the dimensions that can statistically separate those dimensions using techniques such as multidimensional scaling or principal components analysis are linguistically understandable and other times, they are more abstract. However, when trying to understand what humans actually are doing during perception, it is difficult.

Even perceptually simple distinctions can be distinguished by a combination of many underlying dimensions. For instance, Lisker (1986) demonstrated that there are at least 16 acoustic correlates contributing to the perception of voiced and voiceless stops in English.

Some complex distinctions can be distinguished using easily verbalizable or clear linguistic labels. For instance, using multidimensional scaling, Gandour (1979) demonstrated that Cantonese tones can be described by three dimensions: contour, direction, and height.

Mandarin syllables have three components which may interact with one another: onset (consonant), rime (vowel), and tone (Tong et al., 2008). Different researchers have used different
linguistic labels to break down the acoustic dimensions contributing to the perceived tone component. Pitch height or pitch direction may separate the tone categories, but other dimensions may exist (Gandour, 1978, 1983). Additionally, some researchers have defined the dimensions as pitch height and pitch direction (Maddox & Chandrasekaran, 2014) and others as pitch height or pitch contour (Liu & Holt, 2015). The individual definitions of how these dimensions are computed may also be different. Liu and Holt (2015) for instance defined pitch height as mean F0 and pitch contour as the F0 range with directionality. However, the individual computations of these dimensions may differ between researchers, making the true definition of these dimensions unclear.

Category contrasts like the native English difference between /r/ and /l/ can also be described in terms of—less verbalizable—dimensions of F3 onset frequency and F2 onset frequency. While F3 onset frequency gives almost completely accurate discrimination between English /r/-/l/, native Japanese productions of the same contrast could only be partially separated by F2 and F3 onset frequencies (Lotto et al., 2004).

Complex visual categories also may have this underlying issue of latent dimensions being present and contributing to perception. It is difficult to tap into what learners are perceiving, especially when the linguistic labels for that perception may not be obvious. For instance, what distinguishes between a more masculine face and a more feminine face? This is difficult to describe verbally, but there are potential underlying components of this distinction that can come from the signal (Abdi, Valentin, Edelman, & O’Toole, 1995). However, this is a distinction that is likely easy for participants to categorize, even without the overt knowledge of the features involved. Additionally, the higher-level feature of masculinity/femininity may be how people are distinguishing the categories, or they may be using some lower level dimension.
These examples demonstrate that what is present in the signal and what people are using is not always completely clear. There are many complex or abstract features that can be composed of more concrete, simple features. However, it is not a simple task to understand what people are doing when they are interacting with these stimuli. As experimenters, we make assumptions about what people are doing because it makes understanding this easier. However, our assumptions may be missing what people are doing because our own hypothesis space is wrong. This limits the conclusions that we can make about what people are doing during learning.

Finally, features themselves can be learned as a consequence of category learning (Rodrigues & Murre, 2007; Schyns & Rodet, 1997). Thus, emergent dimensions can rise up out of existing psychological dimensions or completely novel dimensions can be created as a function of what is present in the environment and how it matches previous experience.

Now that we have considered the effects of the warping of psychological dimensions on the perception of dimensions—their salience, verbalizability, and the presence of latent dimensions—we will now turn to understanding the effects on learners’ abilities to use the dimensions during learning.

**Effects on the use of dimensions**

Prior experience with dimensions in multidimensional spaces can affect how individuals are able to use those dimensions and spaces in future learning contexts. There is a common assumption in perceptual learning experiments that regardless of the category problem or dimensional space, learners will learn to use the available cues optimally according to task demands. This assumption has been made explicit in some cases (Johannesson, 2001; Nosofsky, 1986).
There is support for this assumption in the literature. Learners can direct attention to diagnostic features over irrelevant features (Baruch, Kimchi, & Goldsmith, 2012; Blair, Watson, & Meier, 2009; Blair, Watson, Walshe, & Maj, 2009; Chen, Meier, Blair, Watson, & Wood, 2013; Le Pelley, Mitchell, Beesley, George, & Wills, 2016; Yang, Little, & Hsu, 2014).

For visual category learning, after learning the categories and their features, learners can optimize their attention and focus on relevant information, spending less time looking at irrelevant information (Blair, Watson, & Meier, 2009; Chen et al., 2013). And the amount of time people look at features is correlated with attentional weights (Blair, Watson, Walshe, et al., 2009).

However, attention is not always allocated optimally to cues based on their informativeness or relevance. Learners may be using an attentional strategy that takes into account costs and benefits of focusing the attention, rather than simply attending only to fully relevant features (Matsuka & Corter, 2008). Learners do not always completely disengage from irrelevant information and still will direct attention to non-diagnostic dimensions even after some learning has occurred (Rehder & Hoffman, 2005).

There are many examples where participants do not respond statistically optimally according only to relevance, accuracy, or reward maximization (Ackermann & Landy, 2014; Ashby, Waldron, Lee, & Berkman, 2001; Blair & Homa, 2005; Francis, Kaganovich, & Driscoll-Huber, 2008; Holt & Lotto, 2006; Idemaru et al., 2012; Rehder & Hoffman, 2005; Rosas, Wagemans, Ernst, & Wichmann, 2005).

Attention given to or directed to individual dimensions can change based on prior knowledge (Kim & Rehder, 2011), type of feedback given (Little & Lewandowsky, 2009), and even something as simple as the mood of the participant (Zivot, Cohen, & Kapucu, 2013). This
can also change based on individual cost-benefit analysis of the subject or some other subjective weighting function (Ackermann & Landy, 2014; Matsuka & Corter, 2008).

And as I have argued in the first section of this paper, the amount of attention or what participants are able to perceive and do with dimensions during learning is dependent on experienced statistical regularities, task demands, and other aspects of prior experience with particular dimensions. The ability of participants to interpret and use dimensions in a way that is required by a particular learning task is fragile and is potentially affected by any of these aspects at any given time. It is absolutely worth considering how these aspects specifically impact the ability of participants to allocate attention to individual features or more broadly, use or weight dimensions during learning. Sometimes learning does not only require complete selective attention to a dimension but is a product of more complex weighting of cues in a way that reflects the relationship of the categories in dimensional space. And the view that individuals optimally weight dimensions during learning is problematic because we know of cases where learners do not optimally allocate attention to the cues that are statistically the most informative for learning. We must consider how individuals’ behavior is shaped by their past experiences and how this past experience may then contribute to the optimal or suboptimal weighting of cues in categorization, category learning, and other behavior.

I will argue that a framework that presents two kinds of psychological representations will help us understand why sometimes learners can optimally weight dimensions and other times they do not optimally weight dimensions. Versatile psychological representations allow for learners to flexibly adjust their cue weights optimally or nearly optimally based on the requirements of the task. In contrast, rigid psychological representations do not allow for cues to be optimally reweighted. The weights may be able to be adjusted during learning based on the
training or exposure, but overall, the weighting is not optimal. Moreover, I believe that this distinction between versatile and rigid psychological dimensions is not a binary one—rather, it is a continuum. Thus, some dimensions in psychological representations are more versatile or more rigid, and importantly, this is dependent on the nature of a learner’s prior experience.

I will now delve deeper into the distinction between versatile and rigid psychological representations to flesh out the distinction and under which circumstances dimensions can be classified as versatile or rigid.

**Versatile psychological representations.** As defined above, more versatile psychological representations allow for the flexible weighting of cues based on the task requirements. There are two key components in the definition of versatile psychological representations of dimensions: the optimal weighting of cues dependent on task requirements and the length of training required to change the representations.

**Optimal weighting of cues.** Versatile representations are able to be used in a more flexible way, such that the cue weights are adjusted and shifted optimally, according to the task demands.

Certain kinds of category learning (so-called unidimensional rule-based categories) require that learners selectively attend to a relevant dimension and completely disengage from or ignore an irrelevant dimension (Ashby et al., 1998; Maddox, Ashby, & Waldron, 2002; Maddox & Dodd, 2003). Learning these unidimensional rule-based categories can lead to decreased processing of irrelevant dimensions or a reduction in perceptual noise on relevant dimension (Maddox et al., 2002; Maddox & Dodd, 2003). In this case, where the task requires selective attention, learners are able to shift cue weights optimally and successfully.

This optimal weighting of dimensions is also true in many other category learning cases where participants can learn to attend to one cue and ignore another (Folstein et al., 2015;
For these versatile representations, the cues that are the most informative or the most valid are the most heavily weighted (Himmelfarb, 1970; Peterson, Hammond, & Summers, 1965). This is also seen in the weighting of multimodal information in perceptual judgements. Humans and non-human primates are able to nearly optimally weight multimodal information in decision making, judgement, and categorization tasks (Alais & Burr, 2004; Ernst, 2007; Ernst & Banks, 2002; Helbig & Ernst, 2007; Jacobs, 1999; Knill & Saunders, 2003). Even if learners have initial preferences for another cue, with versatile psychological representations, they are able to shift their attention to and between cues with training (Aliaga-Garcia, 2010; Christensen & Humes, 1996; Francis et al., 2000).

There are certainly many examples of versatile psychological representations that are consistent with these often-used assumptions that cue weighting is optimal and people are able to flexibly shift their attention between cues to be able to learn in whatever their given task is.

**Length of training and speed of changes to representations.** This optimal allocation of attention or optimal weighting of available cues to whatever is most relevant or valid or rewarded in a given context also happens rapidly. Thus, it does not require much training to be able to shift to the optimal cue and listeners can rapidly adjust to the learning environment and its requirements (Liu & Holt, 2015; Schertz, Cho, Lotto, & Warner, 2016). For instance, in category learning contexts, learners are able to shift the cue weights and learn categories within a short (~1 hour) experimental session (Ashby & Maddox, 2005, 2011).
Additionally, these rapid changes at the level of psychological representations can have profound effects on behavior. With more informative experience about the correlation between two dimensions, the ability to discriminate differences that do not follow that correlation can disappear with as little as 7.5 minutes of exposure (Stilp & Kluender, 2011, 2012; Stilp et al., 2010). The changes to the representations occur rapidly and the adjustment of the weighting of the cues or dimensions when they are versatile representations is rapid and flexible.

However, this rapid reweighting is a product of the versatile psychological representations. It is not always the case that rapid reweighting occurs or again, that it is optimal. There are many cases where learning is slower and far from optimal. Thus, we cannot discuss all instances of category learning with the assumption that the underlying psychological representations are versatile. We must consider cases where the underlying representations may be less flexible and more rigid.

**Rigid psychological representations.** As with versatile psychological representations, there are two key components in the definition of rigid psychological representations: the suboptimal nature of weighting of cues and the length of training and speed of changes to those representations. Having rigid psychological representations for dimensions does not allow for cues to be optimally weighted. Learners may be able to adjust weights somewhat based on training or exposure, but the weighting of those dimensions is often far from optimal.

**Sub-optimal weighting of cues.** Critically, unlike versatile psychological representations, when the psychological representations of dimensions are more rigid, learners will not be able to flexibly adjust attention or weighting of cues or dimensions. Instead, they will settle on some suboptimal weighting of those dimensions, even if the task itself provides informative and useful
information about what is optimal. When those underlying representations are more rigid, it is difficult to adjust them.

As with the optimal weighting of cues in the case of versatile representations, there are many examples of cases where learners find it very difficult to optimally weight cues in the case of rigid representations (Ackermann & Landy, 2014; Blair & Homa, 2005; Francis et al., 2008; Holt & Lotto, 2006; Idemaru et al., 2012; Rehder & Hoffman, 2005; Rosas et al., 2005; Schertz, Cho, Lotto, & Warner, 2015). Here, I will focus on one example of rigid psychological representations that serves as a fascinating case to understand the importance of not assuming that the underlying representations are unchanged by experience and easily manipulated or changed.

This specific case is the case of native Japanese listeners learning English speech categories of /r/ and /l/. As discussed earlier, native Japanese listeners use a suboptimal cue to distinguish between the two English categories and thus show profound difficulty in learning the categories. Native Japanese listeners use this suboptimal cue based on their previous experience with Japanese categories.

Native Japanese listeners show continuous perception along a continuum of /ra/ sounds and /la/ sounds that varies only on the dimension that is most informative for English /r/-/l/ distinction, onset of the third formant frequency, whereas native English listeners show categorical perception, meaning they show a sharp categorical distinction between /r/ and /l/ sounds (Miyawaki et al., 1975). However, the native Japanese listeners are able to discriminate different F3 onset frequencies outside of the context of /r/-/l/ (Miyawaki et al., 1975).

Performance on the perception of the English /r/-/l/ categories can improve with training (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lim & Holt, 2011; Lively, Logan, &
Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991). Native Japanese learners can increase their accuracy of identifying /r/ and /l/ categories in English. Additionally, learners are able to improve their use of the optimal, F3 onset frequency cue with training (Iverson, Hazan, & Bannister, 2005; Lim & Holt, 2011; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). However, their performance is still far from optimal. Even with training, these learners are not optimally weighting the F3 onset frequency cue, as native English listeners do. They are still relying on the suboptimal F2 onset frequency cue to distinguish the categories, even though it does not help them at all.

*Length of training and speed of changes to representations.* Unlike versatile psychological representations, for rigid psychological representations, learners must be exposed to a large amount of training to show changes and additionally, may still show extreme difficulty in optimally weighting cues.

Even though the native Japanese listeners are able to improve somewhat in their discrimination of the English /r/-/l/ categories and improve their use of the optimal cue, this training never produces perception that is equivalent to native English perception (Ingvalson, Holt, & McClelland, 2011; MacKain, Best, & Strange, 1981).

This is true even though the training that researchers have used is extensive, often using many sessions over many weeks, including as many as 45 sessions over 3-4 weeks (Bradlow et al., 1997; Lively et al., 1993, 1994; Logan et al., 1991).

Even a shorter video game training paradigm (2.5 hours of training over a five day session) improved learning as much as the extensive 2-4 week training, it still did not lead to native-like cue use or performance (Lim & Holt, 2011). These training effects are modest and require often long and many training sessions.
Compare this to the results obtained when training participants when there are versatile psychological representations—those cue weights can be shifted optimally within an hour-long experiment. Here, even a full month of training failed to shift cue weights optimally.

Additionally, unlike the versatile psychological representations, the rigid psychological representations of dimensions produce more stable cue weights. This means that often they will not change in response to experimental exposure to different statistical distributions (Idemaru et al., 2012) or in response to short-term perturbations of the distributions of the input (Schertz et al., 2015).

Even if we consider the extreme case of exposure, which is amount of time spent in an English-speaking environment, learners do have more accurate perception of English /r/-/l/ categories the longer they are in the English-speaking country (Ingvalson, McClelland, & Holt, 2011). However, there was no relationship between how long the participants had been in the English-speaking country and their reliance on the most informative cue signaling /r/-/l/ differences (F3 onset cue) (Ingvalson, McClelland, et al., 2011). Even listeners who had been in an English-speaking country for more than 10 years demonstrated suboptimal use of the F3 onset dimension. Thus, they must be improving their use of other cues that are not that single most informative cue to get better at performance overall.

Conclusion. I have demonstrated that in the case of Japanese perception of English /r/-/l/ speech categories, there is a large effect of prior experience on the psychological representations of the F3 onset and F2 onset frequency dimensions, such that they are more rigid in their representation. The underlying psychological representations have been formed and warped by native Japanese experience and thus account for this rigidity. Critically, this experience is so strong and entrenched in this case that the underlying psychological representations of the
dimensions in the F2xF3 onset frequency space remain completely rigid, even with extensive training or even years of experience.

Native Japanese listeners perception of English /r/-/l/ categories lies distinctly on the rigid end of the versatile to rigid continuum of psychological representations. However, other experienced spaces certainly lie somewhere between the two ends of this continuum. Importantly, we cannot simply assume that all learning spaces and problems are 1) unaffected by prior experience and 2) are represented in a way that is more versatile.

It is necessary to understand what about prior experience and representations causes the dimensions to lie on the more rigid or more versatile ends of the continuum. Especially if we are to understand the nature of those representations, how changes to those representations occur with training, and how learning is or is not possible in certain psychological spaces. This issue is not a unique one to the Japanese /r/-/l/ problem or to speech category learning in general. This is an issue that extends to all areas of perceptual learning and we must consider its impact on performance and representation.

The issue of native Japanese perception of English /r/-/l/ categories requires researchers reckon with the sheer difficulty of this problem and why it exists. It exists because of their experience in their native language (Kuhl, 2000; Vallabha & McClelland, 2007). Thus, native language learning can be understood as a sort of tuning of the perceptual system to the statistical distributional patterns of that language. While learning a non-native language one has existing representations that are potentially “mis-tuned to the sound structure of the to-be-acquired non-native language” (Bradlow, 2008, p. 287).

This difficulty and lack of correspondence or mistuning is something that language perception researchers have had to understand and grapple with. The nature of the native
language experience can change what is able to be learning and thus the nature of the underlying psychological representations.

It would benefit perceptual learning researchers in general to take a page out of the book of these language researchers. We must acknowledge that prior experience plays a large role in shaping and warping our representations. And we must also acknowledge that this may change the availability of dimensions and the ability of learners to use those dimensions.

Even when we think that we are sampling a novel learning environment that people don’t have entrenched ideas about, we are likely still tapping into existing representations that are marked with their own biases, experiences, and tendencies. These underlying perceptual relationships may introduce constraints on representation and processing that may then affect higher-level processing like category learning (McKinley & Nosofsky, 1996). And it is by understanding what that transformation is that we can truly understand the transformation’s effects on our perception and even more specifically on our category learning behavior.

**Conclusion**

The way that we represent perceptual information is dependent on the natural constraints of our perceptual systems as well as the regularities present in the sensory environment. These regularities can be exploited by the perceptual system to represent information more efficiently. These perceptual and experienced-based constraints on representation affect learning. In new learning environments that better match our prior experience and representations, the information in the signal can be used more flexibly according to task demands. However, if our representations--derived from experience--are in conflict with new learning environments, the underlying psychological dimensions may behave more rigidly, resisting task-dependent reweighting that is necessary for many learning challenges.
Figure 2. Dissertation overview.

Chapter Overview

The goal of this dissertation is to better elucidate the above mechanisms. Figure 2 shows a visual overview of this dissertation and the level at which each chapter falls. Chapters 2 and 3 investigate how factors of the learning environment and feedback structure can impact overall category learning. Chapter 2 demonstrates that aspects of both the perceptual environment and task demands influence category learning and subsequent learned representations. Chapter 3 investigates how experiencing category variability within a single trial impacts learning and representations. Chapter 4 investigates the link between perceptual representations and novel category learning by examining how existing perceptual representations influence how learners form categories in the same perceptual space. Chapter 5 explores the connection between the sensory environment and perceptual representation by investigating how similarity-based representations are altered through passive exposure to highly structured acoustic information.
Chapter 6 brings together sensory experience and novel category learning by examining how passive exposure to highly structured perceptual information affects subsequent category learning behavior, as assessed through accuracy across learning and strategies via decision-bound models. Chapter 7 presents a neural network model that connects these three levels and serves as a starting point in understanding the underlying computational mechanisms that allow sensory experience to shape perceptual representations, which then influence higher-level cognition, such as the processes involved in category learning. Finally, in Chapter 8, I discuss potential implications for these findings and necessary future directions for research.
References


http://doi.org/10.1080/13506285.2012.726450


http://doi.org/10.1016/j.cognition.2009.04.008


Training+non-native+language+sound+patterns:+Lessons+from+training+Japanese+adults+on+the+English+/r/+-/l/+contrast&ots=PK3PqH-YD_&sig=R0T3AnQp2b3vkxO-4BNLJ5B3SXY


http://doi.org/10.3758/BF03211210


http://doi.org/10.1016/j.cognition.2013.06.007


http://doi.org/10.1017/S0140525X15000965


http://doi.org/10.1111/1467-9280.00392


http://doi.org/10.1121/1.2945161


http://doi.org/10.1146/annurev.psych.49.1.585


http://doi.org/10.3758/s13423-016-1209-0


http://doi.org/10.1037/0096-1523.19.5.1105


http://doi.org/10.1037/0096-1523.15.2.212


http://doi.org/10.3758/BF03210870


http://doi.org/10.3758/BF03207084


http://doi.org/10.1162/jocn_a_00243


Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech
http://doi.org/10.1044/jshr.3902.278

http://doi.org/10.1121/1.407649

http://doi.org/10.1121/1.418207

http://doi.org/10.1523/JNEUROSCI.2978-14.2015


Rost, G. C., & McMurray, B. (2010). Finding the Signal by Adding Noise: The Role of
Noncontrastive Phonetic Variability in Early Word Learning. *Infancy, 15*(6), 608–635. 
http://doi.org/10.1111/j.1532-7078.2010.00033.x


http://doi.org/10.1016/S0010-0277(98)00075-4

http://doi.org/10.1037/a0021610


http://doi.org/10.1017/S0140525X98000107


http://doi.org/10.1371/journal.pone.0030845


Chapter 2: Task and sampling distribution affect auditory category learning
(The following text has been adapted from Roark, C. L., & Holt, L. L. (2018). Task and
distribution sampling affect auditory category learning. Attention, Perception & Psychophysics,

Everyday decisions depend on well-learned category representations, whereby
perceptually discriminable experiences are treated as functionally equivalent. For example, one
must be able to categorize an animal encountered on the street as “friendly” or “dangerous” to
decide whether to approach or avoid it. Speech perception can be considered an example of
categorization (Holt & Lotto, 2010) in the sense that perceptually discriminable and acoustically
variable utterances come to be mapped to phonetic categories. Speech presents a challenging
case of auditory perceptual category learning because phonetic categories are defined by multiple
acoustic dimensions that may not be perceptually separable or easily verbalized, the distributions
of which are highly overlapping (Hillenbrand, Getty, Clark, & Wheeler, 1995; Holt & Lotto,
2008, 2010; Jongman, Wayland, & Wong, 2000; Lisker, 1986; Vallabha, McClelland, Pons,
Werker, & Amano, 2007).

An influential cognitive neuroscience framework of category learning developed in the
visual category learning literature (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) has
recently been applied to auditory and speech category learning (Chandrasekaran, Koslov, &
Maddox, 2014; Chandrasekaran, Yi, & Maddox, 2014; Maddox, Chandrasekaran, Smayda, & Yi,
2013; Yi, Maddox, Mumford, & Chandrasekaran, 2014). The COMpetition of Verbal and
Implicit Systems (COVIS) “dual systems” model of category learning posits two distinct
learning systems mediated by the striatum: an explicit system that involves the frontal cortex and
the head of the caudate nucleus and an implicit system that recruits the putamen and the tail of
the caudate. These dual systems are differentially engaged by distinct distributions of category
exemplars (Ashby & Maddox, 2011). Rule-based category distributions—which are thought to engage an explicit, reflective, hypothesis-testing system that relies upon working memory and attention—can be distinguished by a single, simple verbalizable rule (Ashby et al., 1998). Conversely, information-integration category distributions—proposed to engage an implicit, reflexive system that uses procedural learning—can only be distinguished if information from multiple dimensions is integrated at a pre-decisional stage (Ashby & Gott, 1988). Because this integration is pre-decisional, the relationship between information-integration categories is often non-verbalizable. In the dual systems theory, the distribution structure of category exemplars is thought to be the primary determinant of which of the two category-learning systems drives the motor response.

Research on auditory category learning in the context of the dual systems model has focused mostly on non-native speech category learning of Mandarin lexical tones via overt training (Chandrasekaran, Koslov, et al., 2014; Maddox & Chandrasekaran, 2014; Yi et al., 2014). In overt training, participants are aware that they are performing a categorization task, make explicit categorization decisions, and are given explicit feedback about these decisions after each trial. Additionally, with overt training participants are sometimes explicitly informed of the dimensions on which the stimuli vary. Under these conditions, Mandarin tone speech categories are learned best when participants used a reflexive strategy that relies upon the implicit system that learns information-integration categories (Chandrasekaran, Koslov, et al., 2014; Maddox & Chandrasekaran, 2014; Yi et al., 2014). The reasoning is that speech categories—like information-integration categories—are defined by highly variable exemplars signaled by multiple acoustic dimensions in a manner that is difficult to verbalize (Chandrasekaran, Koslov, et al., 2014) and thus are learned better via the implicit, reflexive
learning system. In support of this, functional neuroimaging reveals that the patterns of corticostriatal activation during speech category learning are more consistent with the involvement of implicit, reflexive system posited by COVIS (Yi et al., 2014).

These recent results suggest the promise of dual systems theory in understanding auditory, and in particular, speech, category learning. However, the categorization challenges presented by auditory (and speech) signals are somewhat different from those invited by the visual categories that have served as the principle testing ground for dual systems theory (Holt & Lotto, 2010). There remain important open questions about how well auditory category learning aligns with the predictions of dual systems theory. In the present research, we examine the predictions of dual systems theory in the context of how manipulations of task, distribution sampling, and the category type affect auditory category learning.

We test these questions using novel, artificial nonspeech auditory categories. Although nonspeech categories have not been used as frequently in the context of examining dual systems theory as non-native speech categories (but see Chandrasekaran, Koslov, et al., 2014), they provide us with the control to precisely define and manipulate category distributions, distribution sampling, and the course of learning as a function of different learning tasks. Thus, in the same way that very simple visual dimensions have been used productively to understand the learning systems available to categorizing more complex objects in the natural world, we employ nonspeech sounds to understand the processes available to support learning more complex speech categories. We next describe the rationale for focusing on task, category distributions, and distribution sampling in the present research.

**Task**

Nearly all studies of visual or auditory category learning from a dual systems perspective
have used an overt training task (Ashby, Maddox, & Bohil, 2002; Chandrasekaran, Yi, et al., 2014; Dunn, Newell, & Kalish, 2012; Ell, Ing, & Maddox, 2009; Maddox, Filoteo, Hejl, & Ing, 2004; Maddox, Love, Glass, & Filoteo, 2008; c.f. for a discussion about unsupervised learning see Ashby, Queller, & Berretty, 1999). Participants are told how many categories exist, they are instructed that the goal is to categorize the stimuli, and they are provided with corrective trial-by-trial feedback on category decisions. In an exception that used an unsupervised category training paradigm without overt feedback, Ashby et al. (1999) found that rule-based (RB) categories distinguished by a single, simple verbalizable rule could be learned without feedback, but that information integration (II) categories requiring information from multiple dimensions to be integrated at a pre-decisional stage could not. The researchers concluded that learning II categories is critically dependent on feedback, whereas learning RB categories can occur without feedback. Thus, due to the inability of participants to learn II categories without feedback, the majority of research from the dual systems literature has utilized supervised learning tasks with overt feedback.

Yet, recent research in auditory category learning suggests an alternative approach that is neither wholly explicit nor unsupervised. *Incidental learning* occurs without instructions to categorize, overt category decisions, or explicit feedback. Instead, *sound categories are learned incidentally by virtue of their relationship to success in performing a primary task* distinct from auditory category learning (Gabay, Dick, Zevin, & Holt, 2015; Lim & Holt, 2011; Lim, Lacerda, & Holt, 2015; Liu & Holt, 2011; Vlahou, Protopapas, & Seitz, 2012; Wade & Holt, 2005). For example, when auditory categories’ exemplars are presented in a manner that correlates with where a visual ‘x’ will next appear on the screen in a visual detection task, participants incidentally learn complex auditory categories, including speech, in the course of performing the
visual detection task (Gabay et al., 2015; Liu, 2014). Incidental auditory category learning is also apparent across more challenging primary tasks, such as navigating a videogame environment in which sound categories are correlated with aspects of the input that support success in the primary, game navigation, task (Gabay et al., 2015; Lim & Holt, 2011; Lim et al., 2015; Liu & Holt, 2011; Wade & Holt, 2005). Inasmuch as this incidental category learning proceeds even without instructions about the importance of the sounds, knowledge of the existence of auditory categories, overt category decisions, or explicit feedback about categorization, it may better model aspects of category learning in natural environments whereby correlated objects, events and actions across modalities are available as structure that may guide learning (Gabay et al., 2015; Wade & Holt, 2005).

Although prior studies of category learning in a dual systems theory framework have relied nearly exclusively on overt training, there is extensive evidence to demonstrate the importance of task variables on category learning. Investigations of the distinction between II and RB category learning have emphasized the significance of feedback timing (Dunn et al., 2012; Ell et al., 2009; Maddox, Ashby, & Bohil, 2003; Maddox, Ashby, Ing, & Pickering, 2004; Maddox & Ing, 2005; Smith et al., 2014; Worthy, Markman, & Maddox, 2013), amount of feedback and feedback type (Ashby et al., 2002, 1999; Ashby & O’Brien, 2007; Dunn et al., 2012; Goudbeek, Cutler, & Smits, 2008; Goudbeek, Swingley, & Smits, 2009; Maddox et al., 2008), and changing instructions to participants (Chandrasekaran, Yi, Smayda, & Maddox, 2016; Grimm & Maddox, 2013) However, the question of whether category learning for II and RB categories differs across overt training compared to incidental training has yet to been investigated.

We hypothesize that II categories—which are difficult to verbalize and require pre-
decisional integration—may benefit more from incidental training tasks in which attention is directed toward a primary task, and away from decisions about category exemplars. In contrast, consistent with the dual systems theory, RB categories may benefit more from an overt training task in which attention can be directed toward the stimuli and features that distinguish categories. The overt task may encourage more explicit, verbalizable hypothesis testing to support learning RB categories. Since this kind of explicit strategizing can be detrimental for II category learning (Grimm & Maddox, 2013), incidental training may be beneficial for learning II categories.

Category Distribution Sampling

The categorization challenges presented by speech—and, indeed, most natural categories—almost always involve complex, probabilistic category exemplar distributions that overlap in acoustic space (Kuhl et al., 1997; Lotto, Sato, & Diehl, 2004; McMurray & Jongman, 2011; Peterson & Barney, 1952). Yet, many studies of speech and nonspeech auditory category learning have examined learning across non-overlapping, deterministic distributions well differentiated in acoustic space and characterized by a small number of exemplars experienced repeatedly across training (Holt & Lotto, 2006; Kluender, Lotto, Holt, & Bloedel, 1998; Kuhl, 1991; Lim & Holt, 2011; Mirman, Holt, & McClelland, 2004; Wade & Holt, 2005). Even when more probabilistic distributions of natural speech productions have been used to study category learning among non-native listeners (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Yi et al., 2014), the impact of distribution sampling on learning has not been a focus of investigation.

The approaches using probabilistic and deterministic distributions differ on several dimensions. However, each approach is meant to approximate some sampling that is similar to real-world categories, such as speech. Deterministic distributions are highly stylized, the
categories do not overlap, and there are relatively few exemplars. Probabilistic distributions are randomly sampled, the categories are often overlapping, and there are many possible exemplars. It is important to understand what, if any, effect sampling from these different kinds of distributions has on category learning. To the extent that each effectively approximates sampling from naturalistic categories, then approach to distribution sampling should not have an impact on learning. However, it is entirely possible that sparser, non-overlapping, more stylistically sampled distributions may be learned in a different manner than denser, overlapping, randomly sampled distributions.

This issue has not been investigated thoroughly even with the artificially constructed visual categories upon which the dual systems model of categorization is based. To our knowledge only one study has addressed something similar to the issue of probabilistic versus deterministic category distribution sampling in the visual domain. Ell and Ashby (2006) examined the impact of category overlap on learning. The degree to which exemplars from different categories were drawn from overlapping versus entirely distinct regions of stimulus space impacted learning of visual II categories, but not visual RB categories. Specifically, when categories’ exemplars were moderately overlapping across II distributions, participants were able to use optimal II strategies in category learning; however, with too much or too little overlap of II distributions, participants relied on suboptimal, RB strategies. This indicates that at least some aspects of the sampling distribution, specifically overlap, may influence the course of learning. Thus, it is important to examine the potential learning differences between carefully sampled deterministic distributions and randomly sampled probabilistic distributions, especially in light of the fact that the auditory category learning literature has employed them somewhat interchangeably.
In the present study, we manipulate whether the II and RB category distributions are sampled probabilistically or deterministically in acoustic space. Examining the interaction of category distribution type (II versus RB) and task (incidental versus overt) and distribution sampling (deterministic, non-overlapping versus probabilistic, overlapping) is important given the ubiquity of probabilistic, overlapping category distributions in speech and other natural categories, including visual categories (e.g. Nosofsky, Sanders, Meagher, & Douglas, 2017).

In the present study, we investigate learning across highly stylized, deterministic distributions of sound category exemplars like those that have characterized most auditory category learning studies to date. We also examine learning across categories defined more probabilistically. There is not a large literature to support strong predictions about the effect of these different sampling distributions on learning. Drawing from Ell and Ashby’s (2006) results, one might predict that category overlap in deterministic compared to probabilistic sampling will impact learning of II categories, but not RB categories. Participants learning categories sampled deterministically may be biased toward explicit strategies thereby impeding II learning and benefitting RB learning. To the extent that explicit strategies influence the learning, there may also be an interaction between task type (incidental, overt) and distribution sampling. Compared to probabilistic distributions learned through incidental training, learning via overt training may be better across deterministic, non-overlapping category distributions that are easy to learn with verbalizable rules and for which the feedback is perfectly consistent with optimal strategies.

Summary

In the present experiment, we examined the impact of task and category distribution sampling on learning four auditory categories defined by either II or RB stimulus distributions. We trained separate groups of participants with either a traditional, overt categorization
paradigm with explicit feedback on every trial or with an incidental paradigm in which neither categorization decisions nor feedback were explicit task demands. Finally, we varied the nature of distribution sampling to examine the influence of the probabilistic or deterministic nature of the category distributions on learning. We directly examined the influence of category type, task, and distribution sampling on auditory category learning using within-training metrics, as well as a common overt labeling task administered post-training to assess generalization of learning to novel category exemplars.

**Methods**

**Participants**

One hundred and sixty-six adults ages 18-25 (89 females, 77 males) affiliated with Carnegie Mellon University participated for partial course credit or a small payment ($10). All participants had normal or corrected-to-normal vision and reported normal hearing. There was a total of eight conditions that varied by training task, category distribution, and distribution sampling (Table 1). Participants were trained on either an incidental task or an overt task, learned either to categorize rule-based or information-integration category distributions, and the distributions were either probabilistic or deterministic in their sampling. An additional five participants were run but excluded from all analyses because of equipment failure.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Participants in Each Condition</strong></td>
</tr>
<tr>
<td><strong>Deterministic Sampling</strong></td>
</tr>
<tr>
<td>Task</td>
</tr>
<tr>
<td>Incidental task</td>
</tr>
<tr>
<td>Overt task</td>
</tr>
<tr>
<td><strong>Probabilistic Sampling</strong></td>
</tr>
<tr>
<td>Task</td>
</tr>
<tr>
<td>Incidental task</td>
</tr>
<tr>
<td>Overt task</td>
</tr>
</tbody>
</table>
Stimuli

The learning challenge differed across four conditions defined by the category input distributions. Each condition had four categories separated by optimal decision boundaries, as shown in Figure 1. Participants trained on either information-integration (II) distributions (Figure 1b, 1d) distributions or rule-based (RB) distributions (Figure 1a, 1c) that were sampled deterministically in acoustic space (Figure 1a, 1b) or probabilistically (Figure 1c, 1d). For the deterministic distributions, it is possible to define optimal decision boundaries that classify exemplars with perfect accuracy because there is no overlap between the categories. However, no decision boundary results in 100% accuracy for the probabilistic distributions because exemplars can probabilistically belong to more than one category due to category overlap. Optimal performance for the probabilistic RB and II conditions was 90.5% and 90.25%, respectively. We used these moderate levels of overlap to reflect similar levels of overlap in Ell and Ashby (2006), who found that moderate levels of overlap did not hinder learning for either II or RB categories.
The two-dimensional acoustic space from which stimulus exemplars were sampled was defined by two dimensions: center, or carrier, frequency (CF) and modulation frequency (MF). CF can be described approximately as the pitch of the tone and MF as the warble of the tone. We chose these particular acoustic dimensions because they were used in an existing auditory category learning study (Holt & Lotto, 2006), demonstrating that listeners are able to learn categories defined across these acoustic dimensions. Additionally in this previous study, these dimensions were psychoacoustically matched for discriminability across the approximate range.
used in the present study (Holt & Lotto, 2006). Further, manipulation of sounds across these dimensions creates highly artificial exemplars that participants are unlikely to have heard previously. In the same manner that Gabor patches provide a simple stimulus to manipulate parametrically in the visual domain, these simple acoustic stimuli provide us with a testbed for auditory category learning. Each 300-ms stimulus was defined by a CF modulated with a depth of 100 Hz at the corresponding MF, with overall energy RMS matched across exemplars and all synthesis accomplished using MATLAB R2014a (The MathWorks, Inc., Natick, Massachusetts).

The stimulus distributions for the Deterministic, RB categories were adapted from Holt and Lotto (2006). The stimulus distributions for the Deterministic II categories were generated by rotating the RB stimuli counterclockwise in acoustic space by 45 degrees. The Deterministic categories were highly stylized, in the manner of previous auditory studies (Holt & Lotto, 2006; Kluender et al., 1998; Kuhl, 1991). The Deterministic categories also did not overlap and had a relatively small number of exemplars per category. For the Deterministic categories (Figure 1a, 1b), there were 48 exemplars per category plus the centroid of each category. Half of the exemplars were used during the training phase (24 exemplars/category, 96 total stimuli) and half of the exemplars plus the centroid were reserved for the generalization test phase (100 stimuli).

To create the Probabilistic stimulus distributions, we defined the underlying distributions to have the same means in CFxMF acoustic space as each of the Deterministic category distributions. We increased the number of exemplars and the variance of the category distributions to manipulate sampling. We then sampled randomly from the defined distribution; the random sampling resulted in means, variance, and covariance that varied somewhat from those defining in the underlying distribution (Appendix A). The probabilistic distributions were created with MATLAB R2012. The Probabilistic distributions (Figure 1c, 1d) had 100 exemplars
per category. Half of the exemplars were used during the training phase (50 stimuli/category, 200 total stimuli) and half of the exemplars were reserved for the generalization test phase (200 stimuli). Unlike the Deterministic conditions, not all of the exemplars from the Probabilistic distributions were played in the test phase. Rather, they were randomly sampled for each participant for a total of 100 trials.

**Task**

Participants were trained on one of two training tasks: incidental or overt (see Figure 2). After training, all participants were tested on an overt generalization post-test, which included exemplars not experienced during training. By adding an overt generalization post-test, we were able to compare learning between the Incidental and Overt training tasks and to better understand learning that extends beyond individual exemplars experienced in training, to include generalization to novel exemplars consistent with the category distribution.
Training Task: Incidental. Research in our laboratory has demonstrated the effectiveness of a simple incidental learning task, the SMART task, in training listeners to categorize sounds (Gabay et al., 2015; Liu, 2014; Liu & Holt, n.d.). This paradigm was adapted as a highly simplified version of a videogame training paradigm successful in training speech and nonspeech categories incidentally, without overt training or feedback (e.g., Lim & Holt, 2011; Wade & Holt, 2005). In the SMART task, the primary objective is to rapidly detect the appearance of a visual target in one of four possible screen locations by pressing a key corresponding to the screen location.

Within-trial category-exemplar variability. On each trial, five unique exemplars drawn
from one of the four sound categories preceded the appearance of the visual target. For the Deterministic categories, the sound categories perfectly predict the location of the upcoming visual target and, consequently, the action required to complete the visual detection task. The overlap of the Probabilistic categories makes it such that the exemplars played on a single trial may not be equally representative exemplars of the category based on the optimal boundary between the categories. However, the exemplars played on each trial were always drawn from a single category, so the sound category was predictive of the visual target location. Prior research demonstrates that participants learn auditory categories incidentally in the SMART task and generalize learning to labeling novel exemplars in an overt, post-training labeling task (Gabay et al., 2015). The generalization to novel sound category exemplars underscores the important point that this learning is not a simple sound-to-location mapping. The inherent variability in sound category exemplars encourages participants to learn sound categories that robustly generalize to guide subsequent responses to novel, unfamiliar exemplars (Gabay et al., 2015).

In the present experiment, we used the SMART task to assess incidental auditory category learning and generalization across deterministic and probabilistic II and RB category distributions with a covert reaction-time measure during online learning and an overt, post-training category labeling task.

**Covert reaction time measure of learning.** Participants were instructed that their task was to indicate where a visual target (a red X) appeared on the screen by responding with the corresponding keyboard button (responses were U, I, O, and P buttons; see Figure 2). They were told that sounds would precede the X, but no mention was made about the relationship between the sounds and the location of the X.

Participants experienced three Training Blocks (96 trials/block) in which each of the four
sound categories predicted the appearance of the visual target in one specific location.

Subsequently, they completed a brief Random Test Block (48 trials), in which the mapping of sound categories to visual target locations was fully random; any sound category exemplar could precede the presentation of the visual target in any position with equal probability. Following the approach of Gabay et al. (2015), this block was shorter to avoid extensive exposure to a random mapping that may erode category learning across the Training Blocks. The final block was one last Training Block. This block was included to reinstate category learning prior to the post-training Generalization Post-Test.

On each trial, the sound category was chosen pseudo-randomly (random shuffle of a fixed number of trials for each category per block). Then five exemplars were randomly selected from the pool of category exemplars. The 300-ms exemplars were each presented once, with a 50 ms inter-stimulus silent interval. The final exemplar was followed by a 500 ms silence, after which a red X appeared in the location is associated with the sound category. The trial structure was identical for the Random Test Block, except that the red X appeared in a randomly-selected location as opposed to the location associated with the sound category in training blocks. Participants responded to indicate the location of the red X by pressing the associated button as quickly as possible. Reaction times were measured as the time lapsed from the onset of the visual detection target to the press of the response key. After each experimental block, participants were encouraged to rest briefly.

The Random Test Block provides for a covert measure of incidental category learning to be collected online during training. If participants have incidentally learned sound categories in service of guiding visual detection behavior, then eliminating the consistent relationship between category and location in the Random Test Block should slow response to the visual target. As
evident in prior studies (Gabay et al., 2015; Liu, 2014), learning should be apparent in a reaction time (RT) Cost ($RT_{block\ 4} - RT_{block\ 3}$) between Block 4 (random) and Block 3 (consistent). This RT Cost serves as our covert reaction time measure of learning across incidental training conditions.

**Overt measures of learning and generalization.** An overt labeling task immediately followed the SMART task. Before the start of the Generalization Post-Test, participants were informed that, the location of the X had been associated with the sounds that preceded it in the SMART task and that they should respond with a key press to guess the location the visual target would be most likely to appear. On each of 100 trials (25 trials/category), five novel sound category exemplars not experienced in the SMART task were randomly selected from the pool of generalization stimuli and presented with the same timing characteristics as in training. As in training, the Generalization Post-Test had within-trial variability. But, no visual targets appear in this task, thereby providing no feedback. This overt labeling task provides an explicit test of category learning and its generalization to novel exemplars not experienced in training.

**Training Task: Overt.** The overt task modeled the training approach taken in most studies of category learning (Ashby et al., 2002; Yi et al., 2014), while aligning closely with many of the task details of the incidental SMART task (Gabay et al., 2015). In the Overt task, participants experienced the same kind of multi-modal location-to-sound category mapping as the Incidental (SMART) task participants, but were explicitly informed of the relationship between sounds and visuomotor targets, made explicit categorization decisions, and were given overt corrective feedback following each categorization decision.

As in the Incidental training task, participants first experienced three Training Blocks (96 trials/block) in which sound categories predicted the position of the visual target. However,
Unlike the Incidental task, participants did not receive a Random Test Block because such a block played no role in covert assessment of learning in this overt feedback version of the paradigm. In sum, the Overt task involved four Training Blocks followed by the Generalization Post-Test.

For each trial, the sound category was first chosen pseudo-randomly (random shuffle of a fixed number of trials for each category per block). As with the Incidental task, there was within-trial variability in the Overt task. Five category exemplars were then randomly selected from the pool of exemplars for that category. The 300-ms exemplars were sequentially presented at the onset of the trial, with 50-ms silent intervals. Participants pressed a button (U, I, O, or P) to indicate which visual location they believed to be associated with the sound. A 500-ms silence followed the response, after which a red X appeared in the visual location associated with the sound category presented on that trial as feedback about category identity.

After each block, participants were encouraged to rest briefly. Button presses were considered correct if they corresponded to the correct visual location mapped to the trial’s sound category, providing a measure of accuracy across blocks.

Overt generalization post-test. Immediately after the last block of training, participants engaged in an overt categorization test with within-trial variability that was identical to the Generalization Post-Test described for Incidental training.

Results

We describe the results separately for the Incidental and Overt training conditions because some measures were task-specific. The Incidental training task provided covert online measures of learning via reaction time (Figure 3), whereas the Overt task did not. For the Overt task, the relevant behavior is accuracy across training blocks (Figure 4). Across Incidental and Overt
conditions, there was a common post-training overt labeling task to assess generalization of learning (Figure 5).

**Training Task: Incidental**

**Reaction time filtering.** We filtered the reaction times to include only trials on which participants were accurate in responding to the X on the screen and for which reaction times were less than 1500 ms or greater than 100 ms. A total of 3.77% of trials were excluded across all conditions (2.74% of trials were excluded for II-Probabilistic, 3.74% for RB-Probabilistic, 4.46% for II-Deterministic, 4.11% for RB-Deterministic).

**Covert reaction time measure of learning.** The covert measure of learning, Reaction Time Cost (RT Cost), provided an online measure of incidental category learning. We predicted that eliminating the consistent relationship between the sound category and the upcoming location of the visual target established across Blocks 1-3 would slow reaction times to detect the visual target in Block 4, compared to Block 3, as expressed by a positive RT Cost ($RT_{Block4} - RT_{Block3}$). Since exemplars vary on a trial-by-trial basis, this is indicative of sound category learning.

Figure 3 shows the average reaction times for each condition. Following the approach of prior research (Gabay et al., 2015), we first examined the RT Cost by conducting paired-samples t-tests comparing Block 4 and Block 3 reaction times for each condition. There were significant RT Costs, indicative of incidental auditory category learning, for both Deterministic II ($M=33.7$ ms, $t(20) = 3.39, p = .003$, Cohen’s $d = .56$) and RB ($M=22.3$ms, $t(19) = 2.79, p = .012$, Cohen’s $d = .38$) distributions. For the Probabilistic distributions, only the RB condition resulted in a significant RT Cost indicative of incidental auditory category learning ($M=31.2$ ms, $t(20) = 3.39, p = .003$, Cohen’s $d = .45$). The RT Cost for the Probabilistic II condition was not significant
(\(M=12.3\) ms, \(t(19) = 1.10, p = .29, \text{Cohen’s } d = .10\)). Thus, with the exception of the Probabilistic II condition, each group exhibited significant incidental auditory category learning, as indexed by the covert, online RT Cost measure.

We next asked whether the magnitude of the RT Cost varied as a function of the learning challenges presented by the different conditions. It did not. A 2 (Category Distribution) x 2 (Distribution Sampling) between-subjects ANOVA revealed that the magnitude of the RT Cost did not depend on either Category Distribution (II vs. RB; \(F(1, 78) = 0.15, p = .70, \eta_p^2 = .002\)) or Sampling (Deterministic vs. Probabilistic; \(F(1, 78) = 0.42, p = .52, \eta_p^2 = .005\)), and there was no interaction \((F(1, 78) = 2.48, p = .12, \eta_p^2 = .031)\). Although all conditions except for Probabilistic II demonstrated evidence of incidental category learning by the RT Cost measure, the magnitude of the RT Cost was not dependent on either Category Distribution or Sampling.

**Average Reaction Time.** We also compared average RTs across all training blocks as a function of condition. Participants learning Deterministic category distributions were marginally faster \((M=381\) ms, \(SE=11.4\)) to respond to the visual targets than participants in learning Probabilistic category distributions \((M=413\) ms, \(SE=11.4\), \(F(1, 78) = 4.06, p = .047, \eta_p^2 = .049\)). It appears that the more highly overlapping Probabilistic category distributions slowed visual target detection somewhat relative to simpler, more coherent Deterministic category distributions. One possibility is that participants may be sensitive to the deterministic versus probabilistic structure of the category input distributions. However, group differences cannot be ruled out in this between-subjects design. There was no effect of Category Distribution (II, RB) on the average RT \((F(1, 78) = 0.56, p = .46, \eta_p^2 = .007)\) and only a marginal interaction between Category Distribution and Sampling \((F(1, 78) = 3.63, p = .060, \eta_p^2 = .045)\).
Figure 3. Average reaction time during the Incidental training task. Ribbon error bars represent the standard error of the mean. Individual points represent individual participant averages. Participants in the II condition are shown as blue circles and participants in the RB condition are shown as red squares.

We note that we did not have any *a priori* predictions that the conditions would differ by average reaction time. Thus, examining average reaction time serves as a manipulation check to make sure that the different conditions did not differ on RT. However, there is one subject in the Probabilistic II condition who was consistently slower than the other participants. We also ran the analyses excluding this subject, who was more than three standard deviations above the mean on four out of five blocks. No other subject was more than three standard deviations above the mean on any individual block. Examining the reaction time data without this subject had largely the same results, except for on the average RT. Excluding the outlier from the Probabilistic II
condition, the effect of Sampling distribution on average RT disappears, such that there were no significant differences between the Probabilistic and Deterministic average RT (Probabilistic $M=404$ ms, $SE=9.6$; Deterministic $M=381$ ms, $SE=11.4$; $F(1, 77) = 2.94, p = .090, \eta_p^2 = .037$).

Similar to the results including the outlier, the effect of Category Distribution and interaction were not significant after excluding the outlier from the Probabilistic II condition (Category Distribution: $F(1, 77) = .042, p = .84, \eta_p^2 = .001$; Interaction: $F(1,77) = 2.52, p = .12, \eta_p^2 = .032$).

After excluding the outlier in the Probabilistic II condition, there were no differences between the groups on average RT, indicating that any differences in learning are not tied to differences in reaction time.

**Training Task: Overt**

Normalization. Recall that an optimal observer would achieve 100% accuracy in the Deterministic conditions, but only 90.25% or 90.5% accuracy in the Probabilistic conditions. To account for this difference, we first computed normalized accuracy values for data from the Probabilistic learning conditions as (Normalized Accuracy = Raw Accuracy / Optimal Accuracy) with optimal accuracy (.9025 and .905, for Probabilistic II and RB, respectively). All comparisons were conducted with these normalized accuracy values. We note that none of the qualitative patterns of results changed as a result of normalization; it simply provides for equitable cross-condition comparison.

Accuracy across blocks. In the Overt training task, (normalized) accuracy across blocks is the principle measure of category learning; Figure 4 plots these results. Examining performance across blocks with a 2 x 2 x 4 repeated-measures ANOVA [Category Distribution (II, RB) x Sampling (Deterministic, Probabilistic) x Block], we found that independent of Category Distribution or Sampling, participants generally improved with training across Blocks ($F(2.6,$
Participants in the Deterministic condition learned more across blocks than participants in the Probabilistic condition ($F(2.6, 209.2) = 3.25, p = .029, \eta^2_p = .039$). Performance across blocks was not impacted by the interaction of Category Distribution (II, RB) and Sampling ($F(2.6, 209.2) = 1.40, p = .25, \eta^2_p = .017$) and there was no advantage in learning II versus RB Category Distributions across training, ($F(2.6, 209.2) = 0.54, p = .63, \eta^2_p = .007$). Generally, this may indicate that Sampling – whether categories are sampled Deterministically, or Probabilistically – is the main driver of the difference in improvement across conditions, rather than Category Distribution or the interaction between Category Distribution and Sampling. Deterministic auditory categories were more readily learned than probabilistic auditory categories. The overall pattern of learning was differentiated by the distribution Sampling, not Category Distribution.

---

1 Huynh-Fedlt corrected because Mauchly’s test of sphericity was significant, $p < .001$
Figure 4. Block-by-block performance during the Overt training task. The dotted line denotes chance performance (25%). Ribbon error bars represent the standard error of the mean. Individual points represent individual participant averages. Performance for participants in the II condition is shown in blue circles and performance for the RB condition is shown in red squares.

In examining overall accuracy, rather than performance across blocks, we found that participants learning RB categories had higher average accuracy than participants learning II categories ($F(1, 80) = 11.58, p = .001, \eta^2_p = .13$). Deterministic category input distributions were more easily learned than Probabilistic category input distributions ($F(1, 80) = 21.87, p < .001, \eta^2_p = .22$). There was no interaction between Category Distribution and Sampling ($F(1, 80) = 1.12, p = .30, \eta^2_p = .014$), indicating that RB conditions were learned better than II conditions for
both Deterministic and Probabilistic category distributions. Moreover, Deterministic
distributions were better learned than Probabilistic distributions for both II and RB categories.
These results support the prediction that overt training should benefit RB category learning. We
also predicted that RB might not be affected by sampling distribution. In contrast to our
predictions, the sampling distribution affected learning for both II and RB categories such that
Deterministic sampling led to better overall category learning than Probabilistic sampling.

We also note that learning was quite rapid. Significant learning was evident in the first 96-
trial block of training in each of the conditions (chance = 25%; Deterministic II: $t(20) = 13.15, p
< .001, M=56.4\%$, Cohen’s $d = 5.88$, Deterministic RB: $t(20) = 12.62, p < .001, M=64.7\%$,
Cohen’s $d = 5.64$, Probabilistic II: $t(20) = 12.41, p < .001, M=51.0\%$, Cohen’s $d = 4.97$,
Probabilistic RB: $t(20) = 13.57, p < .001, M=64.2\%$, Cohen’s $d = 5.66$).

**Generalization Post-Test: Incidental and Overt Training**

Both Incidental and Overt training results in auditory category learning, as assessed by the
task-specific measures during training. Here, we examine the post-training measure of
generalization of category learning assessed using a common task across the groups trained
incidentally and overtly. Performance during the Generalization Post-Test is shown in Figure 5.
Figure 5. Generalization test performance for all conditions. The dotted line denotes chance performance (25%). Error bars represent the standard error of the mean. Individual points represent individual participant average.

**Normalization.** As for the Overt training task, we normalized generalization test accuracies. We computed normalized accuracy values for data from the Probabilistic learning conditions as (Normalized Accuracy = Raw Accuracy / Optimal Accuracy) with optimal accuracy (.9025 and .905, for Probabilistic II and RB, respectively). We did not compute normalized accuracies for the Deterministic conditions because the optimal accuracy was 100%. We note that none of the qualitative patterns of results changed as a result of normalization; it simply provides for a fair cross-condition comparison.
To compare learning in all conditions, we ran a 2 x 2 x 2 ANOVA of Training Task (Incidental vs. Overt) x Category Distribution (II vs. RB) x Sampling (Deterministic vs. Probabilistic). This allowed us to determine the aspects of the training task and/or stimulus components that drive the differences among conditions. The COVIS model predicts that the main driver of differences in performance will be Category Distribution because II and RB categories are learned by distinct neural systems. Additionally, we examined the impact of Training Task and distribution Sampling on performance. We predicted that learning differences would depend on Category Distribution, but also on Training Task and Sampling. In comparing generalization test performance, we found a marginally significant three-way interaction among Training Task, Category Distribution, and Sampling ($F(1,158) = 3.75, p = .055, \eta_p^2 = .023$). To understand the causes of this marginal interaction, we looked more closely at the two-way interactions.

We predicted that performance on II and RB Category Distributions would depend on Training Task, such that Overt training would better support learning RB categories and Incidental training would support learning II categories. We did not find support for this hypothesis; the interaction was not significant ($F(1,158) = 0.002, p = .96, \eta_p^2 = .000$). Instead, we found significantly better generalization of category learning for Overt training compared to Incidental training, irrespective of whether the categories were RB or II. Ignoring distribution Sampling, for both II and RB, Overt training resulted in significantly greater generalization of category learning than Incidental training (RB: $t(81) = 3.78, p < .001$, Cohen’s $d = .84$; II: $t(81) = 3.41, p = .001$, Cohen’s $d = .76$).

We also predicted that Incidental and Overt training tasks might have different effects on Probabilistic and Deterministic category learning. We predicted that Overt training would lead to
better performance for Deterministic categories than Probabilistic, but that Incidental training would lead to better performance for Probabilistic categories than Deterministic. We found that performance for Probabilistic and Deterministic distributions did depend on Training Task \( (F(1,158) = 5.18, p = .024, \eta^2_p = .032) \). In line with our predictions, Overt training led to better generalization of category learning for Deterministic category distributions than Probabilistic distributions \( (t(82) = 2.37, p = .020, \text{Cohen's } d = .52) \). Ignoring category type for Overt training, the Deterministic conditions had an average test accuracy of 68.2% and Probabilistic conditions had an average test accuracy of 54.1%. However, Incidental training did not result in significant differences in generalization across learning Deterministic and Probabilistic distributions \( (t(80) = 0.75, p = .46, \text{Cohen’s } d = .17) \). Ignoring category type for Incidental training, the Deterministic conditions had an average test accuracy of 49.4% and the Probabilistic conditions had an average test accuracy of 47.4.6%. These findings cannot be accounted for directly by the differences in difficulty between Deterministic and Probabilistic distributions because we used normalized accuracies in these analyses. In line with our predictions, Overt training led to better learning for Deterministic relative to Probabilistic categories. However, we predicted that Probabilistic categories would be learned better during Incidental training and we found that Deterministic and Probabilistic categories were learned equivalently during Incidental training.

Our third prediction was that performance of participants learning RB categories, Deterministic and Probabilistic sampling might not affect performance, but for II categories, the Deterministic sampling would be learned better than the Probabilistic sampling. We found an interaction between Category Distributions (II, RB) and the Distribution Sampling on generalization of category learning, \( (F(1,158) = 13.31, p < .001, \eta^2_p = .078) \). In support of our prediction, generalization of RB categories did not differ for Deterministic and Probabilistic
distributions ($t(64.1) = 1.56, p = .13$, Cohen’s $d = .39$, corrected for inequality of variances). We also note that this is different than what we found during overt training, where Deterministic was learned better than Probabilistic. After Incidental and Overt training, generalization of RB categories did not differ between Deterministic and Probabilistic distributions. Also in support of our prediction, for II categories, the Deterministic distributions resulted in significantly higher generalization accuracy than the Probabilistic distributions ($t(76.3) = 3.00, p = .004$, Cohen’s $d = .69$, corrected for inequality of variances). The ability to generalize RB categories is not affected by differences in sampling distributions, but generalization of II categories is worse with Probabilistic distributions than Deterministic.

![Confusion matrices for information-integration conditions in the generalization test.](image)

*Figure 6.* Confusion matrices for information-integration conditions in the generalization test. Each column represents the actual category identity of the exemplars played on a trial and each row represents the category response that the participant made. The shading gradient and
percentages within each cell represent how frequently participants gave a particular response for each category. Columns sum to 100%. To the right is a schematic diagram of the information-integration category structures (also shown in Figure 2).

Figure 7. Confusion matrices for rule-based conditions in the generalization test. Each column represents the actual category identity of the exemplars played on a trial and each row represents the category response that the participant made. The shading gradient and percentages within each cell represent how frequently participants gave a particular response for each category. Columns sum to 100%. To the right is a schematic diagram of the rule-based category structures (also shown in Figure 2).
Though overall generalization accuracy can give us some clues as to what participants were able to learn about these categories through overt or incidental training, this does not allow for a full understanding of the category representations that participants learned. To gain a better understanding of these representations, we constructed confusion matrices for each condition learning II (Figure 6) and RB (Figure 7) categories. These confusion matrices demonstrate participants’ response behavior in the generalization test based on the actual category that was presented to them. For correct responses, the actual category on a trial (columns) and the category of the participant’s response (rows) converge (Figures 6 and 7 on the positive diagonal). For incorrect responses, we can observe a clear pattern of confusion among multiple categories or a random confusion across all categories. Confusion matrices allow us to quantify similarities and differences among categories based on categorization errors during the generalization test.

The pattern of results in the confusion matrices for II categories demonstrates a tendency for participants in all four conditions to respond in a way that groups categories A and B together and groups categories C and D together (Figure 6). The confusable categories are not distinguished easily by either dimension used to construct the categories. Instead, this particular pattern of response is consistent with responses informed by integration along the positive correlation between the two dimensions. Note that this pattern of responses was similar across II conditions, despite quantitatively different levels of overall performance.

The pattern of results in the confusion matrices for RB categories demonstrates a different tendency (Figure 7). Participants in the RB Overt Deterministic condition were the most consistent in their responses across the four categories. These participants did not demonstrate a clear pattern of confusion among any of the categories in the generalization test, which may have stemmed from their higher accuracy in the generalization test. Participants in the other three RB
conditions demonstrated varying levels of confusion between categories B and C. Participants demonstrated clear response patterns that distinguished categories A and D from the other categories and there was some confusion between categories B and C. Categories B and C differ on both stimulus dimensions used to construct the categories, just as the distinct categories of A and D. In these three RB conditions, participants did not appear to be using simple unidimensional rules to separate the four categories into two groups. Instead, there is a more complex pattern of responses that even includes confusion of two categories that differ on both acoustic dimensions. The pattern of responses—and particularly of the errors—in the generalization test provides us some information about how participants represent the categories and the relations between the categories as a function of task and distribution sampling.

**General Discussion**

We examined learning and generalization of auditory categories across incidental and overt training tasks, likewise assessing the influence of probabilistic versus deterministic sampling of category distributions defined by a simple rule or requiring integration across dimensions. To our knowledge, this study is the first to compare dual-systems category learning across II and RB stimulus distributions in an incidental training task and an overt training task and the first to systematically examine the effect of the sampling distribution of the categories and the interaction with training task type. We aimed to understand the extent to which category learning generalizes to novel exemplars since generalization is a central characteristic of categorization. This further served to provide a common measure across the incidental and overt training tasks, which we focus on in discussing the results.

**Incidental versus Overt Training.** The results demonstrate that artificial nonspeech auditory categories can be learned incidentally under conditions in which participants to not
overtly make categorization decisions and are not informed that categories of sound relate to the primary (visual detection) task. Participants were engaged in a simple visual detection task and were not told that the sounds were important or related to the task, that the sounds were drawn from different categories, or that the sounds would later be central in an overt generalization task. The incidental category learning was apparent in overt labeling of novel generalization sounds at post-test, which requires a transfer of incidentally-acquired category knowledge to an explicit category labeling task. Across conditions, incidental training led to successful generalization of category learning across both II and RB stimulus distributions defined deterministically and probabilistically. This tells us that explicit awareness of the relevance of the feedback or even the goal of the task to learn and generalize their category knowledge is not necessary for category learning. This is notable because prior studies have almost exclusively examined learning with training tasks that involve explicit feedback following each overt categorization decision (Ashby et al., 2002; Chandrasekaran, Yi, et al., 2014; Dunn et al., 2012; Ell et al., 2009; Maddox, Filoteo, et al., 2004; Maddox et al., 2008; c.f. for a discussion about unsupervised learning see Ashby et al., 1999).

The COVIS model emphasizes the importance of feedback in driving learning, particularly in the case of learning II stimulus distributions. In this context, it may seem surprising that there was such robust incidental learning of II stimulus distributions. However, although the incidental training paradigm does not utilize feedback in the traditional manner of overt training tasks it should not be considered to lack feedback entirely. The consistent correlation of category exemplars with the location of visual targets presents a situation for which auditory categorization supports predictions regarding the primary visual detection task. These predictions are either correct or incorrect, as indicated by the ultimate appearance of the
visual target. In this way, categorization is incidentally associated with outcomes via the primary visual detection task. We have argued previously that this form of feedback may relate more closely to how sound categories are used in the world; they allow listeners to use variable sensory input to make predictions that support behavior in the larger environment, which sometimes lead to positive outcomes (Gabay et al., 2015; Lim & Holt, 2011). The present results demonstrate that this alternative, less overt, form of feedback is sufficient to support category acquisition across both RB and II stimulus distributions when they are sampled either probabilistically or deterministically. Even for II distributions, which COVIS posits to rely more heavily on feedback, neither overt awareness about the category learning task nor explicit feedback appear to be necessary for category learning.

There are important implications for theory. Based on the prior literature on visual and auditory category learning and the COVIS model, we predicted that categories defined by II stimulus distributions would be better learned via Incidental training than Overt training and, conversely, that categories defined by RB stimulus distributions would be better learned under Overt training, relative to Incidental training. Specifically, since the incidental task is speeded visual detection and not auditory categorization it directs attention away from overt categorization decisions. Thus, we hypothesized that learning II stimulus distributions would benefit from incidental training because overt reasoning is thought to hinder II learning (Ashby & Maddox, 2011). The data do not support the prediction; there was no interaction of training task and category stimulus distribution. Both RB and II stimulus distributions were learned better in the Overt, relative to the Incidental training task.

Overt training lead to better performance than Incidental training regardless of category type. One factor possibly contributing to this finding is that in the Incidental training task
involved a brief block in which the relationship between sound category and visual location as randomized (to covertly assess learning online). This short block may have been enough to differentiate the Incidental training condition from Overt training to influence generalization performance. Another possible explanation for the Overt training advantage is that the simple visual detection of the SMART incidental training task may not be fully tapping into the procedural learning system that best learns II categories. Therefore, caution is warranted in concluding that learning via overt training is necessarily always superior to learning via incidental training.

**Category Distribution Sampling.** The sampling distributions defining the categories impacted learning and generalization performance. This finding is critical because many speech and nonspeech auditory category learning studies have used highly stylized, deterministic distributions, whereas natural categories, including speech, are defined by more variable and probabilistic distributions. We predicted, based on the visual category learning results of Ell and Ashby (2006), that Sampling might affect learning of II categories, but not RB categories. Our generalization test results were consistent with Ell and Ashby’s (2006) findings that overlap affected learning of II categories, but not RB categories. We found poorer category generalization accuracy for II stimulus distributions defined Probabilistically compared to Deterministically. In contrast, category generalization accuracy was equivalent across Probabilistic and Deterministic RB stimulus distributions. While our results are consistent with the general premise from Ell and Ashby (2006)—that RB category learning is unaffected by differences in overlap and II category learning is affected—our finding that generalization for Deterministic distributions was better than generalization for Probabilistic distributions is inconsistent with their findings. Ell and Ashby (2006) found that moderately overlapping
categories, such as our Probabilistic distributions led to better II learning than categories that did not overlap, such as our Deterministic distributions. Of course, the sampling manipulation in the current study involved more than just overlap, which may account for the differences between our study and Ell & Ashby’s (2006). Additionally, this difference may have been driven by the stimuli themselves. It is possible that simple, verbalizable visual dimensions may be used differently by participants during learning than the auditory dimensions used in the current study. Further research is needed to disentangle the effects of overlap or sampling distribution on auditory II category learning.

Our results provide further evidence of the applicability of COVIS to auditory category learning and the instantiation of the multiple systems theory for auditory category learning. Although Ell and Ashby (2006) did not test generalization of learning to novel category exemplars, this finding is in accord with their conclusion that category overlap, one of the differences between our sampling distributions, affects II, but not RB, category learning.

In future research, it will be necessary to disentangle the potentially interacting effects of the factors defining Deterministic and Probabilistic category distributions, including overlap, number of exemplars, and stylistically sampled versus randomly sampled distributions. The Deterministic distributions mirroring those used in many nonspeech and speech category learning studies have fewer exemplars and less exemplar overlap both between and within categories compared to our Probabilistic distributions that were meant to more closely approximate natural category distributions. The differences in learning that were explained by the Sampling distributions underscores the significance of this factor in category learning. If our goal is to understand natural category learning, whether it is visual or auditory, it is critical to best approximate the natural structure of those categories in future experimental studies.
These results caution that reliance on simple, carefully designed deterministic input distributions may not capture the learning challenges involved in acquiring speech categories characterized by highly overlapping distributions across complex and multidimensional input dimensions (Hillenbrand et al., 1995; Swingley, 2009). If we are to generalize the conclusions about II categories and the mechanisms that are used to learn them, we also must carefully consider differences in distributions that can define different existing real-world speech categories (see Wanrooij & Boersma, 2013 for a similar argument about frequency distributional learning).

**Interaction of Sampling and Task.** We predicted that Sampling distribution may also interact with training task such that Incidental training might be better across Probabilistic distributions, whereas Overt training might be better across Deterministic distributions. In line with our predictions, Overt training led to better generalization of category learning for Deterministic, compared to Probabilistic, stimulus distributions. In contrast to our predictions, Incidental training resulted in equivalent generalization of category learning across Deterministic and Probabilistic stimulus distributions. This interacted with the type of stimulus distribution sampling, as well. For II stimulus distributions, the learning advantage of Overt training over Incidental training held for both Probabilistic and Deterministic distributions. For RB stimulus distributions with Deterministic sampling, there was an Overt training advantage. However, this advantage was not apparent for Probabilistic RB stimulus distributions. This highlights that important differences in category learning occur with different distribution sampling, training, and category types.

For the Probabilistic distributions, participants are receiving information about the category boundaries that is inherently less consistent, compared to feedback given for the
Deterministic distributions. All category exemplars in the Deterministic distributions fall perfectly within the hypothetical boundaries within acoustic space defining the respective categories. For the Probabilistic categories, there are a minority of exemplars from each category that cross these hypothetical boundaries, leading to category overlap. This means that the category-consistent feedback (incidental or overt) available in training is not as well-aligned with exemplar similarity in the Probabilistic, compared to Deterministic, conditions. The ambiguous nature of the alignment of the feedback signal with acoustic similarity may lead to less clear category representations, especially around the category boundary. Thus, this leads to a specific benefit of Deterministic over Probabilistic distributions in the Overt task when information is available to explicitly process feedback and incorporate it in future category decisions. On the hypothesis that feedback given in a deterministic manner depends more on explicit memory systems (Seger & Cincotta, 2005), the poorer alignment of exemplar similarity and feedback associated with the Probabilistic distributions may be less impactful during learning in the Incidental task if it draws from learning via more implicit procedural learning systems. The nature of the category distributions and the complexity of category sampling are important aspects to consider because they can greatly impact learning outcomes.

**Within-trial variability.** A key difference between the current study and previous studies investigating the dual systems of category learning is that we used within-trial variability in our training and testing paradigms. On each trial, participants heard five unique exemplars from within a single category. In typical dual system experiments, whether visual or auditory, participants encounter a single exemplar on each trial. This methodology allows experimenters to model the decision bound strategy response based on how a participant responds to each exemplar.
Previous research with auditory category learning, including speech, has demonstrated an overall benefit in generalization performance following training with high within-category variability (Bradlow et al., 1997; Iverson, Hazan, & Bannister, 2005; Liu & Holt, n.d.; Logan et al., 1991). This appears to be particularly potent when within-category variability is aligned with trial-level feedback. Using the same incidental training task as the present study, Gabay et al. (2015) found superior learning when participants experienced category exemplar variability within a trial, and therefore tightly coupled with task-driven predictions and feedback. Participants who experienced the same overall exemplar variability across individual trials in the experiment learned less. However, whereas within-trial variability is likely to have promoted learning and generalization in the present study, it also precluded the use of decision bound modeling to assess individual participant response strategies during learning. Current iterations of decision bound models map an individual’s decision boundary based on the location of a single exemplar in the stimulus space given their response. In future work, it will be useful to build decision bound models that can incorporate within-trial variability.

Strategy use during category learning across within-trial exemplar variability remains an open question for future research. Among many possible strategies, for example, it could be the case that participants use only one exemplar out of five that they experience on a trial to make their decision, or that the average similarity space of exemplars experienced within a trial influences decisions. Since distributional sampling had an influence on learning in the present research it will be informative to direct future research toward understanding how trial-level distributional statistics and longer-term distributional statistics that must be accumulated across an experiment interact to influence category learning.

To take a step in this direction, we examined the patterns of responses in the
generalization test to obtain a broad sense of participant strategy. Participants’ category confusions across the generalization test provide a window through which to approximate the kinds of representations learned. The confusion matrices make clear that similar overall performance in the generalization test can be arrived at via distinct paths. For the II categories, the pattern of confusability implies that listeners tended to group categories in a way that suggests integration across the dimensions, particularly in a positive-going direction. Intriguingly, this same pattern may be evident in the confusion matrices for RB categories. Rather than confuse RB categories distinguished by a single dimension in the stimulus space, listeners tended to make errors consistent with dimension integration across the positive-going dimension correlation. The apparent pattern of reliance on a positive-going integration strategy is consistent with recent results demonstrating a learning advantage for categories defined by a positive-going, compared to a negative-going, slope in this same stimulus space (Roark & Holt, submitted).

**Implications.** Although dual systems theory has been largely developed in the context of empirical data regarding visual category learning, recent work has very successfully applied it to auditory and speech categorization and yielded important insights (Chandrasekaran, Koslov, et al., 2014; Chandrasekaran, Yi, et al., 2014; Maddox & Chandrasekaran, 2014). Because the categorization challenges presented by auditory (and speech) signals are somewhat different from those of visual categories (Holt & Lotto, 2010), this also presents the opportunity to examine first-principles of the model in greater detail through the lens of auditory category learning. We view the present research as a necessary bridge between the auditory category learning research that has focused on the representations acquired in category learning and the highly influential COVIS approach that is beginning to influence auditory category learning.
research. Our results highlight that small differences in task demands result in quite different patterns of learning that interact with the sampling of category exemplars in acoustic space. Overt categorization decisions and explicit awareness of the category-learning task were not necessary for learning II or RB categories. In the present work, the most effective training approach involved overt training across deterministic category distributions. Since the majority of studies informing theoretical development have relied on just such category learning challenges, it is important to consider that laboratory-based studies may tend to overestimate the ease of category learning under more natural conditions that involve probabilistically-defined categories learned across incidental conditions. This is true for both auditory and visual studies. While we have used a specific pair of acoustic dimensions here, future work should examine the effect of these aspects of the training stimuli on learning with other acoustic dimensions and visual dimensions. The dual systems approach has not yet investigated the effect of these aspects of distributions and training task with either auditory or visual dimensions. Next-generation models of category learning will need to consider the nature of the complexity and overlap of sampling distributions, along with their interaction under more incidental learning situations, to better characterize how the system reacts to real-world category learning challenges.
References


http://doi.org/10.1016/j.bandc.2012.11.006

http://doi.org/10.1093/cercor/bhu236
Appendix A

Category Distribution Means, Variances, and Covariances

Deterministic Category Distribution Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (CF, MF)</th>
<th>Variance (CF, MF)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>II: Category A</td>
<td>(674, 197)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
<tr>
<td>II: Category B</td>
<td>(865, 312)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
<tr>
<td>II: Category C</td>
<td>(1056, 197)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
<tr>
<td>II: Category D</td>
<td>(865, 82)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
<tr>
<td>RB: Category A</td>
<td>(730, 278)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
<tr>
<td>RB: Category B</td>
<td>(1000, 278)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
<tr>
<td>RB: Category C</td>
<td>(730, 116)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
<tr>
<td>RB: Category D</td>
<td>(1000, 116)</td>
<td>(3377.5, 1213.3)</td>
<td>0</td>
</tr>
</tbody>
</table>

Probabilistic Category Distribution Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (CF, MF)</th>
<th>Variance (CF, MF)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>II: Category A</td>
<td>(701.1, 199.4)</td>
<td>(6306.0, 1629.8)</td>
<td>-15.5</td>
</tr>
<tr>
<td>II: Category B</td>
<td>(853.1, 310.7)</td>
<td>(7032.6, 2281.8)</td>
<td>494.9</td>
</tr>
<tr>
<td>II: Category C</td>
<td>(1059.0, 196.5)</td>
<td>(6677.8, 2207.2)</td>
<td>-36.7</td>
</tr>
<tr>
<td>II: Category D</td>
<td>(863.1, 86.8)</td>
<td>(6156.6, 2123.3)</td>
<td>119.8</td>
</tr>
<tr>
<td>RB: Category A</td>
<td>(735.8, 277.1)</td>
<td>(7198.3, 2600.3)</td>
<td>-201.7</td>
</tr>
<tr>
<td>RB: Category B</td>
<td>(922.4, 279.8)</td>
<td>(7496.4, 2082.0)</td>
<td>-804.6</td>
</tr>
<tr>
<td>RB: Category C</td>
<td>(749.0, 116.7)</td>
<td>(5081.2, 2251.4)</td>
<td>156.3</td>
</tr>
<tr>
<td>RB: Category D</td>
<td>(1007.8, 114.3)</td>
<td>(5102.4, 1891.0)</td>
<td>367.6</td>
</tr>
</tbody>
</table>

The Probabilistic category distributions were created by defining a two-dimensional Gaussian distribution with the same means and increased variances relative to the Deterministic category distributions. Then, 100 random samples from that underlying distribution were taken to form the Probabilistic category distributions. Thus, the means, variances, and covariances between categories in the Probabilistic category distributions vary relative to the underlying multidimensional Gaussian distributions.
Imagine a cat. A specific cat may come to mind—maybe your pet cat or a cat that you saw recently. Or, a more generic cat may pop into your mind. Now, imagine if you had only ever seen orange tabby cats. In this case, it may be a reasonable conclusion that all cats must be orange and white because every cat you had ever seen had been orange and white. Our conceptualization of specific categories is defined by the instances we have experienced from within a category. These instances may be very similar to one another, like orange tabbies, or relatively more variable, like all cat breeds. How we experience within-category variability in the course of category learning ultimately impacts perception and understanding of the categories.

The influence of within-category variability on learning and generalization has been a particular focus in studies of auditory and speech category learning. For example, experiencing exemplars with high acoustic-phonetic variability across category learning has proven to be successful in fostering category acquisition and generalization among adults learning difficult non-native speech categories (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991; Wang, Spence, Jongman, & Sereno, 1999). This acoustic variability typically arises from sampling speech exemplars produced by multiple talkers across multiple phonetic contexts that results in a distribution of experience that samples a wider region of the multidimensional acoustic input space (Lively et al., 1993; Logan et al., 1991) than exemplars drawn from a single talker or phonetic context (Strange & Dittmann, 1984). High variability acoustic sampling has been successful in
facilitating learning in several category-learning contexts, including training native Japanese listeners on difficult English /r/-/l/ categories (Bradlow et al., 1999, 1997, Lively et al., 1993, 1994; Logan et al., 1991; Shinohara & Iverson, 2018; Zhang et al., 2009) and training native English speakers on Mandarin tone categories and general pitch contour categories (Perrachione, Lee, Ha, & Wong, 2011; Wang, Jongman, & Sereno, 2003).

Outside of speech category learning, training with highly variable exemplars has been found to improve infant word learning (Apfelbaum & McMurray, 2011; Galle, Apfelbaum, & McMurray, 2015; Quam, Knight, & Gerken, 2017; Richtsmeier, Gerken, Goffman, & Hogan, 2009; Rost & McMurray, 2009, 2010), second language vocabulary learning (Barcroft & Sommers, 2005), and sequence learning (Gómez, 2002). High variability training is also beneficial to learning outside of the auditory domain. Visual categories sampling large ranges across the exemplar space are learned better than categories with smaller ranges (Maddox & Filoteo, 2011; Perry, Samuelson, Malloy, & Schiffer, 2010). Visual categories are also learned better when participants have seen multiple exemplars from within a category, relative to a single exemplar (Finch, Carvalho, & Goldstone, 2014). The general benefit of variability in a training set has also been found in more complex tasks, such as math problem solving (Braithwaite & Goldstone, 2015; Sanders, Gonzalez, Murphy, Pesta, & Bucur, 2002), reading (Apfelbaum, Hazeltine, & McMurray, 2012), and perceptual skill acquisition in pilots (Huet et al., 2011).

Typically, experiencing high variability across exemplars promotes overall learning (Logan et al., 1991; Perrachione et al., 2011) and improves generalization of learning to novel category exemplars not encountered in learning (Hahn, Bailey, & Elvin, 2005; Liu, 2014; Lively et al., 1993; Logan et al., 1991; Maddox & Filoteo, 2011; Posner & Keele, 1968). However, variability in a category environment is not always beneficial for learning. Presenting this kind of
variability may also slow initial learning (Amitay, Hawkey, & Moore, 2005; Hahn et al., 2005; Maddox & Filoteo, 2011; Sanders et al., 2002). In general, training with high within-category variability is thought to improve learning and generalization because a wider range of exemplars allows for a richer representation of within-category information, filtering out more idiosyncratic details of exemplars and providing a better sense of common features within a category. However, although training with high variability is generally successful in inducing robust, generalizable category learning, it is not yet clear which dimensions of variability are significant, or how they benefit learning.

The relationship of variability to the structure of the categorization task may be a factor that plays a role in whether variability helps or harms category learning. When categories are defined across multiple input dimensions, some dimensions may be more informative in signaling category membership than others. There is evidence that experiencing variability along a diagnostic dimension may highlight it in learning, facilitating learning (Antoniou & Wong, 2016). In contrast, experiencing variability along both diagnostic and irrelevant dimensions can hinder learning (Antoniou & Wong, 2016). However, other studies demonstrate that experiencing variability on irrelevant dimensions may support learning by drawing attention to relevant dimensions (Braithwaite & Goldstone, 2015; Rost & McMurray, 2010). Thus, how variability that experienced across an input dimension influences category learning remains unclear.

The relationship of variability to the structure of the training task may be important. Nearly all studies investigating the impact of variability on learning have compared learning across low- and high-variability category exemplars so that the overall variability of a training set differs across groups of learners. However, a recent study demonstrates an impact of
exemplar variability on auditory category learning even when the overall variability of the training set is *matched* across groups, but manipulated *within a trial* (Gabay, Dick, Zevin, & Holt, 2015). One group of listeners experienced multiple repetitions of a single category exemplar prior to responding. Another group experienced variable exemplars drawn from the same category prior to responding. Under such a contrast, a simple exemplar-accumulation model might predict no differences in learning across these learning contexts, since *overall* variability in the training set was equivalent. In contrast, Gabay et al. observed that training across trials with within-trial variability led to more robust and generalizable category learning compared to training with no within-trial variability. Thus, Gabay et al. reported that experiencing stimulus variability tightly coupled with feedback and prediction in a learning task supports learning compared to learning the same training set organized without this coupling.

In contrast to nearly all other studies investigating the role of variability in training, Gabay et al. (2015) examined learning across *incidental* training. Participants were not informed about the existence of auditory categories, they did not make overt auditory category decisions, and they did not receive explicit feedback about categorization; they simply made a speeded response to report the location of the visual target. In their incidental learning task, five sound exemplars from within an auditory category preceded the appearance of the visual target on each trial. Although it was not directly relevant to performance of the simple visual detection task, the auditory categories perfectly predicted the location of the upcoming visual target. Thus, if participants learned the auditory categories it was possible for them to use the auditory categories to predict visual target location and response. Category learning was covertly monitored via a reduction in reaction time to report visual target location across trials and a cost to this reaction time when the category-to-location regularity was destroyed.
In the case for which within-trial sound category variability was present, listeners heard five unique category exemplars prior to the appearance of the visual target. For these trials, the category exemplar variability was tightly coupled with the visual target and the participant’s motor response to report its location. Gabay et al. (2015) found that the tight coupling of variability with incidental feedback facilitated learning and generalization, relative to training with no within-trial acoustic variability for which a single category exemplar was repeated five times in a trial. This was true even though participants experienced the same overall category exemplar variability across the course of the two types of training; only the relationship of exemplar variability to the incidental task demands of the visual target detection task differed.

However, since incidental category learning may involve mechanisms distinct from the explicit training with overt feedback (Lim et al. 2014), it is not clear how training with within-trial variability might affect learning on more traditional supervised learning tasks, in which exemplars are more overtly coupled with category decisions and feedback. If it is the case that experiencing exemplar variability drives learning due to its coupling with task demands and prediction in the incidental learning task, the results should be replicable for learning across overt category decisions and explicit feedback. One objective of the present research is to test this hypothesis to understand how exposure to variability during learning shapes category representations.

Another clear difference between this task and traditional categorization paradigms is that five exemplars are presented on each trial. In traditional paradigms, it is more typical for a single exemplar to be presented once before a response. In this way, this paradigm is similar to blocked training methods (Carvalho & Goldstone, 2014a, 2014b, 2015). In this incidental training paradigm, participants experienced five exemplars from a single category before the made their
response, thus experiencing stimuli within a trial ‘blocked’ by category. Carvalho and Goldstone (2014a) demonstrated that learning categories blocked by category brings attention to similarities within a category. This kind of training may then be beneficial for categories for which it is especially difficult to detect similarities within a category (Carvalho & Goldstone, 2014a). Thus, we might predict that participants will benefit from this kind of variability because the similar exemplars are presented together within a trial.

An additional objective of the present study is to understand how variability may impact category learning when the categories make greater or lesser demands on selective attention to individual input dimensions. Studies of adult speech category learning would seem to indicate benefits of acoustic variability, whatever its source. However, at a fundamental level input variability is variability across specific input dimensions. So far, understanding the nature of the dimensions that contribute to acoustic variability in the highly multidimensional acoustic space characterizing speech categories has not been a major focus of research. Yet, we know from parallel literatures that category learning can be highly sensitive to the way that exemplars are sampled across input dimensions, even to the extent that learning may be dominated by different underlying mechanisms as a function of the sampling of category exemplars. For example, the Competition between Verbal and Implicit Systems (COVIS) model proposes that there is an explicit, hypothesis-testing system and an implicit, procedural learning system involved in category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) and that these dual systems are optimally suited for learning of distinct samplings of categories across input dimensions.

The explicit system is thought to involve selective attention to single input dimensions, hypothesis testing, rule application, and rule-switching and thus is best suited to learn rule-based (RB) categories, which can be distinguished by verbalizable rules across single input dimensions.
RB categories are thought to benefit from overt reasoning about the rules that distinguish the categories (Ashby & Maddox, 2011; Maddox, Ashby, Ing, & Pickering, 2004). This explicit system involves neural structures that are involved in hypothesis testing, working memory, and feedback processing, such as the prefrontal cortex and head of the caudate nucleus in the basal ganglia (Ashby et al., 1998; Ashby & Maddox, 2011; Nomura et al., 2007).

In contrast, the implicit system is thought to involve procedural learning mechanisms and is best suited to learn information-integration (II) categories, which require pre-decisional integration across multiple input dimensions not benefited by selective attention to individual dimensions. The implicit system involves procedural learning mechanisms in the body and tail of the caudate nucleus and the putamen (Ashby et al., 1998; Ashby & Maddox, 2011; Nomura et al., 2007). As opposed to RB categories, learning II categories does not benefit from overtly reasoning about the relations between the categories and doing so may even impair learning (Ashby & Maddox, 2011; Grimm & Maddox, 2013).

Formulated for understanding visual categorization, the COVIS model has been expanded recently into the auditory system and to both nonspeech and speech categories (Chandrasekaran, Koslov, & Maddox, 2014; Chandrasekaran, Yi, & Maddox, 2014; Goudbeek, Cutler, & Smits, 2008; Goudbeek, Swingley, & Smits, 2009; Roark & Holt, 2018). Researchers have proposed that some speech categories may function as II categories because they are defined by multiple dimensions, are difficult to describe with verbalizable rules, and often require integration across multiple acoustic dimensions (Hillenbrand, Getty, Clark, & Wheeler, 1995; Yi, Maddox, Mumford, & Chandrasekaran, 2014). However, it is also possible that some speech categories may be better described as RB categories that benefit from selective attention to acoustic information (Holt, Tierney, Guerra, Laffere, & Dick, 2018). Especially once they are
well-learned, some speech categories have dominant dimensions that may function similarly to unidimensional rule-based judgments. For example, to make distinctions between /b/ and /p/ or /d/ and /t/ listeners strongly rely on voice-onset time (VOT; Abramson & Whalen, 2017). This strong reliance may function similarly to selective attention, making selective attention to VOT an RB categorization problem. Here, we attempt to bridge studies on speech category learning that have focused on within-category variability across an experimental training set with studies of category learning through the dual systems perspective. Specifically, experience of within-category variability and the effects on subsequent representations may depend on the learning mechanisms involved, driven by how the categories relate to the dimensions in the stimulus space and the variability across which they are sampled.

Several studies have investigated how visual II and RB category learning is impacted by high variability across the range of stimuli. Experiencing category exemplars that sample a large range across stimulus space slows initial learning of II categories relative to exemplars sampling a smaller range, but improves generalization to novel stimuli and exemplars from novel regions of the stimulus space (Maddox & Filoteo, 2011). In contrast, RB category learning is unaffected by altering the variability in the range of the stimuli encountered during training (Maddox & Filoteo, 2011; Maddox, Filoteo, & Lauritzen, 2007; Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005). The visual II categories in these studies behave similarly to speech categories from other studies—high variability in the training environment initially slows learning, but ultimately leads to better generalization performance relative to low variability. However, as with speech category learning, there have been no investigations of the effects of within-trial variability on II and RB category learning. One aim of the current study is to examine the effect of within-trial variability exposure on II and RB category learning.
There are reasons to believe that experiencing within-trial exemplar variability during training may differentially affect RB and II category learning because the two systems optimal for learning these two category structures differ in their sensitivity to the timing of feedback. RB category learning is robust to disruptions in feedback, including delays in feedback of up to 10 seconds between response and feedback (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005; Worthy, Markman, & Maddox, 2013). In contrast, II category learning is very sensitive to the timing between the presentation of the stimulus, the response, and feedback (Maddox et al., 2003; Maddox & Ing, 2005; Worthy et al., 2013). The sensitivity of II category learning to feedback timing is thought to be due to the nature of the implicit system and the procedural learning mechanisms involved in learning the stimulus-response associations (Worthy et al., 2013). Due to the timing of the signaling in the procedural learning system, stimuli and responses that occur within 500 ms of the feedback can be reinforced (Valentin, Maddox, & Ashby, 2014; Worthy et al., 2013). For this reason, exposure to multiple exemplars within a single trial may not allow for the procedural learning system to make associations among the multiple exemplars, the motor response, and the feedback. Instead, it could be the case that only the stimulus that occurs within the 500 ms window of the feedback might be strengthened in the association. If this is true, II category learning optimally driven by this implicit system may not benefit from exposure to high within-trial variability.

A resulting prediction is that that II categories will not benefit from within-trial variability due to of the constraints of the timing of the procedural learning system. The timing of the system is such that only the final exemplar played on a trial will become bound with the feedback and form the stimulus-response association in a meaningful way. The other exemplars experienced in a trial will be too distant in time (> 500 ms) from the response or feedback to
become linked in a meaningful way. Thus, the prediction is that learning II categories will not be affected by high within-trial variability.

In contrast, another prediction is that RB categories may benefit from within-trial variability because learning of these kinds of categories via the explicit system is robust to disruptions in feedback. Thus, any delay from presentation of an exemplar to the feedback should not affect how the categories are learned. Additionally, RB categories may benefit from within-trial variability because hearing variability from within a category on a trial may allow participants to discover rules that distinguish categories and guide them to what is relevant and irrelevant in the variation among exemplars.

The current study investigates the role of exposure to variability in category learning and the formation of representations with the aim of better understanding the factors that drive efficient and robust learning. To control for effects of stimulus variability, we equate global exemplar variability across the experiment while differing individuals’ exposure to variability within a single trial. We utilize nonspeech dimensions and artificial auditory categories to gain complete control over the precise distributions of stimuli, and variability across input dimensions, that participants experience. We expect that understanding the basic learning principles that drive auditory category learning will inform the mechanisms available to speech category learning. We examine the impact of training with within-category variability on a single trial on learning information-integration and rule-based categories defined across two nonspeech acoustic dimensions. Generalization performance is thought to be supported by variability and has been shown to improve when participants experience within-category variability. Here, we investigate whether the course of learning and its generalization to novel exemplars may be different for information-integration and rule-based categories.
Methods

Participants

Participants were 87 undergraduates ages 18-25 (56 female, 30 male, 1 unreported) from Carnegie Mellon University who were given either $10 or partial course credit for participation. Participants were assigned to one of four conditions. The conditions varied on two factors: structure of the to-be-learned categories (information-integration or rule-based) and whether there was within-trial variability during training (present or absent). An additional three subjects were run, but excluded, from analyses because of equipment failure (1 subject from the rule-based without variability condition, 1 subject from the information-integration without variability condition, 1 subject from the information-integration with variability condition). There were 21 participants in each of the four conditions. Data for the variability-present conditions were also reported in another manuscript (Roark & Holt, 2018).

Stimulus Distributions Defining Categories

Figure 1 illustrates the sampling of acoustic exemplars across an input space in two acoustic dimensions to define four auditory categories. Rule-based categories (Figure 1a) were defined distributionally such that two orthogonal decision boundaries, one across each stimulus dimension, could optimally differentiate the categories. Information-integration categories (Figure 1b) were defined across distributions that required integration of the acoustic dimensions for optimal differentiation of the four categories (diagonal decision boundaries).
Figure 1. Stimulus distributions for the (a) rule-based categories and (b) information-integration categories.

The category distributions for each condition were constructed with MATLAB R2014a such that each category was defined by 100 stimulus exemplars sampled from a bivariate normal distribution of which the means, variance, and covariance matrix were defined for each category (Table I). With four categories per condition, 400 unique stimulus exemplars defined the category learning challenge for each condition. Half of the exemplars from each category were randomly selected for presentation in training; the other half was withheld for testing in the generalization test phase. Sampling was the same for all participants within a condition. Thus, the acoustic variability experienced was the same for all participants within the rule-based conditions and also the same for all participants within the information-integration conditions, whether variability was present on a trial-wise basis or not.
TABLE I

Category Distribution Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (CF, MF)</th>
<th>Variance (CF, MF)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A</td>
<td>(701.1, 199.4)</td>
<td>(6306.0, 1629.8)</td>
<td>-15.5</td>
</tr>
<tr>
<td>Category B</td>
<td>(853.1, 310.7)</td>
<td>(7032.6, 2281.8)</td>
<td>494.9</td>
</tr>
<tr>
<td>Category C</td>
<td>(1059.0, 196.5)</td>
<td>(6677.8, 2207.2)</td>
<td>-36.7</td>
</tr>
<tr>
<td>Category D</td>
<td>(863.1, 86.8)</td>
<td>(6156.6, 2123.3)</td>
<td>119.8</td>
</tr>
</tbody>
</table>

Rule-Based

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (CF, MF)</th>
<th>Variance (CF, MF)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A</td>
<td>(735.8, 277.1)</td>
<td>(7198.3, 2600.3)</td>
<td>-201.7</td>
</tr>
<tr>
<td>Category B</td>
<td>(922.4, 279.8)</td>
<td>(7496.4, 2082.0)</td>
<td>-804.6</td>
</tr>
<tr>
<td>Category C</td>
<td>(749.0, 116.7)</td>
<td>(5081.2, 2251.4)</td>
<td>156.3</td>
</tr>
<tr>
<td>Category D</td>
<td>(1007.8, 114.3)</td>
<td>(5102.4, 1891.0)</td>
<td>367.6</td>
</tr>
</tbody>
</table>

Stimuli

The acoustic category exemplars were frequency-modulated tones differing in center, or carrier, frequency (CF) and modulation frequency (MF). The dimensions and range of the category spaces were developed based on Holt & Lotto (2006). Each stimulus was created from a sine wave tone with a particular CF modulated with a depth of 100 Hz at the corresponding MF. The stimuli were root-mean squared (RMS) matched. Stimuli were synthesized in MATLAB R2014a.

Procedure

Participants were trained in an overt classification task, under the control of E-prime software (Psychology Software Tools, Inc., Sharpsburg, PA) with acoustic category exemplars presented diotically over Beyer DT-150 headphones. Figure 2 shows details of the procedure.

Training Task. Participants were informed that four auditory categories characterized the stimuli and made an overt decision about category identity with a button press. They received overt feedback about their response in the form of a red X in the location of the correct answer.
Participants were instructed to use this feedback to inform future responses.

Each of the four training blocks had 96 trials for a total of 384 trials. On each trial, five 300-ms sound exemplars played with a 50-ms inter-stimulus interval. In this way, each trial contained five blocked exemplars from one category. Participants saw four boxes on a computer monitor mounted at eye level. Participants could respond after the final sound. After each response, there was a 500-ms pause after which feedback appeared, in the form of the red X in the location associated with the sound category from which the sound exemplars had been drawn. Previous research has demonstrated that a 500 ms delay impairs neither II category learning nor RB category learning (Worthy et al., 2013). The feedback remained on the screen for 500 ms and was followed by a 1000 ms ITI.

**Generalization Test.** Participants completed a generalization test after training in which they categorized 96 novel category exemplars not experienced in training drawn from the same category distributions on which individual participants were trained (II or RB). Five 300-ms (50-ms ISI) sounds played on each trial, as in training. Responses were possible immediately following the final sound. Participants received no response feedback during the generalization test.
Figure 2. Details for the training task including block breakups, assignment of categories to locations on screen, and variability manipulation.

**Within-Trial Variability.** The presence or absence of within-trial variability is a main variable of interest. As illustrated in Figure 2, five distinct sound exemplars drawn from a single category defined a trial in the Variability-Present condition. The other half of subjects experienced no within-trial category exemplar variability. For this Variability-Absent condition the same exemplar was repeated five times in a single trial. Stimuli during the experiment were selected randomly without replacement. There was no systematic selection of exemplars for Variability-Present trials; each of the five stimuli belonged to a single category and were randomly selected without replacement from that category’s exemplars. As there was no structure in the variability within a trial, on each Variability-Present trial, exemplars varied on both the CF and MF dimensions, which both varied within each of the categories being learned.

The intent of the variability manipulation was to manipulate experience with within-category exemplar variability within, versus between, trials. The overall global exemplar
variability experienced across the experiment was kept relatively constant between conditions, subject to differences emerging from the randomization scheme. However, there were no statistically-significant differences in the mean CF and MF experienced or the variability (standard deviation) of CF and MF values experienced across the conditions. This was true for both II and RB categories.

**Results**

**Training Task**

**Data Processing and Normalization.** There were slight differences in the optimal possible accuracies that could be achieved across conditions, as a function of the distributions and decision boundaries that defined the conditions. The II distributions had an optimal accuracy of 90.25% and the RB distribution had an optimal accuracy of 90.5%. To more fairly compare categorization accuracy across conditions, we computed a normalized accuracy whereby

\[
\text{Normalized Accuracy} = \frac{\text{Raw Accuracy}}{\text{Optimal Accuracy}}
\]

Normalized accuracies, with a logit transform, were used in all analyses.

**Block-by-block Accuracy.** Figure 3 shows the block-by-block categorization accuracy across conditions. A 2 x 2 x 4 repeated-measures ANOVA examining the effect of category distribution (II, RB), within-trial variability (Present, Absent), and block (1-4) revealed effects of learning across training. Categorization accuracy differed significantly across blocks \((F(3,240) = 4.93, p = .002, \eta_p^2 = .058)\), but the trajectory of learning across blocks did not depend on category distribution \((F(3,240) = 2.23, p = .085, \eta_p^2 = .027)\), within-trial variability \((F(3,240) = 0.18, p = .91, \eta_p^2 = .002)\), or the interaction between category distribution, variability, and block \((F(3,240) = 0.24, p = .87, \eta_p^2 = .003)\). The general pattern of learning was similar for all conditions across training blocks. According to Bonferroni-corrected post hoc comparisons,
participants demonstrated similar performance in the first and second blocks ($p = .10$).

Performance was significantly better in Blocks 3 and 4 compared to Block 1 (Block 3 vs Block 1: $p = .009$, Block 4 vs Block 1, $p = .024$), but all other comparisons were not statistically different ($ps = 1.0$).

![Figure 3](image-url)

**Figure 3.** Normalized accuracy across the four training blocks for participants learning information-integration (left) and rule-based categories (right) with variability absent or present. Performance is shown relative to chance (25%). Ribbon error bars represent the standard error of the mean. Black dots represent the mean and each individual’s average accuracy for each block is also shown.

It appears that much of the learning occurred within the first block. According to one-sample t-tests comparing performance in the first block to chance (25%), participants across all conditions performed above chance in the first block (II-Present: $M = 51.0\%$, $t(20) = 12.4$, $p < .001$, $d = 5.55$, II-Absent: $M = 51.0\%$, $t(20) = 19.2$, $p < .001$, $d = 8.58$, RB-Present: $M = $
As is evident in Figure 3, categorization accuracy during training depended both on the statistical sampling of the categories and whether exemplar variability was experienced within trials \((F(1,80) = 15.6, p < .001, \eta_p^2 = .16)\). The ANOVA also revealed significant main effects of category sampling \((F(1,80) = 7.25, p = .009, \eta_p^2 = .083)\) and variability \((F(1,80) = 20.6, p < .001, \eta_p^2 = .21)\), but because there was a significant interaction, only the interaction effects are interpreted.

The results revealed that presence of within-trial exemplar variability did not affect performance for II categories \((t(40) = .46, p = .65, d = .14)\). In contrast, categorization of RB categories was significantly more accurate among participants experiencing within-trial variability, compared to no within-trial exemplar variability \((t(40) = 5.17, p < .001, d = 1.60)\). Although above-chance performance emerged early in training, the differences among conditions were consistent across training. Performance was superior when RB categories were experienced with within-trial variability compared to when that variability was absent. In contrast, the accuracy during learning of II categories did not differ based on whether there was within-trial variability.

**Generalization Test Accuracy**

The main variable of interest is the generalization test accuracy—the ability of participants to successfully transfer category learning to the labeling of novel category exemplars not experienced in the context of training with feedback (Figure 4). Just as for the accuracy during the training task, we used the transformed normalized accuracies for analyses of generalization test performance to more fairly compare RB and II category learning.
Figure 4. Normalized accuracy in the generalization test for participants who learned information-integration (left) and rule-based categories (right) with either variability absent or present. Performance is shown relative to chance (25%) with a point for each individual subject and the mean and standard error shown for each condition in black.

According to one-sample t-tests, generalization test categorization accuracy was greater than chance (25%) for all groups, indicative of category learning. Participants who were trained on II categories had an average test accuracy of 52.1% with within-trial variability ($t(20) = 9.59, p < .001, d = 4.29$) and 53.8% without within-trial variability ($t(20) = 18.3, p < .001, d = 8.19$).

Participants who were trained on RB categories had an average test accuracy of 67.6% with within-trial variability ($t(20) = 16.6, p < .001, d = 7.43$) and 56.0% without within-trial variability ($t(20) = 15.7, p < .001, d = 7.03$).

A two-way ANOVA revealed that performance in the generalization test depended on the interaction of sampling of category exemplars across the input dimensions (II or RB) and
whether there was within-trial category variability \((F(1, 80) = 8.40, p = .005, \eta_p^2 = .095)\). As predicted, generalization of II category learning was not impacted by the presence or absence of within-trial variability, but generalization of RB category learning was better when there was within-trial variability. There were also significant main effects of category sampling \((F(1,80) = 14.4, p < .001, \eta_p^2 = .15)\) and variability \((F(1,80) = 5.55, p = .021, \eta_p^2 = .065)\), but because there was an interaction, only interactions are interpreted.

Confusion matrices were constructed to compare how participants responded in the generalization test to the actual category identity of the stimuli presented on each trial. This provides insight into what participants learned about the individual categories and what difficulties they encountered distinguishing between categories. Confusion matrices for participants in the RB conditions are depicted in the top panels of Figure 5, with a key assigning arbitrary letter labels to category input distributions at the right. This visualization demonstrates that the categories that were most confusable by participants in the RB conditions were the B and C categories. These categories differed on both the CF and MF dimensions. Even so, participants frequently responded B when the actual category of the stimulus presented was C, and vice versa. Participants learning RB categories infrequently confused A and D categories with other categories. Examining the confusion matrices across Variability-Present and Variability-Absent training conditions, the benefit of within-trial variability for RB categories appears to be support in differentiating easily confusable categories. Within-trial variability seems to have advantaged generalization of learning to the difficult categories B and C that were more easily confusable according to the confusion matrix data.
Figure 5. Confusion matrices of the response category of the participants in the generalization test relative to the actual stimulus category. (Top) Rule-based distributions. (Bottom) Information-integration distributions. Columns sum to 100% and describe how participants responded relative to the actual category of the stimulus.

Confusion matrices for participants in the II conditions are depicted in the bottom panels of Figure 5, with a key assigning arbitrary letter labels to category input distributions at the right. These categories are distinct from the RB categories as they are arranged differently within the input space. Within the II category distributions, participants learning the II categories demonstrated consistent confusability between the A and B categories and between the C and D categories. These pairs of categories differ on both acoustic dimensions. The confusion matrices
for II participants across Variability-Present and Variability-Absent demonstrate the same pattern of confusability. Whereas training with within-trial variability seemed to increase the distinguishability between the most easily confused categories within the RB distributions, training with variability did not help distinguish confusable categories within the II distributions.

**Discussion**

The results of this study demonstrate that rule-based categories benefit from training with within-trial variability, but information-integration categories do not. Specifically, training with high variability within a single trial appears to reduce the confusability between categories that are the most difficult to distinguish for RB categories. Even though the overall exemplar variability was identical for all participants, learning RB categories benefitted from exposure to variability within a trial whereas learning II categories was not affected.

**Within-trial variability did not affect information-integration category learning**

Based on the timing dynamics of the implicit procedural learning system, which optimally drives II category learning (Valentin et al., 2014; Worthy et al., 2013), we predicted that II category learning would not be affected by within-trial variability exposure. The implicit system is thought to involve a procedural learning process that learns to assign particular motor responses to regions of perceptual space (Ashby, Paul, & Maddox, 2011; but see Cantwell, Crossley, & Ashby, 2015). The process that leads to feedback sensitivity in the implicit system is potentially linked to dopamine and calcium signaling in the striatum (Valentin et al., 2014; Worthy et al., 2013). Levels of dopamine and calcium peak simultaneously in the striatum within 500 ms of stimulus presentation, providing a potential cellular mechanism for feedback sensitivity in II category learning (Valentin et al., 2014). Because of the signaling processes in the striatum, the ideal delay period between the response and feedback for learning II categories
is 500 ms (Worthy et al., 2013). At this delay period, people learning II categories perform the best and use optimal integration strategies to learn the categories (Worthy et al., 2013). In the current study, the within-trial nature of the exemplar variability meant that only the final sound exemplar was presented was within the optimal 500 ms window for II category learning. Consistent with this conceptualization, the results demonstrate that II category learning and generalization are unaffected by the presence of within-trial variability.

Whereas the timing dynamics of the implicit system lay out a clear prediction for the effects of within-trial variability, they make no prediction for manipulations of variability arising from the range of exemplars. Indeed, contrary to the present results, prior studies have observed that II category learning and generalization benefit from variability when it comes in the form of sampling across a wide range of input space (Maddox & Filoteo, 2011). In the current study, the variability experienced across conditions was equivalent. However, the timing of feedback required for learning in the implicit system was not aligned with the within-trial exemplar variability. The present results are consistent with prior research that demonstrates that II categories cannot be learned without feedback and even feedback delays as long as 1 or 2.5 seconds can drastically disrupt learning (Ashby, Queller, & Berretty, 1999; Maddox et al., 2003; Worthy et al., 2013). The nature of variability and how it relates to the timing of feedback is critical in learning categories that require information integration across input dimensions.

Also relevant in this regard, researchers have examined how grouping and deferring feedback during learning impacts II and RB category learning (Smith et al., 2014). Smith et al. presented participants with blocks of six trials. After each of the six trials, participants made a categorization response. However, feedback was only presented after all six trials and not after each response. The link between the stimulus and response was upheld, but the link between the
stimulus-response and feedback was disrupted. In this context, II category learning was strongly impaired compared to when the link with the feedback was preserved (Smith et al., 2014). In the current study, participants experiencing variability heard five exemplars from within a single category and made a response only after all five exemplars were presented, followed immediately by feedback. Thus, the link between the stimulus-response and feedback was only present for the final sound that was played. The current study complements these previous findings by demonstrating that II category learning is not boosted by within-trial variability.

**Within-trial variability improved rule-based category learning**

In contrast to II categories learned optimally through the implicit system, RB categories learned optimally through the explicit system are robust to disruptions in feedback (Ashby et al., 1999; Maddox et al., 2003, 2004; Maddox & Ing, 2005). Additionally, experience of variability within a single trial may make it easier for participants to generate and test hypotheses about the relations among exemplars within a category, which is particularly beneficial for RB categories (Ashby & Maddox, 2011). Thus, we predicted that learning and generalization of RB categories would benefit from within-trial exemplar variability. The results supported this prediction; learning and generalization performance were higher for participants learning RB categories with within-trial variability than those learning without within-trial variability.

RB category learning is unaffected by manipulations of overall exemplar variability across an input space (Maddox et al., 2007, 2005). RB categories are thought to be learned primarily by creating and selecting hypotheses for rules that distinguish categories (Ashby et al., 1998). Thus, increasing variability across the exemplar space does not disrupt or benefit learning because it does not change the rule defining the category boundary, which can be learned quickly and successfully generalized to novel category exemplars. The distinction between within-trial
variability and overall exemplar variability is critical in understanding the difference between the current findings and the previous findings. In the current study, exposure to multiple exemplars within a single trial may have allowed RB learners to more quickly discover the relevant relationship among exemplars in a way that is supported by the timing dynamics in the explicit system, which can be learned without feedback and can withstand delays in feedback of up to 10 seconds without impairment (Ashby et al., 1999; Maddox et al., 2003, 2004; Maddox & Ing, 2005).

Additionally, in contrast to II category learning, RB learning is unaffected by grouping and deferring feedback (Smith et al., 2014). In Smith et al. RB category learning outcomes were not different when the feedback was given only after six trials of stimulus presentation and response relative to when feedback was immediately presented after each stimulus-response pair. In the current study, five stimuli were presented after which a single response was made. We found that presenting variability this way (grouping five stimuli together with one response and one instance of feedback) improved RB category learning and generalization performance.

The presentation of within-category variability within a single trial can also be thought of as ‘blocking’ trials by category. Blocked training is thought to be beneficial for learning categories that require attention to within-category similarities (Carvalho & Goldstone, 2014a, 2014b). By grouping the same-category exemplars within a trial, participants in the Variability-Present conditions may have had their attention drawn to the similarities within the categories. Our study demonstrates that this was beneficial for RB, but not II category learning. The category distributions in the current study were not designed to highlight either within-category similarities or between-category differences. It is possible that RB category learning specifically benefits from highlighting within-category similarities through the blocking of same-category
exemplars within one trial. However, in contrast to traditional approaches of blocked training, the current study does not make a comparison between blocked and interleaved training. All of the participants in the current study were trained with trials ‘blocked’ by category. To better connect to the existing literature of blocked and interleaved training, research on the interaction of this training type and II and RB learning is needed.

Further, to understand how participants distinguish between categories during learning, it will be informative for future studies to address the potential strategies that participants were using during learning. Overt category learning experiments often employ some kind of decision modeling to assess strategies used during learning (Ashby & Maddox, 1993). However, the traditional decision bound strategy modeling approach cannot accommodate within-trial variability, so it is not possible to compare the present data with strategies estimated from traditional models. Specifically, these decision bound models assume that participants draw a decision boundary in the two-dimensional space, which can be inferred based on their response to a given stimulus on each trial. Because the current study presented five stimuli on a trial instead of one, we are unable to extend these models to the present results without deep assumptions about how within-trial stimulus information is combined or aggregated. Many strategies (taking the mean, attending to only the last stimulus, etc.) are possible, and they may not be stable across participants or even across trials. At present there is not enough empirical data to guide these choices. However, adapting decision strategy models to accommodate within-trial variability in training will be important for future studies to address to understand how participants’ response strategies change across learning.

The current results expand what is known about how the COVIS framework fits to auditory categorization and expands understanding of the influence of variability during training.
In this regard, there is one particularly intriguing inconsistency between the present data with prior results. Typically, at least in studies without variability manipulations, RB category learning proceeds faster and results in better ultimate performance than II learning; this is thought to reflect the explicit learning system initially driving response output (see Ashby & Maddox, 2011 for review). Here, the Variability-Absent condition is the closest empirical fit to these prior studies. However, RB and II performance were not different in the Variability-Absent condition. RB performance excelled over II performance only when exemplar variability was present within a trial. This discrepancy could be due to the integration, interacting nature of the stimulus dimensions across which exemplars were sampled in the current study (see Roark & Holt, in press). Much more work will need to be done to understand how mechanisms that seem to function in visual category learning according to the COVIS model are involved in auditory category learning. For now, we extend a word of caution for future work examining the COVIS model with auditory stimuli that the dimensions across which exemplars are sampled (Roark & Holt, in press) and the alignment of exemplar variability to task demands can impact the course of category learning.

An additional component of the current study that should be highlighted is that participants in many of the conditions seemed to improve in performance within the first block and with much slower learning throughout the rest of training. Although there was a main effect of block, participants showed much of the overall learning gains within the first block. The length of training here was very short to be able to learn these four complex, probabilistic categories. It is possible that with longer training times, participants would reach higher levels of accuracy and that with this length of training, it is very difficult to learn these categories (see Reetzke, Xie, Llanos, & Chandrasekaran, 2018 for a recent application of this longer training
approach with Mandarin tone categories). It could also be the case that participants were not highly motivated to overcome this plateau with the feedback they were given. These possibilities should be examined in future studies to understand why the learning patterns with these categories emerge.

**Within-trial variability vs. within-category variability**

Examinations of the effect of variability on category learning outcomes have mostly focused on manipulations of the amount of exemplar variability experienced across training (Lively et al., 1993; Logan et al., 1991; Maddox & Filoteo, 2011; Perry et al., 2010). In speech category learning studies, non-native speech produced by multiple speakers across several phonetic contexts leads to learning that generalizes more robustly than less variable speech (Lively et al., 1993; Logan et al., 1991). This has led to a general understanding that variability aids learning and generalization, very broadly speaking. The present results demonstrate that the specific details of how input variability across category exemplars relates to the statistical sampling of categories in the input, and their relationship to response and feedback in training. The current study examined how variability, not in the overall category space, but within a single trial can affect how well categories are learned. There are important distinctions between the present results and those that have led to the general conclusion that variability supports learning.

It is important to recall that the variability experienced across conditions was equivalent in the present experiment. Participants learned from the same category structure and exemplars. Additionally, the categories were complex; each category had many exemplars and exemplars from different categories overlapped in acoustic space. This contrasts with many other studies that use a simplified category environment with relatively few exemplars and little overlap. For instance, in investigating the effect of within-trial variability on learning categories Gabay et al.
(2015) constructed relatively simple categories with only 6 exemplars per category. The current work extends the prior findings to more complex category distributions.

Whereas manipulations of overall category exemplar variability have demonstrated that training with high variability categories leads to poorer and slower initial learning, but better generalization (Bradlow et al., 1999, 1997, Lively et al., 1993, 1994; Logan et al., 1991; Shinohara & Iverson, 2018; Wang et al., 1999; Zhang et al., 2009), manipulating within-trial variability in the present study led to different results. Learning outcomes depended on the nature of the category being learned and whether it required integration of input dimensions and, correspondingly, likely also the learning mechanism involved. RB categories requiring selective attention to a single acoustic input dimension benefitted both in initial learning and generalization from the presence of within-trial variability. In contrast, II categories were unaffected by within-trial exemplar variability in either initial learning or generalization. In all, this suggests that caution is warranted in concluding very broadly that variability improves learning. In fact, the influence of one form of training variability – that experienced within a trial – differs quite substantially depending upon the nature of the category learning challenge. More broadly considering the present results in light of prior research, it is possible to conclude that increases in training set variability interact quite differently with learning mechanisms than adjustments in the relationship of exemplar variability to categorization decision responses and feedback.

While some have suggested that many speech categories may be well described as II categories, it is not clear that all speech categories are best described this way. Further work with speech categories, for which we have less control over the acoustic environment, will be necessary to identify how these kinds of categories function similarly or differently from II and
RB categories. The current study demonstrates that within a nonspeech acoustic environment, there is a stark distinction in the effects of within-trial variability based on whether the category can be defined as an RB or II category.

**Conclusion**

Variability in experience, and specifically in exposure to within-category variability, changes how categories are learned and the representations that result. We investigated category learning in contexts in which overall category variability was equivalent, but variability was differentially associated with the feedback signal. We found that experiencing category variability within a trial benefitted learning and generalization for rule-based categories but did not affect information-integration category learning or generalization. These results may have implications for speech category learning. Specifically, training second language learners on speech categories or other categories that can be described as rule-based categories with high within-trial variability may prove a successful strategy for improving learning and generalization. The same benefit may not exist for speech categories that can be better described as information-integration categories.
References


http://doi.org/10.3389/fpsyg.2014.00825


http://doi.org/10.1021/acschemneuro.5b00094.Serotonin


Learning to treat distinct perceptual experiences as functionally equivalent is vital for perception, action, language and thought. In the auditory domain, we can interpret a child’s squeal as thrilled or terrified, judge a kettle to be boiling from its gurgle, or understand diverse acoustic signals from different talkers to each be the word ‘thanks.’ Auditory categories, whether speech or nonspeech, are most often complex and defined across multiple acoustic dimensions. Often times, these acoustic dimensions can be quite difficult to describe verbally or to attend to selectively (Francis, Baldwin, & Nusbaum, 2000; Grau & Kemler Nelson, 1988; Hillenbrand, Getty, Clark, & Wheeler, 1995).

Recently, an influential theory of category learning that was originally developed to explain visual category learning has been expanded into the auditory domain (Chandrasekaran, Koslov, & Maddox, 2014; Chandrasekaran, Yi, & Maddox, 2014; Maddox, Molis, & Diehl, 2002). Many behavioral, neuropsychological, and neuroimaging studies have provided considerable evidence for the involvement of at least two distinct systems in visual category learning (Ashby & Maddox, 2005, 2011; Morrison, Reber, Bharani, & Paller, 2015; Smith & Grossman, 2008; but see Newell, Dunn, & Kalish, 2011). The Competition between Verbal and Implicit Systems (COVIS) model specifically posits the involvement of an explicit, hypothesis testing system and an implicit, procedural learning system (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). The explicit, hypothesis-testing system is optimal for learning so-called ‘rule-based’ (RB) categories that can be described with verbalizable rules. The kind of rules that are most often used in the literature vary across a single input dimension. However, rules can be
based on multiple dimensions or require a more complex rule than can be easily verbalized. In
the current study, we will investigate RB categories based on one dimension but acknowledge
that rules can be more complex. This explicit system is thought to involve top-down processes
and involves the prefrontal cortex as well as the head of the caudate nucleus in the striatum
(Ashby & Maddox, 2005). The implicit system learns via slower procedural learning
mechanisms and is optimal for learning ‘information-integration’ (II) categories that require
integration across at least two input dimensions. The implicit system is thought to implement
these processes by involving the body and tail of the caudate nucleus in the striatum as well as
the putamen (Ashby & Maddox, 2005).

The expansion of this model into the auditory modality reveals some of the challenges in
applying visual theories to audition (Roark & Holt, 2018). One issue concerns input dimensions.
Whereas most visual category learning studies have examined learning across simple input
dimensions that are easily described verbally, acoustic dimensions like modulation frequency, or
amplitude envelope, or the formant frequencies of speech, may be difficult for untrained listeners
to describe (Francis et al., 2000; Hillenbrand et al., 1995). Additionally, the visual input
dimensions that have been typically used in research tend to be perceptually separable, in that
they are processed independently and are easy to attend to selectively (Garner, 1974). In contrast,
acoustic dimensions are often integral; they are difficult to attend to selectively (Garner, 1974).
Pitch and loudness, for example, are perceived integrally such they are processed in a unitary
fashion (Grau & Kemler Nelson, 1988; Melara & Marks, 1990b). Although both auditory and
visual dimensions can be separable or integral, a challenge in translating categorization models
developed in the visual modality to auditory and speech category learning is that integral,
interacting, and difficult-to-verbalize acoustic dimensions may differ from easily-verbalized
categorization rules across separable dimensions in the visual modality, such as the frequently-used spatial frequency and line orientation dimensions in a Gabor patch (Maddox, Ashby, & Bohil, 2003), or line length and orientation (Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005).

Across both visual and auditory tasks, integral or interacting dimensions have received much less attention than dimensions that are separable and easy to verbalize. Research on integral and separable dimensions demonstrates that dimensions are processed and used differently depending on how they are perceived by participants (Garner, 1976, 1978; Kemler & Smith, 1979). Whereas separable dimensions are easier to attend to selectively and thus, may benefit rule-based category learning, integral dimensions are more difficult to attend to selectively and may be detrimental for rule-based category learning. For instance, learning RB categories is more difficult across the integral dimensions of saturation and brightness than across separable dimensions like circle size and the angle of a radian inside the circle (McKinley & Nosofsky, 1996). However, other researchers found that participants had higher categorization accuracy when learning RB categories than II categories based on the integral dimensions of saturation and brightness (Ell, Ashby, & Hutchinson, 2012). The investigation of II and RB category learning across integral dimensions has been limited and more research is needed to understand how the separable or integral nature of dimensions impacts II or RB category learning.

One study on auditory category learning with the integral dimensions of locations of spectral peaks in frequency space demonstrated that participants learned II categories that required a negative integration in the stimulus space better than a positive integration (Scharinger, Henry, & Obleser, 2013). However, these researchers did not compare this II
category learning with typical RB category learning. Thus, it is not yet fully understood how auditory dimensions that are difficult to attend to selectively impact category learning across different category structures within the same acoustic space. It is necessary to examine acoustic dimensions that are integral and are difficult to verbalize to evaluate the ability of the COVIS perspective to accommodate the complexities of acoustic dimensions. The goal of the current study is to examine how categories within the same two-dimensional acoustic space are learned when selective attention to the dimensions is difficult and the dimensions are not easy to verbalize. This will allow us to investigate the applicability of the dual systems perspective with complex acoustic dimensions that are similar to many complex acoustic dimensions that define auditory categories in real-world contexts, such as speech.

For acoustic dimensions that interact or are difficult to attend to selectively, we may expect that information-integration categories will be learned better than rule-based categories because the dimensions are difficult to separate perceptually. Likewise, the ability to learn different categories within this space may depend on the precise nature of the relationship between the acoustic dimensions and how this relates to internal perceptual representation of the dimensions, which may differ from the acoustic dimensions.

To test these predictions, we trained participants on auditory categories defined across two acoustic dimensions that past research suggests are difficult to attend to selectively (Holt & Lotto, 2006; Roark & Holt, 2018). We examined participants’ accuracy across training and their ability to generalize to novel sounds. Additionally, we applied decision bound computational models to assess participants’ strategy use in category learning and their propensity to integrate or selectively attend to the dimensions in category decisions at each stage of learning.

**Methods**
We investigated two types of information-integration (II) category learning problems and two types of unidimensional rule-based (RB) category learning problems. Each of these category-learning challenges was defined across the same two acoustic input dimensions. Sampling this input space in four different ways allowed us to avoid assumptions about which dimension, or combination of dimensions, would most impact learning. To anticipate, we found that participants tended to integrate across the dimensions, especially in a way that reflected a positive correlation between the dimensions.

Frequency-modulated nonspeech tones served as category exemplars across each of the four category-learning challenges. We used nonspeech stimuli to control as much as possible for participants’ prior experience with this acoustic space. By using nonspeech sounds, we were able to carefully construct artificial categories and match different category exemplar distributions as much as possible. The acoustic input dimensions across which category exemplar distributions were sampled were center, or carrier, frequency (CF) and modulation frequency (MF). In a previous study of auditory category learning using these same dimensions (Holt & Lotto, 2006), participants were able to adjust perceptual weighting across the dimensions on the basis of what was required for the task. However, participants tended to place some weight on each dimension, even when the category learning task required selective attention to a single dimension. In other words, selective attention to these acoustic dimensions is difficult, even when it is required by the task. We chose this particular pair of dimensions because perceptual reliance on the dimensions is malleable and, at the same time, the perceptual representation of these dimensions may not be entirely separable.

**Participants**

A total of 81 adults (38 females, 43 males) ages 18-24 years and affiliated with Carnegie
Mellon University participated for partial course credit. Participants were randomly assigned to one of four conditions defined by the sampling of category exemplars in the acoustic input space. Three participants were excluded due to equipment error, leaving 78 subjects in the final analysis. There were 20 participants in the rule-based-CF (RB<sub>CF</sub>) condition, 19 in the rule-based-MF (RB<sub>MF</sub>) condition, 19 in the information-integration positive slope (II<sub>Positive</sub>), and 20 in the information-integration negative slope (II<sub>Negative</sub>) condition. All participants reported normal hearing.

**Stimuli**

**Sound Exemplars.** The two-dimensional acoustic space from which stimuli were sampled was defined by center, or carrier, frequency (CF) and modulation frequency (MF). As in Holt and Lotto (2006), each stimulus was created from a sine wave tone with a particular CF modulated with a depth of 100 Hz at the corresponding MF. For example, if the CF was 760 Hz and the MF was 203 Hz, the tone was modulated from 710 to 810 Hz at a rate of 203 Hz. Each stimulus was 300 ms long. Exemplars were synthesized in MATLAB and matched for RMS energy.

**Category Distributions.** Two individual category distributions were created to define the category-learning challenge for each of the four conditions (Figure 1). The information-integration conditions sampled acoustic space such that optimal performance would require integration across the two dimensions. The II<sub>Positive</sub> and II<sub>Negative</sub> conditions are mirror images, differing only in the nature of the correlation between CF, shown on the x-axis in Figure 1, and MF, shown on the y-axis in Figure 1. In the case of II<sub>Positive</sub>, higher CF values were associated with higher MF values and, for II<sub>Negative</sub>, higher CF values were associated with lower MF values. The rule-based categories sampled acoustic space such that they could be optimally
differentiated by selectively attending to one of the two stimulus dimensions that define the categories. The RB<sub>CF</sub> condition requires selective attention to the CF dimension. The RB<sub>MF</sub> condition requires selective attention to the MF dimension.

\[ \text{Information-InTEGRATION} \]

\[ \text{Positive} \]

\[ \text{Negative} \]

\[ \text{Rule-Based: Center Frequency} \]

\[ \text{Rule-Based: Modulation Frequency} \]

\[ \text{Figure 1.} \] Stimulus distributions for the four conditions in this study. The black line represents the optimal decision boundary that separates the two categories.

Each category was defined by 100 distinct stimuli sampled from a bivariate normal distribution across the two input dimensions (Table 1). Half of the stimuli from each category were used during training and the other half were reserved for the generalization test. The exemplars defined as training and test were randomly selected, with consistent sampling across participants.
Table 1

**Category Distribution Information**

<table>
<thead>
<tr>
<th>Category</th>
<th>Integration Positive</th>
<th>Integration Negative</th>
<th>Rule-Based Modulation Frequency</th>
<th>Rule-Based Center Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (CF, MF)</td>
<td>Variance (CF, MF)</td>
<td>Covariance</td>
<td></td>
</tr>
<tr>
<td>Category A</td>
<td>(818.3, 225.8)</td>
<td>(6328, 2278.2)</td>
<td>2260.5</td>
<td></td>
</tr>
<tr>
<td>Category B</td>
<td>(913.8, 168.3)</td>
<td>(6328, 2278.2)</td>
<td>2969.9</td>
<td></td>
</tr>
<tr>
<td>Category A</td>
<td>(913.8, 225.8)</td>
<td>(6328, 2278.2)</td>
<td>-2260.5</td>
<td></td>
</tr>
<tr>
<td>Category B</td>
<td>(818.3, 168.3)</td>
<td>(6328, 2278.2)</td>
<td>-2969.9</td>
<td></td>
</tr>
<tr>
<td>Category A</td>
<td>(866, 156.5)</td>
<td>(6770, 796.7)</td>
<td>53.4</td>
<td></td>
</tr>
<tr>
<td>Category B</td>
<td>(866, 237.5)</td>
<td>(6770, 796.7)</td>
<td>-32.3</td>
<td></td>
</tr>
<tr>
<td>Category A</td>
<td>(914, 197)</td>
<td>(1020.2, 3281.8)</td>
<td>-214.5</td>
<td></td>
</tr>
<tr>
<td>Category B</td>
<td>(816, 197)</td>
<td>(1020.2, 3281.8)</td>
<td>-129.1</td>
<td></td>
</tr>
</tbody>
</table>

**Procedure**

Separate groups of listeners participated in each condition. The task was identical across conditions; only the sampling of stimulus distributions varied (Figure 1). Participants were not informed about the nature of the dimensions.

During the training phase, participants completed four blocks of training (96 trials/block; 384 total trials), with a brief break in quiet separating blocks. The trial structure was largely the same across training and generalization phases of the experiment. On each trial, participants heard a single sound exemplar (300 ms) randomly selected from one of the two categories, repeated five times (50 ms silent inter-stimulus interval). Two boxes on the screen indicated response options corresponding to the “u” and “i” keys on a standard keyboard. Participants indicated which of two equally-likely categories the sound belonged to by pressing a response button. A red X indicating the correct category decision appeared in one of the boxes 500 ms after response. Participants were instructed to use this feedback to inform future categorization.
decisions. A 1 sec inter-trial interval followed the feedback.

After completing the training phase, participants completed the generalization test (100 trials). Participants were instructed that they would now be tested on what they learned during training and that there would no longer be feedback. Instead of a red X indicating the correct category decision, question marks appeared inside each of the two boxes on the screen. With the exception of the feedback, the trial structure was identical to the training phase. During the generalization test phase, participants encountered category exemplars (50/category, 100 total) that they had not encountered in training. Thus, the generalization phase measured the ability to generalize category learning to novel exemplars—a hallmark of categorization.

The task was run in a sound-attenuated booth using E-Prime software (Psychology Software Tools, Inc., Sharpsburg, PA), with stimuli presented diotically over Beyer DT-150 headphones at a comfortable listening level.

**Results**

The analyses focused on accuracy of categorization during training and as assessed in the generalization test. Additionally, we fit a series of decision bound models to categorization responses across training in order to examine response strategies across conditions (for more detailed information about model applications see Ashby & Maddox, 1993; Maddox & Ashby, 1993; Maddox & Chandrasekaran, 2014).

**Behavioral Results**

**Normalization.** Although we attempted to equate the stimulus distributions for RB and II for distance between the means and variance, there remained small differences in the overall overlap of the two categories across conditions. An ideal observer would be able to achieve 96% accuracy in the II conditions, 91% in the RB\textsubscript{MF} condition, and 92% in the RB\textsubscript{CF} condition. Thus,
we computed a normalized accuracy score to account for potential cross-condition differences (Normalized Accuracy = Raw Accuracy / Optimal Accuracy). Below, we report only the normalized accuracies to give a more conservative measure of the differences among conditions.

**Training Accuracy.** We measured categorization response accuracy across the four blocks of training to determine the effect of stimulus distribution condition on training performance (Figure 2). A mixed-model ANOVA with block as a within-subjects factor and condition as a between-subjects factor revealed a significant main effect of block ($F(3,222) = 8.53, p < .001, \eta^2_p = .10$), a significant main effect of condition ($F(3,74) = 44.9, p < .001, \eta^2_p = .65$), and no interaction ($F(9,222) = 0.96, p = .47, \eta^2_p = .038$). We outline the results of the analyses for block and condition separately below.

---

2 Compared to the non-normalized data, the only differences were that with the normalized accuracy RB$_{CF}$ was above chance in the first block ($t(19) = 2.788, p = .012$) and there were no statistically significant differences between II$_{Positive}$ and RB$_{MF}$ in the training blocks (whereas with the non-normalized scores, II$_{Positive}$ and RB$_{MF}$ were significantly different in the first and second training blocks).
Figure 2. Block-by-block average normalized accuracy, normalized according to ideal observer accuracy, for all conditions. Ribbon error bars reflect standard error of the mean. Dashed line represents chance accuracy (50%).

Learning across blocks. Bonferroni-corrected post hoc comparisons indicated that the majority of learning for all conditions occurred across the first two blocks. All blocks had significantly higher accuracy than Block 1 (Block 1 vs Block 2 $p = .049$, $d = 0.30$, Block 1 vs Block 3 $p = .004$, $d = 0.40$, Block 1 vs Block 4 $p = .001$, $d = 0.45$), but Blocks 2, 3, and 4 did not differ from one another (all $ps > .05$, $ds < .20$). The majority of learning gains occurred within the first two blocks for all conditions. Learning occurred early and differences among conditions persisted throughout the experiment.
Differences among conditions. Bonferroni-corrected post hoc comparisons revealed that participants learning the II\textsubscript{Positive} categories performed better than participants learning II\textsubscript{Negative} categories across training ($p < .001$, $d = 4.19$). Categorization accuracy was also higher for participants in the RB\textsubscript{MF} condition relative to the RB\textsubscript{CF} condition ($p < .001$, $d = 1.58$). Further, the II\textsubscript{Positive} and RB\textsubscript{MF} categories did not differ in accuracy over the course of training ($p > .99$, $d = 0.37$) and neither did the RB\textsubscript{CF} and II\textsubscript{Negative} ($p = .06$, $d = 1.06$). Participants in the II\textsubscript{Positive} condition had significantly better performance than RB\textsubscript{CF} in all four training blocks ($p < .001$, $d = 2.25$). However, II\textsubscript{Negative} had significantly lower accuracy than RB\textsubscript{MF} ($p < .001$, $d = 2.83$). In our data, training accuracy cannot be easily explained by classifying the category learning challenge as a rule-based or information-integration category distribution. Instead, we found substantial differences between the two II conditions and differences between the two RB conditions, which we examine in greater detail below.

We found very striking differences in learning across the two II conditions, which had identical category distributions, but required integration across the dimensions in opposite directions. In the first block, average accuracy of both II\textsubscript{Positive} and II\textsubscript{Negative} conditions was above chance (II\textsubscript{Positive}: $78.5\%$, $t(18) = 12.4$, $p < .001$, $d = 2.85$; II\textsubscript{Negative}: $54.7\%$, $t(19) = 3.75$, $p = .001$, $d = 0.84$). However, recall that participants in the II\textsubscript{Positive} condition had significantly higher accuracy than participants II\textsubscript{Negative} condition throughout training ($p < .001$, $d = 4.19$). By the end of training in Block 4, average accuracy for each condition was still above chance (II\textsubscript{Positive}: $t(18) = 14.0$, $p < .001$, $d = 3.22$; II\textsubscript{Negative}: $t(19) = 4.23$, $p < .001$, $d = 0.95$), but participants in the II\textsubscript{Positive} condition reached $82.5\%$ accuracy whereas participants in the II\textsubscript{Negative} condition achieved only $56.3\%$ correct. Across training, we found much better performance for participants learning the II distribution that required an integration along the positive axis compared to participants
learning the II distribution that required an integration along the negative axis.

We also found significant differences in performance across the two RB conditions. The primary difference between these two distributions is the dimension that distinguishes the categories. While both RB_{CF} and RB_{MF} performed above chance even in the first block (RB_{CF}: 57.3%, t(19) = 2.79, p = .012, d = 0.62; RB_{MF}: 73.5%, t(18) = 8.36, p < .001, d = 1.92), the RB_{MF} condition outperformed the RB_{CF} condition (p < .001, d = 1.58). This pattern remained throughout training and by Block 4, participants in the RB_{MF} condition reached 80.4% accuracy and participants in the RB_{CF} condition reached only 64.9% accuracy. Block 4 accuracy remained significantly greater than chance for both the RB_{CF} (t(19) = .590, p < .001, d = 1.32) and RB_{MF} conditions (t(18) = 10.3, p < .001, d = 2.36). Whereas participants in both RB conditions were performing above chance throughout training, those learning categories that required a distinction based on MF outperformed those learning categories that required a distinction based on CF throughout training.

**Generalization Test Accuracy.** After training, participants engaged in a generalization test that involved categorizing novel sound exemplars draw from the distributions experienced in training without feedback (Figure 3). Participants in all four conditions exhibited generalization performance greater than chance (50%) accuracy, indicating category learning (II_{Positive} t(18) = 16.5, p < .001, d = 3.80; II_{Negative} t(19) = 3.38, p = .003, d = 0.76; RB_{CF} t(19) = 6.72, p < .001, d = 1.50; RB_{MF} t(18) = 10.8, p < .001, d = 2.47). Generalization test accuracy varied across conditions (F(3,74) = 27.20, p < .001, η^2_p = .52). Participants in the II_{Positive} condition accurately categorized novel sound exemplars 81.4% on average; participants in the II_{Negative} condition reached 55.4%, participants in the RB_{MF} condition reached 80.2%, and participants the RB_{CF} condition reached 68.8% correct. According to Bonferroni-corrected post hoc comparisons, the
overall pattern of generalization mirrors the patterns of learning during training: \( \Pi_{\text{Positive}} \) generalization accuracy was greater than \( \Pi_{\text{Negative}} \) \((p < .001, d = 3.38)\) and \( \text{RB}_{\text{MF}} \) was greater than \( \text{RB}_{\text{CF}} \) \((p = .006, d = 0.93)\). \( \Pi_{\text{Positive}} \) and \( \text{RB}_{\text{MF}} \) had statistically equivalent performance \((p = .74, d = 0.11)\), but \( \Pi_{\text{Negative}} \) was significantly worse than \( \text{RB}_{\text{MF}} \) \((p < .001, d = 2.48)\). \( \text{RB}_{\text{CF}} \) generalization performance was significantly worse than \( \Pi_{\text{Positive}} \) \((p = .002, d = 1.19)\). In general, generalization performance patterned with relative performance across conditions in training. The only difference in the overall pattern of results compared to training is that in the generalization test, participants in the \( \text{RB}_{\text{CF}} \) condition performed significantly better than participants in the \( \Pi_{\text{Negative}} \) condition \((p = .001, d = 1.32)\).

**Figure 3.** Average generalization test accuracy, normalized based on optimal accuracy for each condition. Dashed line represents chance accuracy (50%). Error bars reflect standard error of the mean around the black dot which represents the mean. Each individual point is an individual participant’s average accuracy.
Computational Modeling

Rationale. Categorization accuracy across training and in the generalization test provides a relatively coarse measure of performance that does not reveal why differences between the II conditions and between the RB conditions persist. To obtain a better understanding of what participants learned over the course of this experiment, we applied and fit decision bound models to each block of each participant’s data (Ashby, 1992a; Ashby & Maddox, 1992, 1993; Maddox & Ashby, 1993). Decision bound models are derived from General Recognition Theory (GRT, Ashby & Townsend, 1986), a multivariate application of signal detection theory (e.g. Green & Swets, 1966). These models have been applied extensively in the dual systems literature with both auditory and visual categories (e.g. Ashby & Maddox, 2005, 2011; Chandrasekaran, Yi, et al., 2014; Maddox, Chandrasekaran, Smayda, & Yi, 2013; Scharinger, Henry, & Obleser, 2013). We provide a brief description of the models applied to the data; more specific details of these models, including the proposed neural instantiation of the models, can be found elsewhere (Ashby & Maddox, 1993; Ashby, Paul, & Maddox, 2011; Maddox & Ashby, 1993; Maddox & Chandrasekaran, 2014).

Model Details. Each model assumes participants create decision boundaries to separate the stimuli into two categories. Our model-based approach involves applying four classes of models, with multiple instantiations possible within a class. We fit a unidimensional model based on decision bounds across the CF dimension ($UD_{CF}$), a unidimensional model based on decision bounds across the MF dimension ($UD_{MF}$), an integration model (GLC) with decision bounds based on both CF and MF dimensions, and a random responder model (RR).

Unidimensional Rule-Based Models. Two unidimensional models instantiate a unidimensional decisional bound that is optimal for either the $RB_{CF}$ or $RB_{MF}$ conditions. The
unidimensional model has two free parameters—the decision boundary (vertical (90°) for \( UD_{CF} \) and horizontal (0°) for \( UD_{MF} \)) and the variance of noise (both perceptual and criterial). An example of a unidimensional rule based on CF might be, “If the tone’s CF is greater than 866 Hz, it belongs to category A; if it less than 866 Hz, it belongs to category B.” Optimal performance in the \( RB_{CF} \) requires a \( UD_{CF} \) decision bound whereas \( RB_{MF} \) requires a \( UD_{MF} \) decision bound.

**Information-Integration Model.** The general linear classifier (GLC) also assumes a linear decision boundary but, in contrast to the unidimensional rule-based models, it requires linear integration of the CF and MF dimensions and is therefore optimal for the \( II_{Positive} \) and \( II_{Negative} \) conditions. For the \( II_{Positive} \) condition, the optimal decision boundary has a positive slope (45°) whereas for the \( II_{Negative} \) condition, the optimal decision boundary has a negative slope (-45°). The specific weight a listener places on one dimension can vary, even when fit by the same GLC model. Thus, we also examine the angle of the decision boundaries in the CFxMF input space as an estimate of the perceptual weight of CF versus MF in categorization decisions. The model has three free parameters: the slope and intercept of the decision boundary and the variance of noise (perceptual and criterial).

**Random Responder Model.** The random responder model assumes that the participant guesses on each trial.

**Model Fitting.** For each of the four experimental conditions, we fit the models separately to each participant’s data from each of the four training blocks and the generalization test. The model parameters were estimated using a maximum likelihood procedure (Ashby, 1992b; Wickens, 1982) and the goodness-of-fit statistic was Akaike’s information criterion (AIC) = \( 2r - 2\ln L \) where \( r \) is the number of free parameters and \( L \) is the likelihood of the model given the data (Akaike, 1974). The AIC allows comparison of model fits because it penalizes a model for extra
free parameters such that the smaller the AIC, the closer the model is to the “true” model, regardless of the number of free parameters. To find the best-fit model, we computed AIC values for each model and chose the model associated with the smallest AIC value. We separately replicated the model fit analyses using the Bayesian information criterion (BIC) as the model selection criterion, which gives steeper penalties for extra free parameters. The qualitative pattern of results was not different with the AIC and BIC model fits and so we focus report results based on AIC selection criterion.

**Modeling Results.** To better understand the pattern of learning across the different conditions, we examined the proportion of participants best fit by each computational model (Figure 4) and a more detailed measure of the boundaries participants drew between the categories in the generalization test (Figures 5 and 6).
Figure 4. Proportion of participants fit by each modeling strategy across all four training blocks and the generalization test. None of the participants were best fit by the Random Responder model, so it is not included in the graph.

Proportion of participants using each strategy. Figure 4 shows the proportion of participants whose categorization decisions were best fit by the information-integration (GLC), \( \text{UD}_{\text{CF}} \), \( \text{UD}_{\text{MF}} \), and RR model, separately for each condition. Note that none of the participants in our study were best fit by the RR model in any block. We found that the strategy participants used in the first block was not independent from condition \( \chi^2(6, N = 78) = 17.4, p = .008 \). We
also examined proportions of participants using different strategies in the generalization test block and found the relation between strategy and condition was significant ($X^2(6, N = 78) = 17.4, p = .008$).

A majority of participants in the II$^{\text{Positive}}$ condition (78.9%) and RB$^{\text{MF}}$ condition (78.9%) were best fit by the integration strategy in the first block. The tendency to integrate across the dimensions for the II$^{\text{Positive}}$ and RB$^{\text{MF}}$ conditions emerged early on and persisted throughout training. In Block 4, 100% of participants in the II$^{\text{Positive}}$ condition and 68.4% of participants in the RB$^{\text{MF}}$ condition were best fit by the integration strategy. Integration was a successful strategy for these participants, as learning was most robust in these conditions. It is of note that this was a successful strategy for RB$^{\text{MF}}$ participants, even though integration was a suboptimal strategy for this RB category type. Even though many participants in the RB$^{\text{MF}}$ condition used an integration strategy rather than an optimal unidimensional strategy, their accuracy was still quite high.

In contrast, more participants in the RB$^{\text{CF}}$ condition were best fit by one of the unidimensional (70%), compared to the integration (30%), strategies in the first block (Bonferroni-corrected comparison, $p < .05$). However, many participants were fit by the suboptimal UD$^{\text{MF}}$ strategy, indicating reliance on the MF dimension that is poorly diagnostic of category membership in this condition. This helps to account for the poor categorization accuracy observed in training and generalization in the RB$^{\text{CF}}$ condition.

We also found that, in the first block across the II$^{\text{Negative}}$ condition, more participants’ categorization responses were fit by a unidimensional strategy (60%) compared to an integration strategy (40%) (Bonferroni-corrected comparison, $p < .05$). Over the course of learning, participants in this II$^{\text{Negative}}$ condition were most often fit by a unidimensional strategy, rather than the optimal integration strategy. This is in sharp contrast to much greater adherence to the
optimal integration strategy among participants in the $I_{\text{Positive}}$ condition, for which the only difference from the $I_{\text{Negative}}$ condition was the angle of the decision bound through MFxCF input space.

*Decision boundaries in the generalization test.* Examining the proportion of participants in each condition best fit by each of the models gives us a general sense of participants’ category decision strategies in learning and generalization. However, only examining the proportion of participants using a given strategy does not present a full picture because many decision boundaries are possible within each class. Depending on how well a participant’s decision boundary matches the optimal decision boundary, even within a class, there can be different effects on categorization accuracy. Figure 5 shows the individual best-fit decision boundaries for each participant according to condition in the generalization test. The dashed line in each panel represents the optimal decision boundary.

By the generalization test phase, many participants in each condition were best fit by the integration model. This constituted the majority of participants for the RB$_{\text{MF}}$ and $I_{\text{Positive}}$ conditions (68% and 95%, respectively), and also around 40% of the participants in the RB$_{\text{CF}}$ and 45% of participants in the $I_{\text{Negative}}$ conditions. Visual inspection of the individual decision boundaries in Figure 5 demonstrates that when participants were best fit by an integration strategy, it was along the positive axis, with the exception of a single participant in the $I_{\text{Negative}}$ condition. In the generalization test, there is a bias to integrate along these two dimensions in the positive direction, rather than selectively attending to either dimension (even when that is optimal for category learning) or integrating in the negative direction (even when that is optimal for category learning).
Figure 5. Individual decision boundaries for each participant in generalization test (after all training blocks). The optimal decision boundary for each category is shown as the red dotted line on each plot. The x-axis represents the Center Frequency dimension, and the y-axis represents the Modulation Frequency dimension.

There are especially stark differences between the two II conditions. These two category types differ only in the direction of integration along the dimensions that is required by the category boundary. Nearly every participant in the II_{Positive} condition used a nearly optimal decision boundary. In contrast, participants in the II_{Negative} condition used a mixture of strategies in the generalization test. Among those using the integration strategy in the II_{Negative} condition, all but one participant used a decision boundary with a positive slope between CF and MF, rather
than the optimal negative slope. Thus, even participants best fit by the so-called optimal strategy (as in Figure 4) were not optimally integrating across the two dimensions. This was especially true in the II\text{Negative} condition.

To better quantify the relative weight that participants placed on each dimension in the different conditions during the generalization test, we also computed the angle of the decision boundary for each participant. We compared the individual angle values to the optimal angle of the decision boundary for each of the conditions (II\text{Positive} = 45 degrees, II\text{Negative} = -45 degrees, RB\text{MF} = 0 degrees, RB\text{CF} = 90 degrees). Figure 6 shows the absolute values of these differences for each participant in each condition. The closer the participant’s decision boundary angle is to the optimal decision boundary angle, the more they were optimally attending to the dimensions appropriate for the categories they were learning. This visualization helps to better understand how participants using the integration strategy differently weighted the two input dimensions in categorization decisions and provides more fine-grained information to quantify how close to optimal participants’ strategies were in the generalization test.

We found that the vast majority of participants in the II\text{Positive} condition had a decision boundary with an angle very close to the optimal decision boundary angle (median difference from optimal is 11.0 degrees). In contrast, participants in the II\text{Negative} condition were very far from optimal (median difference from optimal is 91.9 degrees). Thus, it may not be surprising that participants in the II\text{Negative} condition performed much worse than participants in the II\text{Positive} condition. Participants in the II\text{Negative} condition were less able to find the optimal integration strategy and even when they were best fit by the integration model, they were applying a decision boundary along the positive slope, opposite what is optimal for the category distributions. Thus, integration alone is not enough for successful categorization and
generalization—participants must integrate optimally.

**Figure 6.** Box plots of absolute value difference in participants’ best fit decision bound angles relative to the optimal decision boundary. The optimal decision boundary angle is listed for each condition next to its name and is represented by the dashed line at 0. Each dot represents an individual participant value.

In examining the differences between the two RB conditions, approximately equal numbers of participants in the RB\textsubscript{MF} and RB\textsubscript{CF} conditions had decision boundary angles close to the optimal decision boundary angle (Figure 6). The median absolute difference from the optimal angle for RB\textsubscript{MF} was 31.0 degrees and for RB\textsubscript{CF} was 25.0 degrees. Just looking at the difference
between these decision boundary angles and what is optimal, it is not clear why RB$_{MF}$ participants would outperform RB$_{CF}$ by such a large margin. However, it is clear from looking at the actual decision boundaries in Figure 5 that the decision boundaries for participants using the optimal strategy (unidimensional CF for RB$_{CF}$ and unidimensional MF for RB$_{MF}$) are not fully overlapping with the optimal decision boundary. Thus, it is not the angle that is suboptimal for the RB$_{CF}$ participants, but the placement of the decision boundary on the x axis. Participants in the RB$_{CF}$ condition were unable to place the decision boundary at the optimal position along the CF dimension, even if they were reliant on CF for their decision. So even though many RB$_{MF}$ participants integrated across dimensions and had a decision boundary angle that was far from optimal, it resulted in better category learning than sub-optimally placed unidimensional boundaries.

Across all participants, there was a significant correlation between the absolute difference in the decision boundary angle relative to optimal and the generalization test accuracy ($r = -0.62$, $t(76) = -6.97$, $p < .0001$). However, in examining this correlation within each condition, the correlation was only significant for the II$_{Positive}$ ($r = -0.64$, $t(17) = -3.42$, $p = 0.003$) and RB$_{MF}$ conditions ($r = -0.88$, $t(17) = -7.47$, $p < .0001$). There was no significant correlation for the II$_{Negative}$ ($r = -0.05$, $t(18) = -0.21$, $p = 0.83$) or RB$_{CF}$ conditions ($r = -0.16$, $t(18) = -0.70$, $p = 0.49$). Across the entire group of participants, decision boundaries closer to optimal were associated with higher categorization accuracy.

**Discussion**

We examined learning outcomes and participant strategies for auditory categories defined by dimensions that are difficult to attend to selectively (Holt & Lotto, 2006). The results emphasize the importance of considering not just the physical, acoustic, dimensions that define a
categorization challenge, but also the way that acoustic dimensions are represented perceptually. The nature of dimensions’ perceptual representation greatly affects how categories are learned. Dual systems accounts, developed for visual categorization, have emphasized the importance of the sampling of category exemplars from a stimulus space on category learning outcomes. Our results demonstrate that the dimensions defining a stimulus space also play a fundamental role. We cannot assume that a particular sampling from a physical stimulus space maps linearly to a psychological or perceptual space. Few models of category learning address how prior experience shapes representations and how existing representations constrain category learning (but see models of infant and second language speech category learning: Best, McRoberts, & Sithole, 1988; Kuhl, 1991). Instead, it is common to assume that participants will be able to conquer each learning challenge placed in front of them, shifting attentional weights or decision boundaries flexibly based on the requirements of the task. In the current study, even when sampling was equated across the II\textsubscript{Negative} and II\textsubscript{Positive} conditions, participants demonstrated strikingly different learning outcomes and across the two RB conditions; participants were not easily able to disengage from the irrelevant dimension. **Integration strategies persisted even when they were suboptimal**

We found that many participants integrated across the CF and MF dimensions even early in training and even when integration was suboptimal. Prior studies using separable dimensions have reported that participants demonstrate a tendency to selectively attend to the dimensions (e.g., Ashby et al., 1999; Huang-Pollock, Maddox, & Karalunas, 2011; Smith, Beran, Crossley, Boomer, & Ashby, 2010). In the current study, integration across the input dimensions could be described as the ‘default’ strategy for participants. For the acoustic dimensions of the present study, integration emerged early on and produced the best outcomes in terms of categorization.
accuracy even when integration was not the ‘optimal’ strategy predicted by the categorization challenge defined by the sampling of exemplars across the input dimensions.

Strikingly, the bias toward integration was present even for the two RB conditions, for which unidimensional strategies were optimal based on the sampling of stimuli from the input space. The pattern of strategies revealed by computational modeling indicates that participants in the RB conditions did not easily disengage from the irrelevant dimension, even when training feedback did not align with this strategy. This pattern is not typically observed in the existing research on the dual systems theory, for which dimensions defining RB categories have tended to be separable (Ashby & Maddox, 1990, 2005, 2011; Chandrasekaran, Koslov, et al., 2014; Goudbeek, Cutler, & Smits, 2008; Goudbeek, Swingley, & Smits, 2009; Maddox & Ashby, 2004).

Because many auditory dimensions may be integral and difficult to selectively attend to, it is important to examine category learning where it is difficult or impossible to engage with only the relevant dimension during category learning. Many models of category learning, including the COVIS model, posit that participants shift their attention weights or selectively attend to individual input dimensions during learning (Ashby et al., 1998; Nosofsky, 1986). In the current study, many participants in the two RB conditions were able to categorize the exemplars reasonably well even without selectively attending to the dimensions. The ability to selectively attend to the dimensions was not required for above chance performance in this task.

There are few studies comparing two RB category learning challenges that differ only on the dimension to which selective attention is required (for two exceptions with visual categories see: Ell et al., 2012; Maddox & Dodd, 2003). Instead, experimenters typically choose one of the dimensions and use it as the single representative RB category learning challenge. We
investigated two possible RB categories to avoid the assumption that participants may be equally likely to learn RB categories based on either dimension.

This proved to be informative. Learners in the RB_{MF} condition outperformed learners in the RB_{CF} condition throughout the entire experiment. We did not expect that the RB_{CF} and RB_{MF} conditions would demonstrate such strikingly different performance. In Holt and Lotto (2006), participants placed more perceptual weight on the CF dimension when either CF or MF alone could distinguish the categories. However, even when participants placed more weight on CF, they continued to rely upon both dimensions, suggesting it may be difficult to selectively attend to these dimensions.

Several other studies have found differences in the reliance upon various input dimensions during category learning. Although these studies differ in their details, they collectively demonstrate that some input dimensions are more likely to be relied upon in category learning than others (Ell, Ashby, & Hutchinson, 2012; Goudbeek et al., 2008, 2009; Holt & Lotto, 2006; Maddox & Dodd, 2003; Scharinger et al., 2013). In the current experiment, participants did not demonstrate a bias for one dimension over the other. Instead, participants demonstrated a bias to *integrate* across the two dimensions.

In the current experiment, participants were better able to learn the RB_{MF} categories than the RB_{CF} categories. Participants in the RB_{MF} condition used integration strategies often and performed well, even though these strategies are sub-optimal. Participants in the RB_{CF} conditions used the optimal selective attention more frequently but applied these strategies sub-optimally. Thus, even though RB_{MF} participants were technically using suboptimal decision strategies, the actual decision boundaries they were placing were closer to optimal than the optimal strategy decision boundaries that RB_{CF} participants were placing. This finding is counterintuitive, but
consistent in that it further demonstrates how difficult it is to optimally selectively attend to these acoustic dimensions.

A unique feature of the current study is that we examined learning of distinct category structures within the same acoustic space defined by dimensions that are difficult to attend to selectively. In a study investigating visual category learning with the integral dimensions of saturation and brightness, researchers have demonstrated integration across the integral dimensions, but generally RB categorization performance was better than II performance (Ell et al., 2012). In a study of auditory category learning with perceptually integral dimensions that are closely related to pitch and timbre, researchers found that participants strongly relied upon unidimensional strategies compared to integration strategies (Scharinger et al., 2013). In contrast, in the current experiment, we found that participants were more often integrating across the dimensions and there was no clear benefit for II categories over RB categories.

The present results highlight that it is important to consider the role of perceptual dimensions in category learning. Whereas separable, relatively easy to verbalize acoustic dimensions, such as pitch and duration, may behave in a manner aligned with the COVIS model (Chandrasekaran, Koslov, et al., 2014; Goudbeek et al., 2008, 2009), this may not be the case for integral, interacting, or difficult to verbalize acoustic dimensions. Nonetheless, it is important to acknowledge that the current experiment investigated two specific acoustic dimensions; the observed pattern of results may not be true for all acoustic dimensions. Especially in light of the conflict between the current results and those of Scharinger et al. (2013), other interacting and integral acoustic dimensions should be examined. It is important to investigate learning categories defined by complex and interacting acoustic dimensions because many acoustic dimensions, including those defining speech categories, interact (Francis et al., 2000; Grau &
Kemler Nelson, 1988; Hillenbrand et al., 1995). The present results highlight the need for caution in assuming the psychological relationship among perceptual dimensions involved in category learning.

There was a bias to integrate across the dimensions in a way that reflected a positive correlation between the dimensions.

Not only did the participants demonstrate a propensity to integrate across the dimensions in the current study, they did so in a manner that reflected a positive correlation between CF and MF. This propensity to integrate along the positive correlation axis had a particularly potent impact on the two II conditions. Whereas category learning was robust in the $II_{\text{Positive}}$ condition, it hovered near chance in the $II_{\text{Negative}}$ condition. This is notable inasmuch as the statistical sampling of the acoustic input space was identical, except that as the category boundary was rotated 90 degrees.

The nature of this interaction may stem from the physical dimensions and participants’ prior experience. Although we used nonspeech stimuli to attempt to better control the acoustic environment and minimize participants’ prior experience with category exemplars, it still could be the case that existing representations for these acoustic dimensions, and their relationship, influenced category learning. Listeners are sensitive to statistical correlations among acoustic dimensions (Holt & Lotto, 2006; Liu & Holt, 2015; McMurray, Aslin, & Toscano, 2009; Stilp, Rogers, & Kluender, 2010; Wade & Holt, 2005). It is possible that the positive integration bias we found may be due to a natural correlation between these two physical dimensions. For instance, it might be the case that due to the physics of sound, when a sound has a higher CF, more modulations can be added. This interrelation between the dimensions may lead to a natural positive correlation between CF and MF and may have contributed to the bias to integrate along
the positive axis that we observed here. The current study was not designed to address this
particular relationship between the two dimensions, but our results do point to the need to clarify
potentially pre-existing relationships between perceptual dimensions and how they might
influence learning.

Additionally, while our nomenclature of positive and negative integration refers to the
direction of the decision boundary, the perceptual distinction that is made between the categories
is orthogonal to the decision boundary. For the \( \text{II}_{\text{Positive}} \) category, the boundary has a positive
slope, but the two categories require distinctions between categories across this boundary, which
would be along the negative axis. This is an important distinction to make to understand how the
perceptual information is used and processed by learners. This also further highlights the
importance of considering the underlying perceptual representation of the categories being
learned in addition to the decision boundary in the perceptual space.

The propensity to integrate along the positive axis observed in the current study
specifically benefitted learning of \( \text{II}_{\text{Positive}} \) categories and impaired learning of \( \text{II}_{\text{Negative}} \) categories.
Very few studies have specifically compared learning of two different information-integration
conditions (but see Ell et al., 2012; Scharinger et al., 2013). Scharinger et al. (2013) found that
participants were better able to learn II categories with a negative correlation than a positive
correlation. They argued that performance in that condition was better because the negative
correlation better matched the natural correlation found in speech. The dimensions in the current
study are not directly comparable to speech dimensions. However, our results support Scharinger
et al.’s (2013) conclusion that prior experience with similar or identical dimensions influences
category learning behavior. It could be the case that either prior experience or general familiarity
with these or similar acoustic dimensions could be driving perceptual processing and learning.
In contrast, Ell et al. (2012) found that for the integral visual dimensions of saturation and brightness, the positive and negative axis II categories were learned equally well. They did not observe a difference in their two II conditions because participants demonstrated a bias toward strategies that required selective attention to brightness, while ignoring saturation. Because there was no strong tendency to integrate across dimensions, Ell et al. (2012) did not find differences in their two II conditions.

Many acoustic dimensions demonstrate similar interacting relationships, such as pitch and loudness and pitch and timbre (Melara & Marks, 1990b; Neuhoff, 2004). Thus, investigating how categories defined by complex and interacting dimensions are learned is an important and understudied area of category learning. Instead, typical experiments—for the sake of simplicity—define their categories based on simple, separable dimensions. Use of these simple, separable dimensions may lead to better experimental control, but it comes at a cost to generalizability to real-world dimensions, which are often more complicated.

Additionally, it is possible that individual differences in musical expertise, language experience, or hearing loss may affect how participants interact with acoustic information during learning. Because none of these factors were directly tied to our main question of interest, we did not collect these measures from participants and are thus unable to evaluate whether and which individual differences had any effect on performance. Future studies may be directed at understanding the factors of individual participants that may lead to differences in perceptual processing or category learning in general.

**Implications for models of category learning**

Our results demonstrate that the nature of perceptual dimensions, in terms of their perceptual interaction or non-separability, impacts category learning. The influence of
dimensions was apparent in the course of learning, the strategies participants applied in learning, and in generalization of learning.

The current study demonstrates that for the COVIS dual systems approach to be sufficiently expanded into the auditory domain, the role and influence of integral or interacting dimensions must be taken into account. It will be important to consider not only sampling across input dimensions, but also the mapping of these dimensions to perceptual dimensions. Acoustic dimensions, including those important for speech, are often highly integral and often are not easily verbalizable (Garner, 1974; Grau & Kemler Nelson, 1988; Melara & Marks, 1990b). Speech categories are highly multidimensional, with dimensions that are difficult to describe verbally and are often perceptually integral (Francis et al., 2000; Grau & Kemler Nelson, 1988; Hillenbrand et al., 1995). Researchers have demonstrated that many acoustic dimensions that contribute to speech perception are perceptually integral (Kingston, Diehl, Kirk, & Castleman, 2008; Macmillan, Kingston, Thorburn, Walsh Dickey, & Bartels, 1999). This perceptual integrality of speech dimensions may be driven by physical constraints on articulatory mechanisms that renders acoustic input dimensions to be inter-dependent (Carré, 2009) in the manner of the ‘natural covariation’ that we posit to explain the relationship between CF and MF in the current study. Alternatively, this integrality may also be a consequence of some psychoacoustic similarity between the dimensions (Diehl, 2008; Kingston et al., 2008; Macmillan et al., 1999). The issue of how and why acoustic dimensions covary has been engaged with in speech perception and has produced strong debates about whether this covariation arises from environmental covariation or from similar underlying auditory properties (Diehl & Kluender, 1989; Fowler, 1989). While the dimensions in the current study are not speech dimensions, the theoretical issue is quite similar and general theories of auditory categorization
might benefit from engaging with the evidence from speech.

Although we set out to test predictions of the dual systems perspective with auditory dimensions that are difficult to selectively attend to, the most meaningful patterns in the data were not between II and RB categories. Instead, stark differences within each category type (II\text{Positive} vs. II\text{Negative}, RB\text{MF} vs. RB\text{CF}) emerged that are not easily explained by existing dual systems frameworks. Our results indicate that understanding category learning requires understanding how the dimensions that define the space in which categories are situated are represented perceptually. The present results caution that the categories laid out on the page as orthogonal dimensions (acoustic categories, in this case), may not align with traditional conceptualizations of RB and II categories. Instead, the defining factor in determining whether categories are better described as II or RB relies on the perceptual, not physical space. A potential consequence of this is that category exemplar distributions that may appear to be RB (or II) learning challenges in the physical space may not truly be RB (or II) problems after taking into account representations in the perceptual space.

It is important to note that category learning theories apart from COVIS have addressed the attention to dimensions or how participants weight different cues during learning (Francis & Nusbaum, 2002; Goldstone, 1993, 1994; Nosofsky, 1986). However, these models still assume that learners are generally flexible and adaptable and can learn to shift attention within the input space to respond to the demands of the current category learning challenge. Additionally, many of these models do not propose a neurobiologically plausible account of how attention to dimensions interacts with category learning. Future research will need to address how to integrate theories to advance understanding of the interplay of perceptual encoding, attention, and learning involved in acquiring new categories.
Conclusion

The dimensions that define categories affect the ability to learn those categories. Here, participants demonstrated a bias to integrate across acoustic dimensions in a way that reflected a positive relationship between dimensions. This led to high accuracy for II categories requiring positive integration but was detrimental for learning statistically-equivalent sampling of II category exemplars that required negative integration across dimensions. Participants often integrated along the dimensions even when this strategy was suboptimal for learning, in the case of RB categories. These suboptimal integration strategies were not detrimental for learning in the RB\textsubscript{MF} condition. However, learning in the RB\textsubscript{CF} condition was worse than in the RB\textsubscript{MF} condition. Thus, the dimensions used to define categories and the relationship between those dimensions greatly affected participants’ category learning performance, the strategies they used during learning, and their ability to generalize category learning to novel exemplars. The interaction of dimensions in experience and perception impacts category learning in a way that is currently unexplained by the existing COVIS dual systems framework and other models of category learning. Thus, we caution for the need to consider that the input space is not necessarily homologous with the perceptual space as this has important effects on category learning.
References


Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated


Chapter 5: Limits to the effects of short-term statistical learning and efficient coding

The natural world has structure. For example, rain is nearly always accompanied by dark clouds. It is useful to learn this kind of structure because it keeps us from getting caught outside without an umbrella. Structure also exists in sensory information. Categories of speech sounds share acoustic attributes; categories of animals share visual features. Perceptual systems are sensitive to this structure. There is evidence for ‘efficient coding’ of these regularities in the sense that perceptual systems appear to non-redundantly code information such that they reflect natural signal statistics (Barlow, 1961; Schwartz & Simoncelli, 2001). For instance, distinct neural populations encode information for spatial frequency and line orientation, which maximizes the system’s ability to rapidly process visual input. This efficiency in the encoding of information allows humans to flexibly interact with a seemingly stable world.

Efficient coding is often discussed and studied as it relates to long-term regularities in the natural world (e.g. distributions of frequency of natural sounds, independent visual dimensions such as spatial frequency and orientation). There is evidence that sensory systems efficiently encode long-term natural signal statistics in vision (Simoncelli, 2003; Simoncelli & Olshausen, 2001), audition (Kluender, Stilp, & Kiefte, 2013; Lewicki, 2002; Ming & Holt, 2009; Smith & Lewicki, 2006; Stilp & Lewicki, 2014; Wang, 2007), and multisensory integration (Ernst & Banks, 2002).

Separately, there is also a discussion of whether perceptual systems can be described as ‘efficient’ when short-term conditions deviate from long-term norms, requiring the system to adapt. One prominent example of this phenomenon is the McCollough effect (McCollough, 1965).
In long-term visual experience, there is a zero correlation between the color of objects and orientation of lines. The demonstration of the McCollough effect exploits this structure in the visual world. Instead of this typical zero correlation between color and line orientation, to induce the McCollough effect, one experiences a novel contingency between these two visual dimensions. For example, researchers show a red background with vertical black lines and a green background with horizontal black lines. In this environment, color and orientation of lines are perfectly correlated. After relatively brief exposure, the perceptual system adapts to this novel regularity of the correlation of these two, typically independent, visual features (Dodwell & Humphrey, 1990; Humphrey, 1998). Thus, the McCollough effect occurs—the perceiver is presented with horizontal and vertical black lines with no color and instead of seeing no color, they perceive the opposite color to which they were exposed appearing on the different orientations of lines. After exposure to a red background with vertical lines and a green background with horizontal lines, now vertical lines seem to have a slightly green background and horizontal lines seem to have a slightly red background. Even brief exposure to this novel statistical regularity that violates the statistics of the natural visual world—i.e. a correlation between color and line orientation—causes the system to adapt and briefly reflect this novel regularity.

The occurrence of the McCollough effect demonstrates that the perceptual system is sensitive to the violation of long-term statistical structure in a short-term experience. This kind of rapid ‘efficient coding’ or statistical learning has been investigated in several literatures which have remained generally separate. Three components unite investigations of this kind of learning or adaptation. That is the rapid learning of novel regularities, even in a passive manner.
Rapid adaptation to short-term statistical structure has also been examined in the auditory realm. Researchers have demonstrated that the perceptual system is able to rapidly adapt to short-term statistical structure through passive exposure (Stilp, Rogers, & Kluender, 2010). Stilp et al. (2010) examined the effects of brief passive exposure to correlations between two relatively separable acoustic dimensions—attack-decay and spectral shape. After brief exposure, their participants were better able to discriminate sounds that were consistent with the correlation they experienced and temporarily lost the ability to discriminate sounds that were orthogonal to the correlation they experienced. Stilp et al. (2010) interpreted these effects as rapid encoding of this correlation creating a novel perceptual dimension that best reflects the experienced input regularities. These correlations are then one kind of statistical structure that can rapidly shift perception in a way that changes discrimination behavior.

Indeed, apart from discussion of efficient coding, there is a great deal of evidence that perception is sensitive to regularities in short-term and long-term perceptual input. A separate literature has investigated many kinds of statistical structure than humans are sensitive to including the distributional statistics of perceptual information. Adults and infants, as well as other species, are sensitive to statistical distributions of acoustic information (Aslin, Saffran, & Newport, 1998; Escudero & Williams, 2014; Maye, Werker, & Gerken, 2002; McMurray, Aslin, & Toscano, 2009; Toscano & McMurray, 2010; Wanrooij & Boersma, 2013). Exposure to distinct distributions of sensory information can change how that information is perceived. For instance, infants and adults can better differentiate sounds after passive exposure to a continuum of sounds that is bimodally distributed, compared to when they experience sounds from a unimodal distribution (Escudero & Williams, 2014; Maye et al., 2002; Wanrooij & Boersma, 2013). This general sensitivity to statistical information is thought to be a fundamental
component that contributes to how humans organize knowledge, including knowledge about complex object and speech categories.

Although there is some evidence of ‘efficient’ coding in the short-term to input regularities and a separate literature has made connections with passive exposure to input distributions as a basis for category learning, the connections are relatively tenuous and have not been examined thoroughly. Specifically, it is not well understood how representations of the sensory world change with passive exposure to statistical regularities and, further, how exposure to these regularities might alter broader cognitive processes and behavior, for instance in novel category learning.

The current study will examine how the similarity-based representations of an acoustic space are altered as a result of passive experience with a novel statistical correlation between two acoustic dimensions (Chapter 5). This will provide a more detailed understanding about the changes that do or do not occur to mental representations of an acoustic space with exposure to novel regularities. This will also illuminate details about the plasticity of similarity-based representations after short-term, highly structured experience. In the next chapter, we will investigate how exposure to correlations between two acoustic dimensions affects cognitive processes during category learning to better understand the wide-reaching effects of exposure to correlational structure in the input dimensions (Chapter 6). It is currently not known how passive exposure to a statistical regularity scales to affect more complex learning challenges or has cascading effects that impact subsequent learning. Thus, this experiment will provide insights into the impact of perceptual regularities on higher cognitive processes, such as those involved in category learning.
It is also not yet fully understood the extent to which the learning of novel statistical structure depends on which sensory dimensions are involved and the specific nature of the experienced structure (i.e. correlation). The current studies (Chapters 5-6) investigate the impact of passive exposure to correlations between stimulus dimensions that have not used before in examinations of rapid statistical learning or efficient coding. It is important to examine the generalizability of changes to perceptual representations of dimensions that have not been examined before to illuminate the underlying mechanism at play. If there are limited effects of rapidly adapting to statistical structure in sensory environments, then this would raise questions about the applicability of rapid efficient coding and statistical learning in characterizing broad human behavior.

In Chapters 5 and 6, we investigate the effects of brief, passive exposure to a novel statistical regularity, specifically a correlation between spectral modulation and temporal modulation acoustic input dimensions. In Chapter 5, we examine the flexibility of perceptual representations to this novel statistical structure and measure the change in similarity-based representations through similarity judgments and multidimensional-scaling. The aim of Chapter 5 is to illuminate how passive exposure can alter or adapt perceptual representations. In Chapter 6, we will examine how exposure to this statistical regularity alters complex behavior with the same kind of acoustic information, specifically during category learning. The aim of Chapter 6 is to understand the broader cognitive and behavioral consequences of this exposure to a strong statistical regularity. These studies will allow us to better understand the behavioral consequences of a general sensitivity of perceptual systems to statistical regularities in sensory information.
**Pilot Experiments**

The two dimensions that were used to create all stimuli are spectral modulation and temporal modulation. We conducted two pilot experiments to investigate the *a priori* relationship between these two dimensions and to understand the differences in discriminability across the two dimensions in the ranges specified by other experiments (Reetzke, Maddox, & Chandrasekaran, 2016; Yi & Chandrasekaran, 2016).

In Pilot 1, we used a stepwise AXB discrimination task to understand how the two-dimensional space is perceived across a 100-step continuum. In Pilot 2, we used a same-different judgment task with stimuli varying on either dimension or along both dimensions to understand the discriminability of stimuli in the stimulus space without any prior exposure.

**Experiment 4: Pilot #1**

The purpose of the first pilot experiment was to determine the minimum step-sizes along a continuum of spectral and temporal modulation that are approximately equivalent in terms of participants’ just noticeable differences. This will also demonstrate whether the two dimensions within the range specified in other experimental paradigms (Reetzke et al., 2016; Yi & Chandrasekaran, 2016) are approximately equivalent across the two dimensions. Further, this allows us to identify whether steps along the positive axis and negative axis varying on both stimulus dimensions are equivalent in terms of minimum just noticeable differences.

**Methods**

**Participants**

Participants were 11 (2 M, 9 F) Carnegie Mellon University undergraduates ages 18-25 participating for course credit. One subject was removed from analysis due to equipment failure. Ten participants were included in the analyses.
All participants performed the discrimination staircase procedure for the spectral modulation dimension, temporal modulation dimension, two-dimension positive axis, and two-dimension negative axis. This was a within-subjects design. The order of conditions was counterbalanced across participants.

**Stimuli**

Stimuli were complex sounds varying on spectral modulation and temporal modulation and had a 1 sec duration. Spectral modulation and temporal modulation were defined based on prior experimental procedures, outlined in Reetzke et al. (2016). While these dimensions are perceptually independent at several levels of acoustic processing, it is not clear whether they are psychologically separable. However, there is evidence that the neural populations encoding these dimensions are relatively separable (Depireux, Simon, Klein, & Shamma, 2017; Elliott & Theunissen, 2009; Langers, Backes, & Van Dijk, 2003; Santoro et al., 2014; Schönwiesner & Zatorre, 2009; Singh & Theunissen, 2003; Visscher, Kaplan, Kahana, & Sekuler, 2007; Woolley, Fremouw, Hsu, & Theunissen, 2005).

For each within-subjects condition, there were 100 stimuli varying in equal steps along one or two dimensions. The range of the stimuli along the two dimensions was identical to previous studies using these stimulus dimensions (Reetzke et al., 2016; Yi & Chandrasekaran, 2016). The sounds were generated using a custom MATLAB script (see Appendix). The script defined the following parameters: duration = 1 s; phase = 0 degrees; $F_0 = 200$ Hz; spectral bandwidth was defined based on an equation that bases it on the $F_0$; amplitude modulation depth was set a 0 dB; sampling rate = 44.1 kHz. Stimuli were then root mean square (RMS) amplitude matched at 70 Hz using Praat. The range along the temporal modulation dimension is 4 Hz to 12 Hz. The range along the spectral modulation dimension is .1 oct/cyc to 2 oct/cyc.
For the spectral modulation and temporal modulation one-dimension continua, the other dimension was held constant at the mean value of the dimension. For the temporal modulation condition, spectral modulation was held constant at 1.05 cyc/oct. For the spectral modulation condition, temporal modulation was held constant at 8 Hz.

For the temporal modulation condition, 1 step in the 100-step continuum is equal to 0.081 Hz. For the spectral modulation condition, 1 step in the 100-step continuum is equal to 0.019 cyc/oct. With the two-dimension continua, both temporal modulation and spectral modulation were continuously varying between each of the 100 stimuli. For both the positive and negative correlations, each of the 100 steps were separated by 0.08 Hz on the temporal dimension and 0.019 cyc/oct on the spectral dimension. The stimuli were created in MATLAB R2017a (see Appendix).

**Procedure**

All participants completed all four discrimination procedures for both dimensions and the positive and negative axes that varied in both dimensions on each step. Order of conditions was counterbalanced across participants. Participants completed the experiment in the Gorilla Experimental Builder (www.gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2018). The procedure for each condition was identical and only the stimuli varied.

**Staircasing procedure.** To find the just-noticeable differences in the 100-step continuum for the four within-subjects conditions, we employed a psychophysical staircasing procedure using an AXB design with a maximum of 75 trials. On each trial, participants heard three sounds (each 1 sec in duration with a 500 ms ISI) and, using the mouse, responded whether the first or the third sound was different from the second sound. Participants did not receive feedback. Following correct responses, the program gradually decreased the step sizes between stimuli
along the 100-step continuum until participants were incorrect on the smallest step size difference twice. The smallest step size that participants were able to consistently respond to correctly in the AXB task was coded as the just noticeable difference for that condition for that subject. Once this lowest just noticeable difference level was reached, the task ended.

Participants repeated this procedure for each of the four conditions, received a brief break between each condition, and were given identical instructions between each task. The entire experiment took approximately 10-20 minutes to complete.

**Results**

To compare the just noticeable difference step sizes between the four conditions, a repeated measures ANOVA was run with condition as a within-subjects variable and minimum step size as the dependent variable (Figure 1).

![Figure 1. Minimum accurate step size for each within-subjects condition. Error bars are standard error of the mean. Each individual point is a participant’s value and the black point is the group mean.](image-url)
The minimum step size differed across conditions \((F(3, 36) = 4.14, p = 0.0128)\). Bonferroni-corrected paired t-tests indicated that the minimum step size for Negative was higher than Positive \((p = 0.0067)\) and Spectral \((p = 0.0223)\), but not different from Temporal \((p = 1.0)\). Positive was not different from Spectral \((p = 1.0)\) but was lower than Temporal \((p = 0.0160)\). Temporal and Spectral were not statistically different \((p = 0.0826)\).

Even though there were differences in the minimum step sizes needed for each dimension and along the positive and negative axes, the pattern of results was not easily interpretable. In the second pilot experiment, we examined discriminability along each dimension using a same-different judgment task.

**Experiment 4: Pilot #2**

The purpose of the second pilot experiment was to examine discriminability across the continua of these two dimensions and the two-dimensional positive axis and negative axis continua. This should demonstrate if there are any perceptual discontinuities along the dimensions and will determine the step size at which accuracy is approximately 65% along a single dimension and approximately 70% along both dimensions in the discrimination task, which allows us to match the methods of previous experiments (Stilp et al., 2010).

**Methods**

For Pilot 2, following the methodology of other auditory experiments (Holt & Lotto, 2006; Stilp et al., 2010), we examined the discriminability of stimuli that vary on one dimension, while holding the other dimension at a constant value. We used an AX discrimination task where participants heard two sounds separated by 1, 2, 3, or 4 steps and responded whether they are the same sound or different sounds. The steps were equidistant along the stimulus range as outlined in previous studies with these two dimensions. The stimuli varied along one dimension or the
other or simultaneously varied on both dimensions either in a negative or positive correlation (as in the two-attribute control experiment from Stilp et al. 2010).

Participants

Participants were 80 (25 M, 53 F, 2 prefer not to answer) Carnegie Mellon University undergraduates ages 18-25 participating for course credit or a small payment.

Stimuli

The four stimulus continua were defined based on the range specified in previous studies in this space (Reetzke et al., 2016; Yi & Chandrasekaran, 2016). The stimuli varied along an 18-step continuum with one continuum along the temporal modulation dimension, one along the spectral dimension, one along the positive axis, and one along the negative axis (Figure 2). As in the first pilot experiment, in the one-dimension conditions, the other dimension was held constant at the mean of the range specified by previous studies (Reetzke et al., 2016; Yi & Chandrasekaran, 2016). The stimuli were created in MATLAB R2017a (see Appendix).

The stimulus step sizes for pilot #2 were achieved by evenly dividing the space into 17 spaces, creating a total of 18 stimuli for each continuum. In this new 18-step continuum, each step was separated by 0.47 Hz in the temporal modulation condition, 0.11 cyc/oct in the spectral modulation dimension, and 0.47 Hz/0.11 cyc/oct in the positive and negative conditions.
Figure 2. Stimulus distributions for same-different judgment task of Pilot #2.

Procedure

Participants were randomly assigned to one of four conditions. The procedure was identical for all conditions and the factor that varied was which dimension continuum participants were judging. There were 20 participants in each condition.

Participants made same/different judgments in an AX task. Similar to Holt and Lotto (2006), we tested discriminability of stimulus pairs that are 1, 2, 3, and 4 steps apart on the continuum. There were 496 trials with equal number of same and different trials. There are 80 total stimulus pairs for each continuum, which were presented in a counterbalanced order. Each of the 62 “different” stimulus pairs (17 stimuli 1 step apart, 16 stimuli 2 steps apart, 15 stimuli 3
steps apart, and 14 stimuli 4 steps apart) were repeated four times for a total of 248 “different” trials. Each of the 18 “same” stimulus pairs were repeated approximately 14 times (mean 13.78) for a total of 248 “same” trials.

On each trial, participants heard two sounds (1 sec in duration) separated by a 500 ms ISI. At the onset of the second sound, participants saw instructions on the screen (“Same or Different?”) with the associated response keys directly below each possible response to remind them which key was associated with which response. Across participants, response keys were counterbalanced. Participants did not receive feedback about their response. There was a 1 sec ITI between trials.

Participants received an untimed break halfway through the experiment and were told to rest briefly and regain their attention before they began again. The experiment took approximately 30-45 minutes to complete.

Results

We first present the results without filtering the data and then present filtered data after removing subjects who had a value above 10 on the d’ measure of response bias, beta.

Same-different accuracy across step sizes without filtering any data

In a 5 x 4 mixed-model ANOVA examining the effects of within-subject step distance (0, 1, 2, 3, 4) and between-subject condition (Spectral, Temporal, Positive, Negative), there was an interaction between step size and condition \(F(12, 304) = 2.77, p = 0.00136\). Figure 3 shows the same-different accuracy by step size across the dimension conditions.

To better understand what is driving this interaction, we ran a series of Bonferroni-corrected t-tests comparing conditions at each of the different step sizes. For 0 steps, 1 step, and 2 steps there were no significant differences among the four stimulus conditions. For 3 steps,
there were no differences between groups except the Negative condition had higher accuracy than Spectral condition. For 4 steps, there were no differences except Negative had higher accuracy than Spectral and Positive had higher accuracy than Spectral.

![Figure 3](image.png)

*Figure 3*. Same-different accuracy across step-size for the different stimulus sets, with all participants.

**d’ without filtering data**

According to a one-way ANOVA on d’, there was no main effect of condition with the unfiltered data (Figure 4, $F(3, 76) = 1.08, p = 0.36, \eta^2 = 0.041$). That is, the d’ values for all four dimensions were not statistically different.
Figure 4. $d'$ across different stimulus sets, with all participants.

**Filtering**

In the filtered data, individual participants were excluded on the basis of the $d'$ measure of response bias, beta. We included only participants who had a beta value less than 10. As a result of this filtering, there were 18 subjects in the Negative condition (2 excluded), 16 subjects in the Positive condition (4 excluded), 17 subjects in the Spectral condition (3 excluded), and 17 subjects in the Temporal condition (3 excluded).

**Same-different accuracy with filter based on response bias**

Similar to the filtered data, there was a significant interaction between stimulus condition and step size for the beta-filtered data ($F(12, 256) = 2.587844, p = 0.0029$). The pattern of accuracy across the step sizes differed by condition. Figure 5 shows the same-different accuracy for only the subjects with beta value less than 10. A more meaningful comparison, however, is the $d'$ across conditions.
Figure 5. Same-different accuracy across step-size for the different stimulus sets, only including participants with a beta (bias) value less than 10.

\textbf{d’ Results}

As with the unfiltered data, according to a one-way ANOVA on d’, there was no main effect of condition when excluding participants with beta values higher than 10 (Figure 6, $F(3, 64) = 0.54, p = 0.66, \eta^2 = 0.025$).

Because the average d’ did not differ across conditions, we did not alter the stimulus space in any way for the following experiments. Although the ability to discriminate at higher step sizes is not necessarily equivalent across dimensions or combination of dimensions, because we are giving each individual participant specific experience and examining judgments across the acoustic space, we believe this should not differentially impact learning ability. Because the average d’ is the same across dimensions and we will be collecting individualized similarity space measurements, in Experiment 4 we use a non-normalized space that is identical to the space used in previous experiments (Reetzke et al., 2016; Yi & Chandrasekaran, 2016).
Experiment 4

The goal of Experiment 4 is to understand the effects of short-term passive exposure on the similarity-based representations of a two-dimensional stimulus space. Using multidimensional scaling methods on similarity judgments between stimuli in a grid, we obtained “snapshots” of the representation space. We passively exposed participants to a novel statistical regularity—a perfect positive or negative correlation between the dimensions—and re-examined the snapshot of the representation space in a post-exposure of the same similarity judgments between stimuli in a grid.

Methods

Participants

Participants were 74 (28 M, 44 F, 2 prefer not to answer) Carnegie Mellon University undergraduates ages 18-25 participating for course credit or a small payment. Ten subjects were
not included in analyses due to equipment failure. A total of 64 subjects were included in analyses. Participants were randomly assigned to one of two conditions: Positive correlation exposure (n = 33) or Negative correlation exposure (n = 31).

**Stimuli**

**Stimulus dimensions.** The goal in choosing these stimulus dimensions was to have a stimulus matrix across which sounds separated in a grid are approximately equally discriminable (i.e. the grid approximates discrimination a priori; see pilot experiments). As in the pilot experiments, the stimulus dimensions that define the stimulus space were temporal modulation and spectral modulation.

**Stimulus distributions**

**Grid Distribution.** The grid distribution contains 13 stimuli that span the two-dimensional space (Figure 7). Pairs of stimuli within the grid can vary along the spectral modulation dimension, temporal modulation dimension, positive axis, or negative axis.

![Grid Distribution](image)

*Figure 7. Grid distribution of stimuli in the similarity judgment task.*

**Exposure Distributions.** Two sets of exposure distributions were created in R and reflected a perfect \((r = 1.0, r = -1.0)\) correlation between the two dimensions (Figure 8). The
distributions have identical variance, means, and covariance and differ only in the direction of the correlation between the two dimensions. There are 18 stimuli in both the positive and negative distributions and these distributions are identical to the Positive and Negative distributions in Pilot 2. The use of 18 equidistant stimuli was used in a similar experiment with different acoustic dimensions (Stilp et al., 2010).

![Figure 8](image-url)  
*Figure 8. Distribution of stimuli in the exposure phase with (A) positive correlation and (B) negative correlation.*

**Stimulus generation**

Stimuli were generated in a manner identical to the two pilot experiments using the custom MATLAB script in MATLAB R2017a (see Appendix).

**Procedure**

There were three parts of the experiment (Figure 9). In Part 1, participants completed a similarity ratings task as a pre-exposure measure of similarity-based representations of the stimulus space. In Part 2, participants listened passively to stimuli while drawing whatever they want on a blank piece of paper. In Part 3, participants completed the same similarity ratings task
as Part 1 as a post-exposure similarity measure. We compared the pre-exposure and post-exposure similarity ratings to quantify the change in similarity-based representations induced by the passive exposure phase.

**Figure 9.** Breakdown of the three parts of Experiment 4.

**Phase 1: Similarity judgments pre-exposure.** Participants judged the similarity of pairs of stimuli selected from 13 stimuli in the temporal modulation and spectral modulation acoustic space. Of the 13 stimuli, 5 fall along the positive axis (Figure 9, blue circles & center black circle), 5 fall along the negative axis (Figure 9, red circles & center black circle), and 4 fall in the space adjacent to either axis (Figure 9, black X). A single stimulus falls in the exact center of the positive and negative axis distributions (Figure 9, black circle).

On each trial, participants heard two 1 sec duration stimuli with a 50 ms ISI and judged the similarity on a scale from 1 (“the same”) to 4 (“completely different”) in a procedure that has
been used in other acoustic environments (Emberson, Liu, & Zevin, 2013). In both the pre-exposure and post-exposure phases, participants completed 274 trials, judging the similarity of each of the 91 possible pairs 3 times (78 unique different pairs, 13 same pairs). Participants judged the center stimulus 4 times to bring the total stimuli to an even 274 that was split evenly across two blocks. The stimulus pairs were presented in a counterbalanced order. Participants did not receive any feedback during the task. Each response was followed by a 1 sec ITI. Figure 10 shows a breakdown of a single trial.

![Figure 10. Illustration of the breakdown of a trial during the similarity pre-exposure phase.](image)

**Phase 2: Passive exposure.** In the second phase, participants passively listened to a stream of sounds for approximately 8 minutes. They heard 450 presentations of sounds (25 repetitions each of 18 sounds), a repetition number that has been shown in another stimulus space to affect discriminability (Stilp et al., 2010). These sounds were drawn from either the positive or negative distribution (Figure 8). Each 1 sec duration sound was followed by a 50 ms...
ITI. Participants were given markers and blank pieces of paper and told to draw whatever they would like for this part of the experiment.

**Phase 3: Similarity judgment post-exposure.** Immediately following the passive exposure phase, participants engaged in the identical similarity judgment task as in the pre-exposure phase. This post-exposure similarity judgment phase examines how participants’ similarity judgments of the stimulus space changed as a consequence of experiencing the statistical structure in the passive exposure phase. As in the pre-exposure phase, participants judged each of 91 pairs three times with the exception of the centroid, which they judged four times. The stimuli were presented in a random order without replacement.

**Results**

To understand how passive exposure to these correlations between the stimulus dimensions impacts the representations of stimuli, we examined the changes in similarity judgments from before and after exposure for stimulus pairs that fall along the positive axis, negative axis, spectral modulation dimension, and temporal modulation dimension.

Similarity judgments for individual participants and the average across participants are shown in Figure 11. To better visualize the change from pre-exposure judgments to post-exposure judgments, difference scores for the similarity judgments are shown in Figure 12. We will report results from analyses for the difference scores for the similarity judgments because there are inherent differences in the physical similarity of the stimuli based on the construction of the grid. Specifically, the stimuli along the spectral and temporal dimensions are closer in similarity space, on average, than stimuli along the positive and negative axes. To better understand the change in the representation, we examine only the differences in similarity
judgments in the post-exposure phase minus the pre-exposure phase, rather than the overall similarity judgments.

We ran a 4 x 2 mixed effects ANOVA on these difference scores with dimension varied (spectral, temporal, positive, negative) and exposure correlation (positive, negative) as variables. There was a significant main effect of dimension varied \((F(3, 183) = 6.319, p < .001, \eta_p^2 = .094)\), no main effect of exposure correlation \((F(1, 61) = 0.382, p = .539, \eta_p^2 = .006)\), and no interaction \((F(3, 183) = 0.944, p = .420, \eta_p^2 = .015)\).

There were differences in the change in similarity due to experience with the statistically structured information, but the effect was not dependent on the nature of that experience. Instead, the differences depended only on the dimensions along which the stimuli varied. According to Bonferroni-corrected post hoc comparisons, the stimuli were judged as more dissimilar after exposure when they varied along the positive axis compared to the negative axis \((p < .001)\). Additionally, stimuli were judged as more dissimilar after exposure when they varied along the positive axis relative to the temporal modulation acoustic dimension \((p = .006)\).

For each of the four dimension comparisons, we assessed whether the difference scores collapsed across exposure correlation were significantly different from zero. That is, we assessed whether there a difference in the pre-exposure and post-exposure judgments. We ran one-sample t-tests on the difference scores and found that after structured exposure of either kind (positive, negative), stimuli were judged as more dissimilar than before exposure when they fell along the positive axis \((M = 0.12, t(62) = 4.497, p < .001)\) or along the spectral modulation dimension \((M = 0.062, t(62) = 2.498, p = .015)\). In contrast, after structured exposure of either kind, stimuli were not judged differently than before exposure when they fell along the negative axis \((M = \)
0.019, \( t(62) = .750, p = .456 \) or along the temporal modulation dimension (\( M = 0.0053, t(62) = .176, p = .861 \)).

In summary, there were no effects on similarity judgments that depended on the specific nature of the exposure correlation. While we might have predicted that the exposure correlation would drive differences in similarity judgments in particular directions, instead, exposure to either statistical correlation led to stimuli along the positive axis and spectral modulation dimension being judged as more dissimilar after exposure and stimuli along the negative axis and temporal modulation dimension not changing.

**Figure 11.** Individual subject data and group averages for pre-exposure and post-exposure similarity judgments along the negative axis, positive axis, temporal modulation dimension, and spectral modulation dimension for the negative exposure condition (top) and positive exposure condition (bottom).
Figure 12. Difference scores (post-exposure minus pre-exposure) for similarity judgments based on the dimension that varied in the stimuli being compared. Data are shown separately for negative and positive exposure conditions, though there was no main effect nor interaction of exposure condition on the effect of dimension on difference in similarity. The dashed line reflects no difference between pre-exposure and post-exposure judgments. Negative values can be interpreted as an increase in similarity after exposure and positive values can be interpreted as an increase in dissimilarity after exposure.

Multidimensional scaling

To understand how the stimuli were arranged in a perceptual similarity space and to better assess a more generalized representation of this acoustic space, we conducted
multidimensional scaling analyses with an individual differences scaling procedure (INDSCAL) using the `smacofIndDiff` routine in the smacof library in R (De Leeuw & Mair, 2009). We restricted the number of dimensions to two.

Because we found no effects of the specific direction of correlation exposure on similarity judgments, we ran the MDS analyses collapsing across the negative and positive exposure conditions. The two-dimensional solutions (pre- and post-exposure) of the INDSCAL procedure for both negative and positive exposure conditions are shown in Figure 13.

In Figure 13, each individual point reflects the coordinates of the MDS solution for the pre-exposure judgments. These points are connected to a vector arrow, the end of which represents the coordinates of the MDS solution for the post-exposure judgments. Thus, Figure 13 demonstrates how the similarity space changed before and after exposure for all subjects. We note here that there is no specific direction that all arrows are facing. The direction of effects was specific to placement in the grid and seemed to change only very slightly based on exposure.
**Figure 13.** Results of multidimensional scaling analyses with pre-exposure values as individual points in the two-dimensional space and post-exposure values as the end of the arrow vectors. Data are presented collapsed across both positive and negative exposure conditions. Stimulus colors and shape match that of the grid distribution in Figure 9 (blue = positive axis, red = negative axis, black = adjacent stimuli and centroid).

It is important to note that the MDS dimensions do not necessarily correspond to the underlying acoustic dimensions. To understand the interpretation of this perceptual space and the dimensions of the MDS solution more directly, we constructed biplots using the `biplotMDS` function from the SMACOF library in R (De Leeuw & Mair, 2009). This function uses linear regression to understand how the two dimensions in the MDS space (perceptual dimensions) map on to the acoustic input dimensions (spectral modulation and temporal modulation). The
dependent variable of these regression models are the coordinates of the acoustic input
dimensions and the independent variables are the coordinates of the MDS dimensions
(Dimension 1 and Dimension 2). Figure 14 demonstrates the resulting biplots before and after
exposure for all participants together in the positive and negative exposure conditions.

Figure 14. MDS biplots for before (left) and after (right) exposure, collapsed across positive and
negative exposure conditions. Individual points reflect the coordinate space in similarity space
for the MDS solutions, colored by the corresponding identities in the coordinate grid (Figure 9,
blue = positive axis, red = negative axis, black = adjacent stimuli and centroid). Lines emanating
from the center of the space reflect best-fit lines for the spectral and temporal dimensions in the
output of the linear regression. Lengths of the lines correspond to the amount of variance mapped
onto that dimension and direction of the lines correspond to the axes of the acoustic input
dimension as laid onto the two-dimensional MDS space. The longer the vector, the more
variance is mapped onto that acoustic dimension (Greenacre, 2010).
The points in the space are the same as the pre and post-exposure points in Figure 13. Overlaid on the graph are the lines, which can be interpreted as the axes in the MDS space that best reflect the axes of the acoustic input dimensions, spectral modulation (spectral) and temporal modulation (temporal). The length of the vector arrows corresponds to the amount of variance mapped onto that dimension and the direction of the lines correspond to the axes of acoustic input dimension as laid onto the two-dimensional MDS space. The coordinates of the endpoint of the lines corresponds to the regression coefficient values for Dimension 1 and Dimension 2.

The clearest interpretation of Figure 14 is that the perceptual dimensions that emerged from the MDS analysis do not directly correspond to the acoustic input dimensions (spectral modulation and temporal modulation). That is, the spectral/temporal biplot lines are not parallel to the Dimension 1 and Dimension 2 axes in MDS space. Instead, both perceptual dimensions contribute to both acoustic dimensions. This is true for space before exposure and after exposure.

The contribution of both perceptual dimensions to the acoustic dimensions is evident from the results of the linear regressions (Tables 1 & 2). For the spectral modulation model (spectral = Beta*D1 + Beta*D2, where D1 is dimension 1 in MDS space and D2 is dimension 2 in MDS space), the overall models for pre-exposure ($F(2,11) = 152.7, p = 9.479e-09$, adjusted $R^2 = 0.9589$) and post-exposure ($F(2,11) = 163.0, p = 6.702e-09$, adjusted $R^2 = 0.9614$) were significant. Before exposure, the beta coefficients for the spectral model for dimension 1 (1.57) and dimension 2 (-1.22) were significant. This is also true for the post-exposure spectral model for dimension 1 (-1.68) and dimension 2 (-1.04). The relative size of the dimension coefficients can also be interpreted. According to these dimension coefficients, spectral modulation loads onto both MDS perceptual dimensions because both Dimension 1 and Dimension 2 coefficients...
were significant in the regression. Before exposure, spectral modulation loads onto Dimension 1 slightly more than Dimension 2 (coefficient difference = 0.35). After exposure, the spectral dimension loads even more onto Dimension 1 than Dimension 2 (coefficient difference = 0.64).

The model for the temporal dimension shows similar patterns. The results of the regression models are shown in Tables 1 and 2. For the regression model (temporal = Beta*D1 + Beta*D2), the overall models were significant for both pre-exposure ($F(2,11) = 220.1, p = 1.344e-09$, adjusted $R^2 = 0.9712$) and post-exposure ($F(2, 11) = 141.9, p = 1.399e-08$, adjusted $R^2 = 0.9559$). Before exposure, the beta coefficients for the temporal model for Dimension 1 (1.09) and Dimension 2 (-1.77) were significant. This is also true for the post-exposure temporal model for Dimension 1 (0.96) and Dimension 2 (-1.86) Temporal modulation loads onto both MDS perceptual dimensions. Before exposure, temporal modulation loads onto Dimension 2 slightly more than Dimension 1 (coefficient difference = 0.68). After exposure, the temporal modulation dimension loads even more onto Dimension 2 than Dimension 1 (coefficient difference = 0.90).

Together, these regression results indicate that the two dimensions in the MDS solution, or the dimensions that best reflect the perceptual representational space, do not map directly onto the acoustic input dimensions.

Table 1. Pre-exposure linear regression results.

<table>
<thead>
<tr>
<th>Acoustic input dimension (DV)</th>
<th>Coefficient for Dimension 1</th>
<th>Coefficient for Dimension 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral modulation</td>
<td>-1.5662</td>
<td>-1.2191</td>
</tr>
<tr>
<td>Temporal modulation</td>
<td>1.09056</td>
<td>-1.77382</td>
</tr>
</tbody>
</table>

Table 2. Post-exposure linear regression results.

<table>
<thead>
<tr>
<th>Acoustic input dimension (DV)</th>
<th>Coefficient for Dimension 1</th>
<th>Coefficient for Dimension 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral modulation</td>
<td>-1.6827</td>
<td>-1.0395</td>
</tr>
<tr>
<td>Temporal modulation</td>
<td>0.9600</td>
<td>-1.8565</td>
</tr>
</tbody>
</table>
Discussion

The goal of this experiment was to deepen understanding about the impact of experience with statistical regularities on perceptual representations of an acoustic space. Prior research showed that brief, passive exposure to correlations between two acoustic dimensions enhanced discrimination along the axis of correlation and diminished discrimination along the axis orthogonal to that experience (Stilp et al., 2010). This was taken as evidence for rapid efficient coding based on short-term, passive exposure to input dimension correlations. The current results conflict with these previous findings. This is the first study of rapid efficient coding or learning of statistical regularities to investigate the changes in the similarity-based representations in a stimulus space before and after passive exposure to statistically structured sensory information. We probed the impact of passive exposure with correlation structure at a representational level and did not find changes that depended on the nature of experience. The current study demonstrates that there are limits to the effects of short-term exposure to statistical regularities when there are not carefully controlled perceptual spaces, or one does not truly understand the nature of the perceptual representation of information.

We found that the change in similarity-based representations was not directly related to the direction of the correlation between two acoustic input dimensions. Instead, this exposure led to stimuli that fall along the positive axis or along the spectral modulation dimension to be judged as more distinct after exposure. The similarity of stimuli that fall along the negative axis and along the temporal modulation dimension did not change as a function of passive exposure. Additionally, both before and after exposure, the perceptual dimensions best captured in the MDS solutions were not directly related to the acoustic input dimensions. Instead, the acoustic input dimensions contribute to both perceptual dimensions.
In previous investigations of rapid efficient coding, Stilp and colleagues have used highly controlled stimulus spaces by constructing a perceptually normalized space where the difference between any two stimuli in a grid allowed for a discrimination accuracy of 65% along a single dimension and 70% along two dimensions (Stilp & Kluender, 2011). In the current study, we did not normalize the space in this manner. Instead, we maintained the approximation of the physical sensory dimensions on the x- and y-axis, rather than creating a highly controlled space with arbitrary units on the dimensions. This allowed us to examine how the dimensions of perception map onto the physical acoustic dimensions and maintain the ability to interpret the physical dimensions of the stimuli.

Further, the current study demonstrates that for these dimensions, even before exposure, there was not a one-to-one relation between the sensory dimensions (spectral and temporal modulation) and the perceptual dimensions (in the MDS solution). We selected spectral and temporal modulation dimensions to define our stimulus space because previous research demonstrated that the dimensions are represented by relatively independent neural populations (Depireux et al., 2017; Elliott & Theunissen, 2009; Langers et al., 2003; Santoro et al., 2014; Schönwiesner & Zatorre, 2009; Singh & Theunissen, 2003; Visscher et al., 2007; Woolley et al., 2005). Even so, we found that there was a sort of interaction between the sensory dimensions in perceptual representations. It may even be the case that the kinds of exposure correlations that we chose (positive, negative) more closely match participants’ existing representations of this acoustic space. That is, in representational space, the sensory dimensions are not independently represented. If this is the case, then the actual correlations we exposed participants to may actually better align with existing representations, rather than the correlation reflecting a novel statistical regularity. One way to investigate this possibility further would be to investigate...
exposure in this same stimulus space, but with no correlation in the exposure phase. That is, expose participants to sounds that have a zero correlation between the temporal and spectral modulation dimensions. If the dimensions of representation are closer to the positive/negative axes in the current study and these rapid efficient encoding effects are dependent on the structure being novel, then we should observe contrasting effects with the zero correlation. Instead, if this rapid adaptation or efficient coding is limited in some other way in this space, we might find, as in the current study, that there are limited effects on the plasticity of the representational space.

In addition to not observing effects that were dependent on the direction of the correlation of exposure, we found that when there were differences before and after exposure, they were only present for the positive axis and spectral modulation dimension stimuli. There were no differences between the pre-exposure and post-exposure judgments for the negative axis or temporal modulation dimension stimuli. It is not immediately clear why experience with either correlation direction in this acoustic space should lead to stimuli along the positive axis and spectral modulation dimension to increase in discriminability. In the first pilot experiment, we found that the negative axis and temporal dimension had larger just noticeable differences than the positive axis and spectral dimension. It could be that the perceptual system is better tuned toward the positive axis and spectral dimension already and this experience highlighted those dimensions even further. These similarity changes also could be due to changes in a decision process, rather than a perceptual process. Participants may have changed their decisional criterion of similarity for the spectral and positive axis stimuli. Though we offer some speculation, because we did not predict these effects, the current study was not designed to explain these effects and further research is needed to understand the mechanism of this change.
We believe that it is important to create a broader framework about the correspondence of physical dimensions and what the mind actually represents as perceptual dimensions. This study demonstrates that the mind and/or brain clearly do not rapidly adapt to any regularities in the environment, but rather it likely depends on factors other than the local statistics being presented in the experiment.

To speculate a bit further, we believe that some of the potential factors that could constrain the ability of a sensory system to adapt to statistical regularities rapidly in an experimental context may depend on how the statistics of the current environment relate to the statistics of the previously encountered environments and the existing perceptual representation of the information. It could be the case that in the current experiment, the exposure participants received matched closely to what their representations already look like, so there was little effect in any direction because no real learning or reweighting occurred. This could also be potentially differentiated with a future experiment that looks at a space that trains people on independence between these sensory dimensions, rather than a strong positive or negative correlation.

Limitations and future directions

It is possible that this small amount of passive exposure in this space is not enough to create substantial changes in this perceptual space. This is a question that should be answered in future research to understand what exactly is changing (or not) when participants experience statistical regularities. It could be that there were small changes, but we were not able to observe with our method because they are short-lived. The grid stimuli in our similarity judgment task had no correlation between the dimensions, so our results may reflect a wash out of any effects of the correlation because we conducted our analyses on the average of all of the post-similarity
judgment trials. It will be important to understand the longevity of any effects and the adjustments (if any) made to representations that are task-general or task-specific.

It is also possible that passive exposure might not have been enough to tip the scales to substantially change representations in the similarity-space that we were examining. Passive exposure to distributions of stimuli has been the main way that distributional learning has been investigated, importantly because much of this work has been done in infants. Learning as a result of passive exposure has been investigated extensively in both infants and adults (Adriaans & Swingley, 2017; Aslin et al., 1998; Aslin & Newport, 2014; Escudero & Williams, 2014; Maye et al., 2002; Pajak & Levy, 2012; Saffran, Newport, & Aslin, 1996; Wanrooij, Boersma, & van Zuijen, 2014). Additionally, the investigations of rapid efficient coding also use passive exposure paradigms to expose participants to the regularities (Stilp, Kiefte, & Kluender, 2018; Stilp et al., 2010; Stilp & Kluender, 2012, 2016). The mechanisms involved in adjustments to perceptual representations may be fundamentally different when participants experience the regularities in a passive manner, as in the current study, or in a more involved and interactive manner. For instance, the act of category learning, a more active and interactive task, can alter underlying perceptual and neural representations for dimensions that are relevant or irrelevant to the given task at hand (Ester, Sprague, & Serences, 2017; Folstein, Palmeri, & Gauthier, 2013; Goldstone, 1995, 1998; Goldstone, Lippa, & Shiffrin, 2001; Goldstone & Styvers, 2001; Wong, Folstein, & Gauthier, 2012). Additionally, more naturalistic tasks, such as incidental category learning can alter these learned representations in different ways than overt category learning (Gabay, Dick, Zevin, & Holt, 2015; Leech, Holt, Devlin, & Dick, 2009; Lim & Holt, 2011; Roark & Holt, 2018), indicating that the training task used (passive exposure, incidental learning, or overt category learning) can have drastic impacts on what changes as a consequence of
experience. The differences and similarities underlying the mechanisms at play in response to regularities that are presented in passive or interactive manner are not well understood.

While not passive exposure, a previous study demonstrated that the ultimate category representations are distinct when learning through an explicit categorization task with overt goals and feedback and an incidental categorization task, where participants are exposed to auditory category exemplars that are predictive of events and actions in a task that is seemingly unrelated to the sounds (Roark & Holt, 2018, chapter 2). This is an important area of research because understanding what is changing in perception and representation may depend on how that information is presented in the context of an observer’s broader goals and attention to the given task and information.

This is the first study of rapid efficient coding or learning of statistical regularities to investigate the changes in the similarity-based representations in a stimulus space before and after exposure. Stilp and colleagues have used AXB judgment tasks to examine how participants are or are not able to discriminate stimuli before and after exposure. We were able to examine the entire stimulus space to understand changes to representations that depend on the judgment of similarity between pairs of stimuli. Thus, we have a more graded and fine-grained understanding of the changes across the space than in a discrimination task. Our results are not consistent with Stilp et al.’s (2010) interpretation of their results that exposure to a correlation between two acoustic dimensions creates a novel perceptual dimension.

Stilp and colleagues demonstrated effects on the AXB judgments after the same amount of passive exposure as the current experiment. A future study should examine AXB judgments in this space to better reconcile previous findings and the current results. It could be the case that if we had used an AXB task, we would have found the same pattern of results predicted by an
efficient coding hypothesis. However, the current study demonstrates that there are limits to the impact of exposure to correlations of information on people’s perceptual judgments.

The results of this study and previous investigations of rapid efficient coding may be reconciled if we add an additional layer of complexity to the account. It might be the case that the nature of the existing perceptual representations play such a role in this kind of short-term adaptation and efficient coding that if presented with an environment that completely coheres with the existing representations, there will be few changes and given experience that is completely orthogonal (as with the McCollough effect), then rapid efficient coding occurs that leads to the aftereffects that we see in both the McCollough effect and previous investigations of rapid efficient coding in the auditory system (Lu, Liu, Dutta, Fritz, & Shamma, 2019; Stilp et al., 2018, 2010, Stilp & Kluender, 2012, 2016). However, without a full understanding of existing representations, this account is incomplete. It may not be sufficient to completely control and normalize a stimulus environment and then observe the effects; rather, it is our view that we need to crack open the black box of the transformation from sensory information to perceptual information to inform predictions of what kinds of statistical regularities people will be able to rapidly adapt to and change their underlying representations (in a short-term or long-term manner). The current study cannot fully account for this because it is a single stimulus space. However, because the effects we found were not dependent on the nature of exposure, this highlights the need for more understanding of the nature of the underlying representations.

**Conclusion**

We found that brief, passive experience with a correlation between two acoustic dimensions has some impact on underlying similarity-based representations. However, the changes to the underlying representations were not dependent on the particular regularities
experienced. The experience-driven changes did not differ across experience to a positive versus a negative correlation between input dimensions. In contrast to expectations from prior research (Stilp & Kluender, 2012, 2016; Stilp et al., 2010), the direct impact of passively-experienced acoustic regularities on perceptual similarity-based representations seems to be limited. Although organisms are sensitive to the statistical structure in the world, there are limits to how passive experience to statistical structure can change behavior. In Chapter 6, we will investigate how this passive exposure may impact more complex behavior, such as the use of acoustic dimensions in novel nonspeech category learning.
Appendix

MATLAB code used to create spectral modulation-temporal modulation stimuli

% code to make 2d gratings
% created on 18 June 2018
% by neeraj (neeks) at cmu
% adapted by Casey L. Roark on 19 June 2018

clear all; close all;

% Create a file with this name that has temporal modulation frequency in
% column 1 and spectral modulation frequency in column 2
y=load('Stimuli.txt');

% How many stimuli are included in the *.txt file
NumStim=18;

% Create Category Stimuli
for i=1:NumStim
% Variables here set sampling frequency and duration
Fs = 44.1e3; % sampling rate, max frequency is Fs*1/2
f0 = 200; % fundamental frequency, distance between harmonics
f_resol = 1/50;
N_octave = log2(Fs/2/f0); % spectral bandwidth
oct_scale_axis = (0:f_resol:N_octave-1);
lin_scale_axis = f0*2.^(oct_scale_axis);
N_freq_chan = length(lin_scale_axis);
phi = pi*(rand(1,N_freq_chan)-0.5);
dur = 1; % duration (in seconds)
t = (0:1/Fs:dur);

% ------ ripple params
scale = y(i,2); % n0 = spectral modulation
rate = y(i,1); % m0 = temporal modulation
dB = 0;
mu = 10^(dB/20);
ripple_phi =0; % phase = 0 degrees
% ------ make ripple
S = zeros(N_freq_chan,length(t));
for m = 1:length(t)
    for j = 1:length(oct_scale_axis)
        S(j,m) = 1+mu*sin(2*pi*rate*t(m)+2*pi*scale*oct_scale_axis(j)+ripple_phi);
    end
end
% ------ make the time domain signal
s = zeros(size(S));
carr = zeros(size(S));
for m = 1:size(S,1)
    carr(m,:) = sin(2*pi*lin_scale_axis(m)*t+phi(m));
    s(m,:) = S(m,:).*carr(m,:);
end
sig = sum(s,1);
sig = sig/max(abs(sig));
store_path = ['./StimuliDirectory/']; % directory for saving files
if (~ isdir(store_path))
    mkdir (store_path);
end
audiowrite([store_path 'StimName_' int2str(i),'.wav'],sig,Fs);
end;
References


Chapter 6: Prior exposure to correlations between two acoustic dimensions and the influence on subsequent category learning

Chapter 5 demonstrated that there are limits to the effects of brief, passive exposure to statistical regularities in an acoustic space on similarity-based representations. In this chapter, we investigate the effect of passive exposure to correlated acoustic dimensions on how listeners are able to use the perceptual information during a complex, cognitive task—specifically, category learning.

It is sometimes difficult to draw attention to information that is relevant for categorization, especially in the auditory realm, where input dimensions are often difficult to verbalize and are perceptually integral. The long-term goal of this research is to better understand how experience in a sensory environment may shift representations or alter behavior by changing how perceptual information is used during learning. This has implications for understanding how humans create and mentally organize complex novel object categories or second language speech categories. Specifically, this research may improve understanding about the mechanisms involved in learning difficult language contrasts in adulthood, like Hindi dental-retroflex contrast for native English listeners or English /r/-/l/ contrast for native Japanese listeners. If exposure to short-term regularities can shift attention to information in a manner that allows learners to use information differently during subsequent category learning, this would have broad implications.

Additionally, though sensitivity to broad statistical structure in the world has been investigated in multiple domains and in adults (Escudero & Williams, 2014; Wanrooij & Boersma, 2013), infants (Aslin et al., 1998; Maye et al., 2002; McMurray et al., 2009; Toscano
& McMurray, 2010), and non-human animals (Pons, 2006), there are gaps in understanding its long-term or behaviorally-relevant impact. It has been proposed that the sensitivity to distributional information is a fundamental ability that underlies infant acquisition of language, especially regarding speech sounds (Maye et al., 2002). We will investigate the effects of exposing participants to correlated acoustic information that is consistent with the distinction that will need to be made in category learning compared with exposure to an inconsistent relationship and with naïve, unexposed listeners. We will examine not only gross measures of behavior like accuracy, but also more individualized measures of how listeners use acoustic information to distinguish categories. We also will use decision bound computational models to understand the broad strategies of participants for drawing boundaries between categories and map their actual decision boundaries across training to understand how precise their boundaries are relative to what is required for the categories they are learning.

We are primarily interested in the learning and behavioral differences in the first block of training, when the differences among conditions should be driven by the differences in the pre-categorization exposure, which also may diminish across training due to common category-relevant feedback across groups. The primary goal is to examine whether and how passive exposure to regularities might induce (even a short-lived) bias that is carried forward to differentially impact category learning according to the nature of that exposed regularity and its alignment with the category-learning challenge. While previous research on rapid adaptation or learning of statistical regularities might predict that the statistical structure of these regularities changes how participants approach a perceptual category learning task, the effects on the cognitive processes and behavior have not been examined.
We designed Experiment 4 to measure representation change given different experience with the same stimulus space, but with different relationships between the dimensions within that space. Though we were unable to find reliable changes to those representations as a consequence of this exposure, it is possible that the effects of exposure to the high correlation between the dimensions still may impact other behaviors, such as those during novel category learning. Experiment 5 will examine differences in category learning within the same environment as a function of 1) experience and 2) category distribution type.

Methods

Participants

Participants were 181 (62 M, 117 F, 2 prefer not to answer) Carnegie Mellon University undergraduates ages 18-29 participating for course credit or a small payment ($10). Participants were randomly assigned to one of three exposure conditions (Exposure-Consistent, Exposure-Inconsistent, Naïve) and one of two category types (II-Positive, II-Negative). Those in the Exposure conditions also were exposed to either a positive or negative correlation, as in Experiment 4, and were classified as consistent or inconsistent based on the specific relation between the exposure correlation and the category being learned. For II-Positive categories, exposure to a positive correlation was coded as Consistent and exposure to a negative correlation was coded as Inconsistent. For II-Negative categories, exposure to a negative correlation was coded as Consistent and exposure to a positive correlation was coded as Inconsistent. There were six conditions, with approximately 30 participants in each group. The total number of participants in each condition are shown in Table 1.
Table 1. Number of participants in each condition

<table>
<thead>
<tr>
<th>Exposure Condition</th>
<th>Category Type</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure-Consistent (positive)</td>
<td>II-Positive</td>
<td>29</td>
</tr>
<tr>
<td>Exposure-Inconsistent (positive)</td>
<td>II-Negative</td>
<td>31</td>
</tr>
<tr>
<td>Exposure-Inconsistent (negative)</td>
<td>II-Positive</td>
<td>29</td>
</tr>
<tr>
<td>Exposure-Consistent (negative)</td>
<td>II-Negative</td>
<td>32</td>
</tr>
<tr>
<td>Naïve</td>
<td>II-Positive</td>
<td>30</td>
</tr>
<tr>
<td>Naïve</td>
<td>II-Negative</td>
<td>30</td>
</tr>
</tbody>
</table>

Stimuli

All stimuli were defined based on spectral modulation and temporal modulation and were created in an identical manner to those in Experiment 4. The stimuli were created using a custom MATLAB script (see Chapter 5 Appendix). Stimuli were then RMS amplitude matched at 70 Hz using Praat.

Stimulus distributions

Exposure distributions. The exposure distributions, used for passive exposure, were identical to those used in Experiment 4. Specifically, two distributions were defined based on a perfect positive correlation and a perfect negative correlation between two dimensions.

Category distributions. Two category types were created by sampling a bivariate Gaussian distribution using the `mvnorm` function in the MASS package in R. We sampled for a single category using normalized coordinates and then manipulated and rotated that distribution to create all other categories (Figure 1). The categories were rotated by 45 degrees and 135 degrees around the normalized midpoint (.5, .5) to create the II-Negative-A and II-Positive-A categories, respectively. Both of these categories were then mirrored across the space with a slope matching the optimal boundary between the categories. The slope was -1 for II-Negative and +1 for II-Positive. Thus, both category types—and each individual category—have the identical information in terms of variance, covariance, and the relationship between the two
categories in terms of overlap. These distributions were initially created in a normalized, arbitrary stimulus space that varied from 0 to 1 along both dimensions. Before stimuli were created, these coordinates were subjected to equations to transform from this arbitrary stimulus space to the temporal modulation and spectral modulation stimulus space. These transformation equations are identical to those used in previous studies (Reetzke, Maddox, & Chandrasekaran, 2016; Yi & Chandrasekaran, 2016). These equations are shown below as Equation 1 and Equation 2, where \( x \) is the arbitrary \( x \)-axis value and \( y \) is the arbitrary \( y \)-axis value.

Equation 1: \( \text{temporal modulation (Hz)} = 4 + x \times 8 \)

Equation 2: \( \text{spectral modulation (cyc/oct)} = 0.1 + y \times 1.9 \)

These final transformed categories are shown in Figure 1 and the means, variance, and covariance of the transformed categories are shown in Table 2.

![Figure 1](image)

*Figure 1.* Stimulus distributions for the two category types participants learned during the category learning phase in transformed space.
Table 2: Category Distribution Information for Training Stimuli

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (X, Y)</th>
<th>Variance (X, Y)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-Positive: Category A</td>
<td>(9.35, 0.83)</td>
<td>(2.01, 0.10)</td>
<td>0.190</td>
</tr>
<tr>
<td>II-Positive: Category B</td>
<td>(7.08, 1.37)</td>
<td>(1.80, 0.11)</td>
<td>0.190</td>
</tr>
<tr>
<td>II-Negative: Category A</td>
<td>(8.92, 1.37)</td>
<td>(1.80, 0.11)</td>
<td>-0.190</td>
</tr>
<tr>
<td>II-Negative: Category B</td>
<td>(6.65, 0.83)</td>
<td>(2.01, 0.10)</td>
<td>-0.190</td>
</tr>
</tbody>
</table>

Category distributions for the generalization test were sampled separately with 50 stimuli per category, in a manner identical to the training distributions. The categories were sampled using the same parameters as the training distributions but due to the probabilistic nature of the sampling, the distribution means, variances, and covariances are slightly different (Table 3). The transformed generalization test distributions are shown in Figure 2. The combined means, variances, and covariances for the training and test distributions are shown in Table 4.

Figure 2. Stimulus distributions for category types during the test phase in transformed space.
**Table 3: Category Distribution Information for Test Stimuli**

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (X, Y)</th>
<th>Variance (X, Y)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-Positive: Category A</td>
<td>(9.23, 0.75)</td>
<td>(1.34, 0.077)</td>
<td>0.15</td>
</tr>
<tr>
<td>II-Positive: Category B</td>
<td>(6.73, 1.34)</td>
<td>(1.37, 0.076)</td>
<td>0.15</td>
</tr>
<tr>
<td>II-Negative: Category A</td>
<td>(9.27, 1.34)</td>
<td>(1.37, 0.076)</td>
<td>-0.15</td>
</tr>
<tr>
<td>II-Negative: Category B</td>
<td>(6.77, 0.75)</td>
<td>(1.34, 0.077)</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

**Table 4: Category Distribution Information for All Stimuli (Training & Test)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (X, Y)</th>
<th>Variance (X, Y)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-Positive: Category A</td>
<td>(9.31, 0.80)</td>
<td>(1.78, 0.09)</td>
<td>0.18</td>
</tr>
<tr>
<td>II-Positive: Category B</td>
<td>(6.96, 1.36)</td>
<td>(1.68, 0.10)</td>
<td>0.18</td>
</tr>
<tr>
<td>II-Negative: Category A</td>
<td>(9.04, 1.36)</td>
<td>(1.68, 0.10)</td>
<td>-0.18</td>
</tr>
<tr>
<td>II-Negative: Category B</td>
<td>(6.69, 0.80)</td>
<td>(1.78, 0.09)</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

Both training and test distributions were subjected to decision bound modeling that will also be used to assess the decision bound strategies of participants across learning. This was to ensure that the true optimal model was the one idealized by the experimenter. All training stimulus sets and test stimulus sets were best fit by the optimal model. The results are shown in Table 5 and demonstrate that for all training and test distributions, the best-fit model is an integration (II) model and the slopes match the expected slopes.

**Table 5: Model results for category distributions**

<table>
<thead>
<tr>
<th>Category</th>
<th>Best-Fit Model</th>
<th>Best Fit AIC Value (Next closest)</th>
<th>Slope of Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-Positive-Training</td>
<td>II</td>
<td>77.28 (178.4975, UDX)</td>
<td>1</td>
</tr>
<tr>
<td>II-Positive-Test</td>
<td>II</td>
<td>12.21 (72.125, UDY)</td>
<td>1</td>
</tr>
<tr>
<td>II-Negative-Training</td>
<td>II</td>
<td>77.28 (178.4975, UDX)</td>
<td>-1</td>
</tr>
<tr>
<td>II-Negative-Test</td>
<td>II</td>
<td>12.21 (72.466123, UDY)</td>
<td>-1</td>
</tr>
</tbody>
</table>

**Procedure**

There were two parts in this experiment. Part 1 was a passive exposure phase and Part 2 was a category learning phase. Participants were randomly assigned to one of three experience conditions: Exposure-Consistent, Exposure-Inconsistent, or Naïve. Participants in all conditions then learned one of the two category distribution types: Information-Integration-Positive (II-
Positive) or Information-Integration-Negative (II-Negative). Participants in the two Exposure conditions first experienced the passive exposure phase where they were exposed to a particular relation between the two dimensions—positive correlation or negative correlation, which were coded as Consistent or Inconsistent based on the relation between the correlation in exposure and the category being learned. When the category being learned was II-Negative, the negative correlation was Consistent and the positive correlation was Inconsistent. When the category being learned was II-Positive, the positive correlation was Consistent and the negative correlation was Inconsistent. This exposure training was identical to the passive exposure phase of Experiment 4. Participants in the Naïve conditions did not engage in any exposure training and began the experiment in the category learning phase.

Exposure phase (Exposure conditions only). Participants in the Exposure conditions experienced the same passive exposure phase as participants in Experiment 4. Participants listened passively to a stream of sounds for approximately 8 minutes. Each 1 sec sound was followed by a 50 ms ISI. Participants experienced 450 presentations of sounds, 25 repetitions each of 18 stimuli, presented in a random order.

Category learning phase. Participants in both Exposure conditions and the Naïve condition completed the category learning phase. In this phase, participants were randomly assigned to learn one of two possible category types (Figure 1)—II-Positive or II-Negative. Note that the nomenclature we are using here of II-Positive and II-Negative refers to the placement of the optimal boundary between the two categories; the optimal boundary for II-Positive has a positive slope and the optimal boundary for II-Negative has a negative slope.

Participants learned the categories in a supervised categorization task across four blocks of training with 96 trials per block for a total of 384 training trials. On each trial (Figure 3),
participants heard a single exemplar selected randomly without replacement followed by a screen on which they were prompted about whether they believed the sound belonged to Category A (associated with a box on the left of the screen) or Category B (associated with a box on the right of the screen). Participants indicated their category response with a key press (\textit{u} or \textit{i}). Response keys for each category and category location were counterbalanced across participants. After a response was made there was a 500 ms pause after which participants were given feedback about the correctness of their response (“Correct!” or “Incorrect!”). In addition to the written feedback, a red X appeared in the correct box that is associated with the category. This red X was presented regardless of the correctness of the response. Feedback was displayed for 500 ms before a 1 sec ITI. Participants were told at the beginning of the category learning phase to use this feedback to inform their future decisions about the sound categories.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Illustration of the breakdown of a trial for the category learning phase.}
\end{figure}

\textit{Generalization test phase.} After completing the 384 training trials, participants completed a generalization test phase. In this phase, participants no longer received feedback and
encountered completely novel stimuli (Figure 2). Participants completed a total of 100 trials with 50 stimuli in each category.

**Results**

As a reminder, the main goal of the current experiment was to understand the broader effects on behavior after exposure to a high correlation between acoustic dimensions. If participants are not only sensitive to the statistical regularities of this passive exposure phase, but also if these statistical regularities can affect complex cognitive behavior and decision making, one might predict that categorization behavior and how participants use the perceptual information during learning will depend on the nature of the experienced regularities. Thus, we might expect an interaction between the category being learned (II-Positive, II-Negative) and the nature of the experience (Naïve, Consistent, Inconsistent). Because we did not find any effects that were dependent on the nature of the correlation to which people we exposed on their similarity-based representations in Experiment 4, it could also be the case that this kind of passive exposure to these correlations in this space do not sufficiently cue the perceptual system to rapidly adapt to the regularities in a way that would meaningfully alter behavior such as category learning. If this is the case, we might predict no interaction between the category being learned and the nature of experience.

**Behavioral Results**

**Training.** Using a mixed-model ANOVA, we examined the effect of experience condition (Naïve, Consistent, Inconsistent), training block (1, 2, 3, 4), and category type (II-Positive, II-Negative). Specifically, we coded the experience as a function of the relation of the correlation to the category. For the II-Positive condition, the positive exposure was coded as Consistent and the negative exposure was coded as Inconsistent. For the II-Negative condition,
the positive exposure was coded as Inconsistent and the negative exposure was coded as Inconsistent.

Figure 4. Accuracy across the four training blocks for the six conditions. Chance performance (50%) is denoted by a dashed line. The black points reflect the mean of each condition and the ribbon error bars reflect the SEM.

Accuracy performance across blocks is shown in Figure 4. Mauchly’s test of sphericity was significant ($p < .001$), so we report Huynh-Feldt corrected values. Performance changed across the four training blocks as there was a significant main effect of block ($F(2.43, 423.952) = 20.857, p < .001, \eta^2_p = .106$). According to post-hoc Bonferroni-corrected comparisons, there was improvement in accuracy from block 1 to block 2 ($p < 0.001$) and block 2 to block 3 ($p = 0.034$), but the accuracy in block 3 was not different from accuracy in block 4 ($p = 1.00$).
The learning trajectories across blocks did not differ depending on whether the exposure condition was Consistent, Inconsistent, or Naïve \((F(4.845, 423.952) = 1.801, p = .114, \eta_p^2 = .020)\), nor on which category was being learned \((F(2.423, 423.952) = 1.662, p = .184, \eta_p^2 = .009)\). There was also no interaction between block, experience condition, and category type \((F(4.845, 423.952) = .334, p = .887, \eta_p^2 = .004)\). Overall, participants improved in category learning across the blocks up to block 3 and this pattern did not differ based on prior experience or the category type that was being learned.

In examining the differences between the conditions, collapsing across blocks, there was a significant main effect of category type, such that accuracy for II-Positive categories was higher than for II-Negative categories \((F(1, 175) = 65.435, p < .001, \eta_p^2 = .272)\). Collapsing across blocks, participants in the II-Positive condition were, on average, 11.8% more accurate than participants in the II-Negative condition. There was no main effect of experience type—those who had Consistent prior experience, Inconsistent prior experience, or were Naïve, did not differ in their accuracy \((F(2, 175) = .506, p = .604, \eta_p^2 = .006)\). There was also no interaction between experience condition and category type \((F(2, 175) = .035, p = .966, \eta_p^2 = .000)\). The nature of the experience did not change how the categories were learned.

We were most interested in the effects in the first block to account for the possibility of an extinction of effects of exposure once all participants start receiving category-relevant feedback during learning. We examined the group differences within the first block of 96 trials. To break this first block of exposure down even further, we created two mini-blocks of 48 trials to examine the potential early differences in accuracy between groups (Figure 5).
Figure 5. Accuracy, showing individual subject variability within each condition, in the first block split in halves (miniblocks). Chance performance (50%) is denoted by the dashed line. Individual subjects’ performance is shown as colored points and the group mean and SEM are shown in black for each group. Note that this is an expanded version of the same data shown in Figure 4 for block 1, split in two miniblocks (miniblock 1, on left; miniblock 2, on right).

We ran a mixed-model ANOVA to understand the effects of experience condition (Consistent, Inconsistent, Naive), category type (II-Positive, II-Negative), and miniblock (1, 2). Results are shown in Figure 5.

Participants improved in accuracy from the first miniblock to the second miniblock ($F(1, 175) = 17.383, p < .001, \eta^2_p = .090$). This improvement pattern was not dependent on experience type ($F(2, 175) = .064, p = .938, \eta^2_p = .001$). Participants in Inconsistent, Consistent, and Naive
conditions all improved similarly from miniblock 1 to miniblock 2. There was a significant interaction between miniblock and category \((F(1, 175) = 7.615, p = .006, \eta_p^2 = .042)\). There was greater improvement between miniblock 1 and miniblock 2 for II-Positive than II-Negative learners. For II-Positive categories, miniblock 1 accuracy was 66.9% and miniblock 2 accuracy was 72.9% (difference = 6%). For II-Negative categories, miniblock 1 accuracy was 58.4% and miniblock 2 accuracy was 59.6% (difference = 1.2%). There was not an interaction between miniblock, experience type, and category \((F(2, 175) = .242, p = .785, \eta_p^2 = .003)\).

Collapsing across both miniblocks, we found significant main effects of both experience condition and category type. There was no interaction between experience condition and category type \((F(2, 175) = .146, p = .864, \eta_p^2 = .002)\). As with the overall results, even in the first 96 trials of the experiment, participants learning the II-Positive categories had higher accuracy than participants learning the II-Negative categories \((F(1, 175) = 46.991, p < .001, \eta_p^2 = .212)\). However, in this first set of trials, there was also a small main effect of experience condition \((F(2, 175) = 3.173, p = .044, \eta_p^2 = .035)\). The mean accuracy in Block 1 was 67.1% for the Consistent condition, 64.1% for the Inconsistent condition, and 62.2% for the Naïve condition. Bonferroni-corrected post-hoc comparisons demonstrate that there was no significant difference in accuracy in this first block of trials between Consistent and Inconsistent conditions \((p = .370)\) or between the Naïve and Inconsistent conditions \((p = 1.00)\). However, there was a significant difference between the Consistent and Naïve conditions, such that those who had experienced a Consistent correlation with the category they were learning were more accurate in block 1 than those who had no experience before beginning the category learning task \((p = 0.041)\). If there are any effects of prior passive experience on category learning, the effects are short-lived and relatively small \((\eta_p^2 = 0.035)\).
The learning that occurred in this experiment was also very rapid for all conditions. Even in the first miniblock (first 48 trials), the average accuracy in all conditions was greater than chance (50%). This same pattern remained for all conditions in the second half of block 1 as well. Results for these one-sample t-tests can be found in Table 6.

Table 6. One-sample t-test results for miniblock 1 and miniblock 2 for all conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average Accuracy</th>
<th>One-sample t-test (compared to chance, 50%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve, II-Negative</td>
<td>Miniblock 1: 55.00% Miniblock 2: 57.43%</td>
<td>Miniblock 1: t(29) = 2.878, p = .007 Miniblock 2: t(29) = 4.018, p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miniblock 1: t(29) = 4.018, p &lt; .001</td>
</tr>
<tr>
<td>Consistent, II-Negative</td>
<td>Miniblock 1: 61.39% Miniblock 2: 62.17%</td>
<td>Miniblock 1: t(31) = 6.167, p &lt; .001 Miniblock 2: t(31) = 6.589, p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miniblock 1: t(31) = 6.589, p &lt; .001</td>
</tr>
<tr>
<td>Inconsistent, II-Negative</td>
<td>Miniblock 1: 58.87% Miniblock 2: 59.27%</td>
<td>Miniblock 1: t(30) = 5.547, p &lt; .001 Miniblock 2: t(30) = 4.625, p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miniblock 1: t(30) = 4.625, p &lt; .001</td>
</tr>
<tr>
<td>Naïve, II-Positive</td>
<td>Miniblock 1: 65.49% Miniblock 2: 71.04%</td>
<td>Miniblock 1: t(29) = 5.651, p &lt; .001 Miniblock 2: t(29) = 7.284, p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miniblock 1: t(29) = 7.284, p &lt; .001</td>
</tr>
<tr>
<td>Consistent, II-Positive</td>
<td>Miniblock 1: 69.47% Miniblock 2: 75.29%</td>
<td>Miniblock 1: t(28) = 9.207, p &lt; .001 Miniblock 2: t(28) = 10.537, p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miniblock 1: t(28) = 10.537, p &lt; .001</td>
</tr>
<tr>
<td>Inconsistent, II-Positive</td>
<td>Miniblock 1: 65.88% Miniblock 2: 72.27%</td>
<td>Miniblock 1: t(28) = 5.765, p &lt; .001 Miniblock 2: t(28) = 9.294, p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miniblock 1: t(28) = 9.294, p &lt; .001</td>
</tr>
</tbody>
</table>

Generalization test. After the four blocks of training, participants engaged in a generalization test where they encountered novel category exemplars not encountered in training and no longer received any feedback. We examined the overall pattern of generalization test accuracy across conditions and the extent of generalization by comparing generalization test accuracy to chance (50%) and to block 4 accuracy.
Figure 6. Accuracy, showing variability within each condition, in the generalization test block. Chance performance (50%) is denoted by the dashed line. Individual subject performance is shown as colored dots and the group mean and SEM are shown in black for each group.

We ran a two-way ANOVA on generalization test to examine the effects of the experience (Naive, Consistent, Inconsistent) and category type (II-Positive, II-Negative). As Figure 6 demonstrates, we found that test accuracy was higher for II-Positive ($M = 80.6\%$) than II-Negative ($M = 69.3\%$) categories ($F(1, 175) = 31.122, p < .001, \eta_p^2 = .151$). Test accuracy was not different across the exposure conditions ($F(2, 175) = .081, p = .922, \eta_p^2 = .001$) and there was no interaction between category type and exposure conditions ($F(2, 175) = .013, p = .987, \eta_p^2 = .000$). This pattern of results is the same as we observed in the ANOVA for
performance across training. Thus, the general pattern of results in the test block was the same as
the training blocks.

Participants in all conditions successfully generalized their category knowledge from the
training blocks to this novel set of stimuli. As shown in Table 7, performance in all conditions
was significantly better than chance (50%). Additionally, as shown in Table 8, for all conditions,
accuracy in test block was significantly higher than accuracy in the final training block (block 4).
This demonstrates successful transfer of knowledge and also indicates that the stimuli in our test
distributions may have been inherently easier to classify or participants behaved slightly
differently when they no longer received feedback.

Table 7. One-sample t-test results for the generalization test for all conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average Accuracy</th>
<th>One-sample t-test (compared to chance, 50%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive, II-Negative</td>
<td>69.93%</td>
<td>t(29) = 7.773, p &lt; .001</td>
</tr>
<tr>
<td>Consistent, II-Negative</td>
<td>69.34%</td>
<td>t(31) = 7.210, p &lt; .001</td>
</tr>
<tr>
<td>Inconsistent, II-Negative</td>
<td>68.58%</td>
<td>t(30) = 7.895, p &lt; .001</td>
</tr>
<tr>
<td>Naive, II-Positive</td>
<td>80.77%</td>
<td>t(29) = 12.962, p &lt; .001</td>
</tr>
<tr>
<td>Consistent, II-Positive</td>
<td>80.79%</td>
<td>t(28) = 12.699, p &lt; .001</td>
</tr>
<tr>
<td>Inconsistent, II-Positive</td>
<td>80.17%</td>
<td>t(28) = 12.528, p &lt; .001</td>
</tr>
</tbody>
</table>

Table 8. Paired samples t-test results for the generalization test vs block 4 for all conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average Difference (Test minus Block 4)</th>
<th>Paired samples t-test (test vs block 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive, II-Negative</td>
<td>5.38%</td>
<td>t(29) = 4.078, p &lt; .001</td>
</tr>
<tr>
<td>Consistent, II-Negative</td>
<td>5.48%</td>
<td>t(31) = 3.277, p = .003</td>
</tr>
<tr>
<td>Inconsistent, II-Negative</td>
<td>5.24%</td>
<td>t(30) = 3.615, p = .001</td>
</tr>
<tr>
<td>Naive, II-Positive</td>
<td>6.15%</td>
<td>t(29) = 4.341, p &lt; .001</td>
</tr>
<tr>
<td>Consistent, II-Positive</td>
<td>5.72%</td>
<td>t(28) = 4.938, p &lt; .001</td>
</tr>
<tr>
<td>Inconsistent, II-Positive</td>
<td>5.82%</td>
<td>t(28) = 3.699, p = .001</td>
</tr>
</tbody>
</table>
Computational Modeling

Rationale. We ran several classes of decision bound computation models, which are derived from General Recognition Theory (Ashby & Townsend, 1986), a multivariate version of signal detection theory (Green & Swets, 1966). The methods and overall rationale for these methods were reported in Chapter 4 and elsewhere (Ashby, 1992; Ashby & Maddox, 1992, 1993; Maddox & Ashby, 1993).

We were primarily interested in how the pre-category learning exposure (or lack thereof in the Naïve conditions) may have changed how participants distinguish between categories, especially in the first block, when they have not received a lot of external, category-relevant feedback. Therefore, while we examine the best-fit decision boundary strategies and actual placement of the boundaries for all four training blocks and in test, we will examine strategy use in block 1 in more detail.

Modeling methods. We fit several classes of decision boundary models. The four main classes of models that we fit were: unidimensional rule-based strategy along the x-axis (in this case, temporal modulation dimension), unidimensional rule-based strategy along the y-axis (in this case, spectral modulation dimension), information-integration strategy, and random responder. We fit the models and determined the best-fit strategy for all subjects across each of the four training blocks (96 trials/block), the test block (100 trials), and the miniblock breakdown of block 1 (48 trials/miniblock). See chapter 4 (Roark & Holt, 2019) for more details about the precise specification of the classes of models.

Results. Figure 7 shows the proportion of participants across learning best-fit by the different decision bound model strategies. To assess the precision of these best fit decision boundaries in the acoustic space, we map the best-fit decision boundaries in Figure 8.
Figure 7. Proportion of participants best fit by each strategy in the four training blocks and test block across the six conditions. None of the participants in any block were best fit by a random responder model, so it is not shown on the figure.

Strategy use across training and test. Visual examination of the pattern in Figure 7 reveals that there are large differences in the best fit strategies for the II-Negative conditions (left) and the II-Positive conditions (right). While most of the participants in the II-Negative conditions were best fit by one of the two unidimensional models (x-axis or y-axis), most of the
participants in the II-Positive conditions were best fit by the optimal integration model, even early in learning.

Additionally, when participants are best fit by a unidimensional strategy, it was more often a unidimensional-x strategy, where participants base their decision on the temporal modulation dimension. Collapsing across the three II-Negative conditions, in each of the four training blocks and test block, the total proportion of participants best fit by a unidimensional-x strategy was larger than the proportion best fit by a unidimensional-y strategy (Table 9). This same pattern was also true collapsing across the three II-Positive conditions (Table 9). When participants were using a suboptimal unidimensional strategy, they were doing so more often based on the temporal modulation dimension (unidimensional-x strategy) than the spectral modulation dimension (unidimensional-y strategy).

Table 9. Proportion of participants best fit by unidimensional strategies across blocks

<table>
<thead>
<tr>
<th>Category</th>
<th>Unidimensional-X (Temporal) Strategy</th>
<th>Unidimensional-Y (Spectral) Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-Negative</td>
<td>Block 1: 44.09%</td>
<td>Block 1: 37.6%</td>
</tr>
<tr>
<td></td>
<td>Block 2: 53.8%</td>
<td>Block 2: 24.7%</td>
</tr>
<tr>
<td></td>
<td>Block 3: 53.8%</td>
<td>Block 3: 28%</td>
</tr>
<tr>
<td></td>
<td>Block 4: 52.7%</td>
<td>Block 4: 25.8%</td>
</tr>
<tr>
<td></td>
<td>Test: 38.7%</td>
<td>Test: 22.6%</td>
</tr>
<tr>
<td></td>
<td>Block 1: 25%</td>
<td>Block 1: 10.2%</td>
</tr>
<tr>
<td></td>
<td>Block 2: 14.8%</td>
<td>Block 2: 11.36%</td>
</tr>
<tr>
<td>II-Positive</td>
<td>Block 3: 19.3%</td>
<td>Block 3: 4.5%</td>
</tr>
<tr>
<td></td>
<td>Block 4: 17%</td>
<td>Block 4: 13.6%</td>
</tr>
<tr>
<td></td>
<td>Test: 12.5%</td>
<td>Test: 10.23%</td>
</tr>
</tbody>
</table>

The classification of the strategy used by each participant is useful to illustrate broad patterns across participants. To understand how participants are dividing the categories in the acoustic space at a more fine-grained level, we mapped the actual best-fit decision boundaries relative to the optimal boundary (Figure 8). This provides a better understanding of how precise participants are at placing the boundary. This is an important distinction because, two
participants might be best fit by the same integration strategy model but they could be placing
the boundaries in the space completely differently—for instance, with different slopes.
Additionally, because the optimal strategy of the two kinds of categories we are investigating is
integration, it is clear that we need a more precise understanding of how exactly participants are
integrating across the space. We present the decision boundary information in three ways: 1)
visual inspection of the best-fit decision boundaries plotted relative to the optimal boundary
(Figure 8), 2) comparison of the degree of the angle of the best-fit decision boundary compared
to the optimal angle (Figure 9), and 3) examining the correlation between the difference in the
best-fit boundary angle relative to the optimal angle and accuracy across learning (Figure 10).

*Best-fit decision boundaries.* Figure 8 shows the best-fit decision boundaries for each
subject in each condition across the four training blocks and in test. This figure demonstrates
quite clearly that participants in the II-Positive conditions (Figure 8, right) were much closer to
the optimal category distinction in the space than participants in the II-Negative conditions
(Figure 8, left). This was true even in the first block. Across all of the blocks and even in test, the
majority of participants in the II-Negative condition were not integrating along the two
dimensions. Even when participants were integrating (i.e. Best-fit by the integration model), they
were typically integrating by placing a boundary along the positive axis (positive slope), not the
negative axis (negative slope). An integration strategy with a positive slope is beneficial for the
case of II-Positive categories but is detrimental for learning the II-Negative categories. There
was a general difficulty in distinguishing the categories with a negative decision boundary, even
though this kind of boundary is completely consistent with the feedback given to the participants
in the II-Negative conditions.
Figure 8. Best-fit decision boundary lines for each participant in each block (columns) across conditions (rows). The optimal decision boundary for the II-Positive (right) and II-Negative (left) categories are shown in a dashed red line.

Degree of angle relative to optimal. Figure 9 shows boxplots of the angles of the decision boundaries across the different conditions relative to the optimal angle (dotted line at 135 degrees for II-Negative, 45 degrees for II-Positive). To compare the angles between exposure conditions, we ran a non-parametric Kruskal-Wallis one-way ANOVA on the decision bound angle data. For both II-Negative and II-Positive category types, the decision bound angles for the Consistent, Inconsistent, and Naïve exposure conditions were not significantly different (Table 10). To compare how closely participants in each condition were able to place their decision boundary relative to optimal, we conducted one-sample Wilcoxon Signed Rank Tests, a non-parametric version of the one-sample t-test, to compare the best-fit angles to optimal angle for the two categories (II-Negative, optimal: 135 degrees; II-Positive, optimal: 45 degrees). We used
non-parametric tests because our data are not normally distributed, which violates the assumptions of the one-sample t-test and one-way ANOVA.
Figure 9. Across all blocks, boxplots showing the angles of the best-fit decision bound models for all participants across conditions. The optimal decision boundary angle is shown in the dotted line and is 135 degrees for the II-Negative condition and 45 degrees for the II-Positive condition. This gives a sense of precision of the decision boundary for each participant—the closer to the optimal angle, the closer the participant is to the optimal distinction between the two categories. Note that the unidimensional-y boundary angle was coded as 180 degrees for II-Negative conditions and 0 degrees for II-Positive conditions.

Table 10. Results of Kruskal-Wallis one-way ANOVA on Exposure Condition for II-Negative and II-Positive categories

<table>
<thead>
<tr>
<th>Block</th>
<th>II-Negative</th>
<th>II-Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>$\chi^2(2) = 1.18, p = 0.56$</td>
<td>$\chi^2(2) = 2.52, p = 0.28$</td>
</tr>
<tr>
<td>Block 2</td>
<td>$\chi^2(2) = 1.20, p = 0.55$</td>
<td>$\chi^2(2) = 0.01, p = 0.99$</td>
</tr>
<tr>
<td>Block 3</td>
<td>$\chi^2(2) = 4.66, p = 0.10$</td>
<td>$\chi^2(2) = 0.68, p = 0.71$</td>
</tr>
<tr>
<td>Block 4</td>
<td>$\chi^2(2) = 0.74, p = 0.69$</td>
<td>$\chi^2(2) = 1.56, p = 0.46$</td>
</tr>
<tr>
<td>Test</td>
<td>$\chi^2(2) = 2.14, p = 0.34$</td>
<td>$\chi^2(2) = 0.85, p = 0.65$</td>
</tr>
</tbody>
</table>

Consistent with the visual inspection of the decision boundaries in Figure 8, it is clear that the median angle of the best-fit boundaries for II-Negative category learners was further from optimal than for the II-Positive category learners. Using one-sample Wilcoxon signed rank tests, we compared the median best-fit decision bound angle collapsing across these conditions for all II-Negative learners together. We observed that the median best-fit decision bound angles for II-Negative participants was significantly different from the optimal angle of 135 degrees across all four training blocks and in test (Table 11). Using one-sample Wilcoxon signed rank tests, we compared the median best-fit decision bound angle collapsing across these conditions for all II-Positive learners together. We observed that the median best-fit decision bound angles for II-
Positive learners was also significantly different from the optimal angle of 45 degrees across all four training blocks and in test (Table 11).

Table 11. Results of One-Sample Wilcoxon Signed Rank Tests on Median Best-Fit Decision Bound Angle

<table>
<thead>
<tr>
<th>Block</th>
<th>II-Negative</th>
<th>II-Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>$Mdn = 90, V = 916,$</td>
<td>$Mdn = 56, V = 2821,$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.0001$</td>
<td>$p = 0.0003$</td>
</tr>
<tr>
<td>Block 2</td>
<td>$Mdn = 90, V = 543,$</td>
<td>$Mdn = 57, V = 2784,$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.0001$</td>
<td>$p = 0.00059$</td>
</tr>
<tr>
<td>Block 3</td>
<td>$Mdn = 90, V = 720,$</td>
<td>$Mdn = 57, V = 3023,$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.0001$</td>
<td>$p &lt; 0.0001$</td>
</tr>
<tr>
<td>Block 4</td>
<td>$Mdn = 90, V = 776,$</td>
<td>$Mdn = 58, V = 2671,$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.0001$</td>
<td>$p = 0.003$</td>
</tr>
<tr>
<td>Test</td>
<td>$Mdn = 90, V = 766,$</td>
<td>$Mdn = 56, V = 2656,$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.0001$</td>
<td>$p = 0.0037$</td>
</tr>
</tbody>
</table>

Correlation between difference in best-fit angle and the optimal angle and accuracy across learning. To understand how the precision of an individual’s decision boundary relates to their accuracy across learning, a Spearman’s rank-order correlation was run to determine the relationship between absolute difference of best-fit decision bound angle and optimal angle and accuracy across learning and in test. A reasonable prediction might be that the closer the decision boundary angle is to optimal (i.e. closer to zero in Figure 10), the higher the accuracy will be. We tested this directly for the II-Negative and II-Positive conditions and the results are shown in Table 12.

We first computed the absolute difference between the best-fit decision boundary angle and the optimal angle for the two category types. We plotted this absolute value against the proportion correct across blocks and in test (Figure 10). If the participant’s decision boundary angle was identical to the optimal angle, the x-axis value is zero (vertical dashed line). The further from zero, the larger the distance from optimal the angle is and thus, the less precise the participant was at placing the decision boundary in the space. Again, as above, we coded the
unidimensional-y decision boundary as 0 degrees when the category was II-Positive and 180 degrees when the category was II-Negative.

For the II-Negative category learners, there was a significant monotonic relationship between the difference in angle from optimal and accuracy in blocks 3, 4, and test (Table 12). For the II-Negative learners, there was no significant monotonic relationship in blocks 1 and 2. In contrast, for II-Positive learners, there was a strong, significant relationship between difference in angle from optimal and accuracy in all training blocks and test. That is, the closer the boundary is to the optimal boundary, the higher accuracy participants had across training and in test. The directions of these correlations were always negative. The strength of the correlation was also always stronger for the II-Positive condition than the II-Negative condition.

In general, this pattern is unsurprising. The closer one applies a decision boundary to the optimal boundary, the more accurate one is. However, it is clear from looking at the individual data that an individual subject can still be quite accurate even with a suboptimal decision boundary angle.

Table 12. Results from Spearman’s rank order correlation across blocks and test for II-Negative and II-Positive category learners

<table>
<thead>
<tr>
<th>Block</th>
<th>II-Negative</th>
<th>II-Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>$\rho(91) = -0.048, p = 0.65$</td>
<td>$\rho(86) = -0.58, p &lt; 0.001$</td>
</tr>
<tr>
<td>Block 2</td>
<td>$\rho(91) = -0.12, p = 0.24$</td>
<td>$\rho(86) = -0.55, p &lt; 0.001$</td>
</tr>
<tr>
<td>Block 3</td>
<td>$\rho(91) = -0.23, p = 0.025$</td>
<td>$\rho(86) = -0.57, p &lt; 0.001$</td>
</tr>
<tr>
<td>Block 4</td>
<td>$\rho(91) = -0.22, p = 0.036$</td>
<td>$\rho(86) = -0.54, p &lt; 0.001$</td>
</tr>
<tr>
<td>Test</td>
<td>$\rho(91) = -0.43, p &lt; 0.001$</td>
<td>$\rho(86) = -0.57, p &lt; 0.001$</td>
</tr>
</tbody>
</table>
Figure 10. Across all blocks and in test, the relation between the absolute value of the difference between the optimal angle and actual best-fit decision boundary angle and proportion correct (accuracy) across all participants. The y-axis dashed line reflects chance accuracy (50%) and the x-axis dashed line reflects a decision boundary that is equivalent to the optimal angle (difference = 0). The points are color-coded according to exposure condition.

Computational Results for Miniblocks. With regard to differences among the exposure conditions for each type of category, we again turn to the strategies within block 1 to best compare the effects of any bias induced by the experience before the category learning phase. Figure 11 shows the proportion of participants across the conditions best fit by the different classes of strategies in block 1, split by miniblock. This figure still shows the same pattern as the overall results; the starkest differences were among the II-Negative and II-Positive conditions. Note that there are some differences between the overall block 1 (96 trials) and each miniblock
(each 48 trials) solution of the best fit models. We show the best-fit boundaries for the block 1 miniblocks in Figure 12.

![Proportion of Participants fit by Each Strategy Across Blocks](image)

**Figure 11.** Proportion of participants best fit by each strategy across block 1 broken into two miniblocks. None of the participants in either miniblock were best-fit by a random responder model, so it is not shown on the figure.

*Proportion best-fit by decision bound strategies.* For the II-Negative categories, there are relatively few strategy differences between the conditions and across miniblocks. The vast majority of participants were using a unidimensional strategy (over 75% of subjects in each of
the miniblocks). The Naïve conditions reflect how participants engage with this category information without any prior experience with similar stimuli. In the Naïve II-Negative condition, 90% of participants used a unidimensional strategy in miniblock 1 (50% x-axis, 40% y-axis) and only 10% used an optimal integration strategy. In miniblock 2, 6.7% used an integration strategy, 63.3% used a unidimensional-x strategy and 30% used a unidimensional-y strategy. It appears to be difficult for participants to integrate these acoustic dimensions such that they can draw a decision boundary along the negative axis.

After being exposed to a Consistent, negative correlation, participants learning II-Negative categories are mainly using unidimensional strategies in miniblock 1 (78.1%) and miniblock 2 (87.5%). After being exposed to an Inconsistent, positive correlation, participants learning II-Negative categories were also mainly using unidimensional strategies in miniblock 1 (77.4%) and miniblock 2 (77.4%).

For the II-Positive categories, there are also small differences between conditions, but the pattern and change across the miniblocks is interesting. In all conditions, the prevailing strategy in the first 48 trials (miniblock 1) is a unidimensional strategy. In the Naïve condition, only 30% of participants were best fit by the optimal integration model. In the Consistent exposure condition 34.5% of participants were best fit by the optimal integration model. In the Inconsistent exposure condition 37.9% of participants were best fit by the optimal integration model. In the

In miniblock 2, this pattern changes. In all three conditions, more participants are best fit by an integration model in miniblock 2 than miniblock 1. There was a large shift for the Naïve condition (30% best fit by optimal integration model in miniblock 1 to 60% in miniblock 2). There was a slightly smaller shift for the Consistent condition (34.5% best fit by integration
model in miniblock 1 to 51.7% in miniblock 2). There was the smallest shift for the Inconsistent condition (37.9% best fit by integration model in miniblock 1 to 48.3% in miniblock 2).

The feedback is successful in the II-Positive condition at having participants increasingly use both dimensions, rather than a unidimensional strategy, even in the earliest stage of learning. This is consistent with the II-Positive condition showing such rapid early learning in block 1. This feedback seems the most effective for participants in the Naïve condition who demonstrated the most bias to use a unidimensional strategy in miniblock 1. The exposure conditions may have encouraged participants to use an optimal integration strategy earlier in learning compared to the Naïve condition. Thus, though the Naïve participants do catch up quickly and, in fact, more participants in that condition were best fit by the integration strategy in miniblock 2, there might have been an early and small induction in some bias caused by the exposure phase that cued participants to use both dimensions in the II-Positive condition. However, the fact that this occurred only for II-Positive and not II-Negative does suggest that there are other, significant factors at play, such as pre-existing biases that might already have made II-Negative a much harder mapping to create than II-Positive.

*Decision boundaries relative to optimal.* To uncover the precision of the decision boundaries in the earliest stage of learning, we mapped the best-fit boundary for each participant for the two miniblocks of block 1 relative to optimal (Figure 12).
Figure 12. Best-fit decision boundary lines for each participant in block 1 split by miniblocks only across conditions. The optimal decision boundary for the II-Negative (top) and II-Positive (bottom) categories are shown in the dashed red line.

Degree of angle relative to optimal. Figure 13 shows the decision bound angles in the two miniblocks of block 1 for II-Negative (left) and II-Positive (right) relative to optimal (dotted line at 135 degrees for II-Negative and 45 degrees for II-Positive). To compare the decision boundary angles between exposure conditions, we ran a non-parametric Kruskal-Wallis one-way ANOVA on the decision bound angle data. Across the two miniblocks of block 1, both II-Negative and II-Positive category types, the decision bound angles for the Consistent, Inconsistent, and Naïve exposure conditions were not significantly different (Table 13). Using one-sample Wilcoxon signed rank tests, we compared how the median angle of the decision boundaries across participants collapsed across exposure conditions for II-Negative and II-Positive categories differed from the optimal angles of boundaries for the two category types. We observed that the best-fit decision bound angles for the II-Negative participants differed from the optimal angle of...
135 degrees across both miniblock 1 and miniblock 2 (Table 14). For the II-Positive participants, there were no significant differences between the optimal angle of 45 degrees and the median best-fit angle in the first miniblock, but there was a difference in miniblock 2 (Table 14).

**Figure 13.** Across the two miniblocks of block 1, boxplots showing the angles of the best-fit decision bound models for all participants across conditions. The optimal decision boundary angle is shown in the dotted line and is 135 degrees for the II-Negative condition and 45 degrees for the II-Positive condition. This gives a sense of precision of the decision boundary for each participant—the closer to the optimal angle, the closer the participant is to the optimal distinction between the two categories. Note that the unidimensional-y boundary angle was coded as 180 degrees for II-Negative conditions and 0 degrees for II-Positive conditions.
Table 13. Results of Kruskal-Wallis one-way ANOVA on Exposure Condition for II-Negative and II-Positive categories across the miniblocks of block 1

<table>
<thead>
<tr>
<th>Block</th>
<th>II-Negative</th>
<th>II-Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miniblock 1</td>
<td>$\chi^2(2) = 1.33, p = 0.51$</td>
<td>$\chi^2(2) = 0.72, p = 0.70$</td>
</tr>
<tr>
<td>Miniblock 2</td>
<td>$\chi^2(2) = 1.45, p = 0.48$</td>
<td>$\chi^2(2) = 0.17, p = 0.92$</td>
</tr>
</tbody>
</table>

Table 14. Results of One-Sample Wilcoxon Signed Rank Tests on Median Best-Fit Decision Bound Angle in the two miniblocks of block 1

<table>
<thead>
<tr>
<th>Block</th>
<th>II-Negative</th>
<th>II-Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miniblock 1</td>
<td>$\text{Mdn} = 90, V = 660, p &lt; 0.001$</td>
<td>$\text{Mdn} = 56.98, V = 1888, p = 0.77$</td>
</tr>
<tr>
<td>Miniblock 2</td>
<td>$\text{Mdn} = 90, V = 810, p &lt; 0.001$</td>
<td>$\text{Mdn} = 56.04, V = 2788, p &lt; 0.001$</td>
</tr>
</tbody>
</table>

Correlation between difference in best-fit angle and the optimal angle and accuracy across learning. A Spearman’s rank-order correlation was run to determine the relationship between absolute difference of best-fit decision bound angle and optimal angle and accuracy in the miniblocks of block 1 (Figure 14). Even in the earliest stage of training, there was a strong negative relationship for the II-Positive category learners in both miniblock 1 ($\rho(86) = -0.71, p < .001$) and miniblock 2 ($\rho(86) = -0.64, p < .001$). In contrast, there was no significant monotonic relationship between difference in angle from optimal and accuracy for the II-Negative category learners in either miniblock 1 ($\rho(91) = -0.041, p = 0.70$) or miniblock 2 ($\rho(91) = -0.0047, p = 0.96$). Within the first miniblock, those in the II-Negative condition were relatively successful with a suboptimal strategy, but accuracy seems to depend greatly on how well that strategy is applied as there is a lot of individual variability.
Figure 14. Across the two miniblocks of block 1, the relation between the absolute value of the difference between the optimal angle and actual best-fit decision boundary angle and proportion correct (accuracy) across all participants. The y-axis dashed line reflects chance accuracy (50%) and the x-axis dashed line reflects a decision boundary that is equivalent to the optimal angle (difference = 0). The points are color-coded according to exposure condition.

Discussion

We investigated the impact of prior experience with correlations between acoustic dimensions on subsequent auditory category learning. We found that differences in experience in the acoustic environment prior to category learning did not substantially change the learning trajectories or the way that learners interacted with the environment while learning information-integration categories. Prior experience with correlations in the acoustic space did not lead to differences in overall accuracy, decision bound strategies, decision boundary placement, or the angle of decision boundaries, relative to Naïve participants. However, there were some potential, small effects that were present in the earliest stages of learning that may indicate that after exposure to a correlation that is consistent with the category boundary that needs to be learned,
one might have a slight earlier advantage in finding the optimal strategy and utilizing an appropriate decision boundary, which then leads to slightly higher accuracy. These small effects disappeared by block 2, demonstrating that it is difficult to create strong perceptual change through passive exposure that is not nearly immediately overridden by overt feedback during category learning.

One consistent and robust finding from this study is that it was easier for participants to learn II-Positive categories (optimal decision boundary with a positive slope) than II-Negative categories (optimal decision boundary with a negative slope). This persistent difference indicates that there may be some additional biases in the sensory system that make learning in this space difficult and not necessarily penetrable based on brief, novel, structured experience (see Chapter 5). These sensory biases may exist in many sensory spaces (e.g. Chapter 4, Roark & Holt, 2019) and should be seriously considered in models of perceptual category learning.

It could be the case that to see broad, robust effects on categorization behavior, a change in perceptual representations may be required. In Chapter 5, we did not observe experience-dependent effects on perceptual representations of this acoustic space. It is an open question whether or not representational changes are required to observe effects in category learning or if category learning can be biased based on short-lived perceptual biases. The current set of experiments examines effects in a single acoustic environment and finds that passive exposure to statistical regularities did not cause changes in perceptual representations and also did not have long-lived effects on categorization behavior.

A result that is of particular interest is that we did not find effects in opposite directions for Consistent and Inconsistent exposure. An interpretation of Stilp et al.’s (2010) results is that the brief, passive experience with the correlation of the acoustic dimensions creates a novel
perceptual dimension that reflects this correlation. If a novel dimension were created based on this passive exposure in the present experiment, then presumably participants would have been able to use that dimension in category learning. Our results are inconsistent with this interpretation because participants who experienced the correlation of the acoustic dimensions were not better able to use the axis of that correlation as a true dimension during categorization. These results demonstrate that passive exposure to structured information in the environment does not automatically make that information available to cognition.

Several other studies have used this spectral modulation/temporal modulation acoustic environment to train participants on auditory categories. Yi and Chandrasekaran (2016) found that when learning four categories in this space that were separable (rule-based) or inseparable (information-integration), that with full feedback (as in the current study), the category types were learned equally well at about 57% and 58% accuracy, respectively, across 600 trials. Chance performance for four categories is 25% and for two categories is 50%, so it is not surprising that overall accuracy in the current study was slightly higher than in Yi and Chandrasekaran (2016). They did not report learning patterns across blocks, so we cannot compare how performance unfolded across trials. Reetzke et al. (2016) investigated four-category rule-based learning in 7-12-year-old children, 13-19-year-old adolescents, and 20-23-year-old adults. Examining only their adult group, to better compare with our adult group (though it should be noted we included 18-29-year-old adults in the current study), they found a stronger effect of block with steeper learning curves than the current study. However, even in their first block, adults had around 40% accuracy, which is higher than chance performance (25% for four categories). In the final block of trials (trials 500-600), their adult participants were performing at around 59% accuracy. Because these studies used four categories and the current
study used two, the comparisons are not perfect, but the studies seem to agree and demonstrate that some learning that occurs in this acoustic space occurs very rapidly and that even after 600 trials (Reetzke et al., 2016; Yi & Chandrasekaran, 2016) or 384 trials (current study), learners are not performing at ceiling on these difficult, probabilistic categories.

We did not find substantial changes to the similarity-based representations, but we have some understanding about what the similarity-based representations in this space look like based on the results of Experiment 4. We found that the dimensions in the MDS analysis did not map perfectly on to the acoustic dimensions of spectral modulation and temporal modulation. Instead, one dimension appeared to mostly contain stimuli that varied along the negative axis (Chapter 5, Figure 14, Dimension 1) and the other dimension appeared to mostly contain stimuli that varied along the positive axis (Chapter 5, Figure 14, Dimension 2). The clearest result from the current study was that learning the II-Positive categories was easier than learning the II-Negative categories. A question arises about what makes learning the II-Positive distinction easier than the II-Negative distinction. These categories are statistically identical; the only difference is the rotation in this two-dimensional space. Nonetheless, participants engaged with the information differently based on the feedback they received (decision bound strategies) and had much higher accuracy for II-Positive categories than II-Negative categories, regardless of prior experience with structure in the space. We can only speculate about what makes one kind of category easier than the other. One possibility is that it could be the case that there exists some processing or representational bias as a result of long-term prior experience that makes mapping easier in one case and not the other. What is clear is, just as in Chapter 4 (Roark & Holt, 2019), there is some pre-existing bias in the system that makes some kinds of category learning much easier than others. Neither the current study nor the entire dissertation have cracked this code. This work...
provides more evidence that these biases exist and can constrain learning. In this case, learning is constrained much more significantly by the rotation of the categories in space than based on prior experience with structured correlations of acoustic dimensions. In the next chapter, we present a neural network model that attempts to uncover the mechanism of this bias.

An interesting pattern in the data is that the learning that did occur in this experiment occurred quite early in the experiment. Though there were significant differences in accuracy across the first three blocks of training, a vast majority of the accuracy gains that participants had were in the first block of 96 trials. However, participants on average achieved an accuracy of only about 70-80%. It could be the case that these categories are simply not completely learnable without much more training. It could also be that participants, unmotivated to learn more about the idiosyncrasies of the categories to increase their accuracy, were content with the accuracy they achieved. One way to better understand the ultimate learnability of these categories is to conduct a study examining several participants who learn these categories until they reach a criterion of automatization (Reetzke, Xie, Llanos, & Chandrasekaran, 2018). If these two category types remain unlearnable to a ceiling performance even after several sessions of training, this could provide even further insight into the constraints of the learning system. However, if automatization is possible, but takes more training, it might be clear that in these single session category learning experiments, we are only tapping into the earliest learning processes and quite different mechanisms may be at play to build longer-term representations that can be built across multiple sessions.

The participants in Chapter 5 and the current experiment were given passive exposure to the structured acoustic information. It is entirely possible that the passive nature of this experience contributed to the null effects obtained across the two experiments. It would be
interesting to examine how other kinds of experience with structured information might be able to induce some of the differences that we initially predicted. The process of category learning itself can have broader effects on perceptual organization. Learning categories that highlight one dimension over another are thought to have perceptual effects of highlighting or stretching the relevant dimensions, while down-weighting or shrinking the irrelevant dimensions (Feldman, Griffiths, & Morgan, 2009; Goldstone, 1994, 1995, 1998; Goldstone, Lippa, & Shiffrin, 2001). Novel dimensions are also thought to be created through a process of categorization, where one might create a novel dimension to best distinguish categories through feedback or other experience through a general process referred to as unitization (Goldstone, 1998). Visual category learning also alters the neural representations of category-relevant and irrelevant features in visual cortex (Ester, Sprague, & Serences, 2017; Folstein, Palmeri, & Gauthier, 2013). While in the current set of experiments we did not find large or sustained effects of prior experience with structured acoustic information, this is not to say that any and all experience would not affect broader behavior. The current findings provide some support for the idea that there are limits to the effects of passive exposure on restructuring perceptual representations or on impacting cognitive processing. Future studies should examine more interactive forms of exposure to understand how the mind and brain use information in the environment to affect future behavior. One might predict that when encountering structured information that is directly or incidentally linked to meaningful actions or events in an environment, that this might have more profound or long-lasting effects on behavior than the passive exposure used in the current experiments.
Conclusion

We found that passive exposure with a positive or negative correlation between spectral modulation and temporal modulation had limited effects on the subsequent learning of auditory categories comprised of these dimensions. These findings highlight the need for understanding the limits of short-term passive exposure on restructuring of perceptual representations that influence significant cognitive processing and decision making, such as that involved in complex category learning. Humans may be sensitive to structure in the sensory world, but the ultimate impact of this sensitivity should be investigated more closely to better understand how the sensory world, perceptual representations, and cognitive processes influence one another.
References


Chapter 7: A neural network model of the effect of prior experience with regularities on subsequent category learning

Research from visual and auditory category learning suggests that there may be at least two kinds of learning mechanisms that are better suited for specific kinds of category structures (Ashby & Maddox, 2005, 2011; Chandrasekaran, Koslov, & Maddox, 2014). So-called rule-based (RB) structures require selective attention to a single dimension that defines the categories, or a complex, but verbalizable rule that separates the categories. Information-integration (II) structures require pre-decisional integration across two dimensions defining the categories (Ashby & Maddox, 2005, 2011).

The distinction between RB and II categories has received a lot of attention in the literature. Considerably fewer resources have been devoted to understanding how underlying perceptual representations of the sensory information may influence cognitive processing during category learning. This is an important concept because, as my previous research has shown, what an experimenter may draw on a page as clearly separable, orthogonal dimensions in sensory space may not truly be orthogonal in perceptual space (Roark & Holt, 2019; chapters 4, 5, and 6). To fully understand cognitive processing and decision-making during category learning, it is necessary to bring perception into the picture and to understand the role perception and perceptual representations of sensory information play in these higher-level processes. Without understanding the nature of the mental and neural representations of incoming sensory information, it is not possible to understand how a category learning mechanism will interact with the sensory information.
Perceptual representations might be shaped through experience with statistical structure in sensory input. It is possible that long-term experience with the statistical regularities and relationships between sensory dimensions in naturalistic environments may constrain category learning involving those same dimensions. Research on second-language speech category learning has thoroughly addressed the role of prior experience on category learning. Whereas young infants are able to discriminate sounds from the world’s languages, at 10 months they lose the ability to discriminate speech contrasts that are not present in their native language, while also improving their ability to discriminate contrasts in their native language (Kuhl et al., 2006; Werker & Tees, 1984). This change in perceptual ability is thought to be accompanied by warping of perceptual representations of these sound categories (Iverson & Kuhl, 1995; Kuhl, 1991). This early, native language experience fundamentally changes perceptual representations of speech and impacts human speech perception into adulthood. In adult second-language speech category learning, speech contrasts that are not present in one’s native language are still difficult to discriminate, as they are at 10 months old (Flege, 1995; Miyawaki et al., 1975). Researchers have proposed that the precise nature of the conflict between one’s own native language and the sound contrast in the second language determines the difficulty (Best, 1995; Best, McRoberts, & Goodell, 2001; Best, McRoberts, & Sithole, 1988; Iverson, Kuhl, Akahane-Yamada, & Diesch, 2003). For instance, learning a sound contrast where the sounds are present, but not distinguished in one’s native language (e.g. rhotic flap in Japanese becomes two sound categories of /ɾ/ and /l/ in English) is extraordinarily more difficult than learning a sound contrast with entirely novel sounds (e.g. Zulu click categories for native English listeners).

While the role of prior experience in learning has been investigated thoroughly in speech perception, exactly how prior experience impacts learning mechanisms remains unknown. To
fully understand the mechanisms of perceptual category learning, it is my view that we also must understand the nature of perceptual representations of the sensory information involved in learning and how these existing representations interact with cognitive processing during learning.

The goal of the current neural network model is to illuminate the black box of perceptual representations from a theoretical perspective and to make predictions about how the nature of perceptual representations of sensory information can influence category learning. In this chapter, I trained the network on five kinds of relations in a sensory environment that will mimic long-term statistical relationships in experience in naturalistic environments (such as those in a native language environment). I then trained a categorization model that used the resulting learned, internal (hidden) representations as input. I trained this categorization model on four novel category learning environments (representative of those one might encounter in non-native language environments). I predict that the precise nature of the trained relationship (prior experience) will have drastic consequences on how the novel category learning problems are learned. I will outline these specific predictions in the Methods section.

**Methods**

There were two overall phases to the neural network training: representation learning and category learning. In the representation learning phase, I used an autoencoder model—in which each input is reconstructed over the output via a smaller number of hidden units—to train representations of a hidden layer to reflect the regularities present in five separate stimulus environments. In the category learning phase, I trained the model with the frozen hidden unit representations on four unique two-category problems in similar stimulus environments as the training environments.
**Representation Learning**

**Stimuli.** I created five distributions of stimuli to train the model during the representation learning phase. The resulting hidden unit representations symbolically reflect the perceptual representations that exist in mental representations of sensory information that can affect the ability to perceive and use different sensory information. The five training distributions are shown in Figure 1 and are labeled by type (Independent, A, B, C, and D). Each distribution has 289 stimuli. Table 1 shows the means, variance, and covariance of the training distributions.

*Figure 1.* Distributions of stimuli for training representations by type.
Table 1. Training Distribution Information.

<table>
<thead>
<tr>
<th>Distribution Type</th>
<th>Mean (X, Y)</th>
<th>Variance (X, Y)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>(.5, .5)</td>
<td>(0.015, 0.015)</td>
<td>0</td>
</tr>
<tr>
<td>Type A</td>
<td>(.5, .5)</td>
<td>(0.03, 0.03)</td>
<td>-0.03</td>
</tr>
<tr>
<td>Type B</td>
<td>(.5, .5)</td>
<td>(0.03, 0.03)</td>
<td>0.03</td>
</tr>
<tr>
<td>Type C</td>
<td>(.5, .5)</td>
<td>(0.00094, 0.06)</td>
<td>0</td>
</tr>
<tr>
<td>Type D</td>
<td>(.5, .5)</td>
<td>(0.06, 0.00094)</td>
<td>0</td>
</tr>
</tbody>
</table>

The Independent distribution is an equally spaced grid which was designed to reflect a situation where the sensory input dimensions are independent (the variance along one dimension is equal to the variance along the other dimension) and there is no directional relationship between the dimensions (the covariance is zero). The Type A and Type B distributions were designed to reflect a situation where the sensory input dimensions have contingent, directional relationship and are represented in such a way that they have nearly collapsed onto a single perceptual dimension. Type A reflects an underlying dimension with a negative slope and Type B reflects an underlying dimension with a positive slope. The Type C and Type D distributions reflect a situation where a single sensory dimension is clearly and robustly represented in perceptual space but the other dimension is less well represented. Type C reflects a robust representation of the y-axis dimension and Type D reflects a robust representation of the x-axis dimension.

Model architecture. The full model architecture is shown in Figure 2. The components that are relevant for the representation learning component of the model are the sensory input, perceptual representation hidden layer, and autoencoder output.
The sensory input and autoencoder output for the representation learning model are identical. Each component has 20 units—10 units represent the x-coordinate value and 10 units represent the y-coordinate value. The activation for the specific units was determined by coding the value of the dimension into an unnormalized Gaussian function where each unit takes on the value along that Gaussian distribution such that the overall activation of the bank of 10 units is equal to one, but the value of a single unit is larger when it is closer to the actual dimension value. This allows for graded input which reflects population encoding of information sensory cortex. The function used to create the input based on the coordinates was: \( \exp\left(-((x - m)^2)/0.025\right) \) where \( m \) is normalized by the unit from 0.05 to 0.95 in increments of .1 across the ten units and where \( x \) is the x- or y-coordinate value in the two-dimensional space. Additionally, this input layer was given a small amount of noise (range = 0.1) to reflect a small amount of noise in the perceptual encoding of a stimulus.
The 20 sensory input units are connected to a 10-unit hidden layer, which reflects the perceptual representation of the incoming sensory information after learning. This hidden layer is then connected to an autoencoder output layer which is identical to the sensory input layer.

**Training procedure.** I trained the autoencoder network for 50,000 epochs on batched learning across the 289 stimuli within each training type with a learning rate of 0.0001 on each of the five training distributions. The learning algorithm used was Doug’s momentum and momentum was set to 0. These learning parameters are extremely conservative and were chosen solely to ensure that representation learning was stable and effective. The initial range of weight randomizations was set to 0.1. I then saved the weights of the sensory input layer to the hidden layer so they could be used and frozen in the category learning network (described below).

**Category learning**

**Stimuli.** The stimulus distributions for the category learning environments are identical to the information-integration category distributions used in the human experiment in Chapter 6 with the addition of two rule-based category types. For the context of training the model, I reduced the scale in a normalized space so that the category distributions fit more clearly within the training environment space, which was normalized from 0 to 1 along the two axes. The four category distributions are shown in Figure 3.

There are four category environments: information-integration along the positive axis (II-Positive), information-integration along the negative axis (II-Negative), rule-based along the x-axis (RB-X) and rule-based along the y-axis (RB-Y). Consistent with nomenclature throughout the category learning literature, the two II categories can be distinguished based on a boundary that requires both sensory dimensions and the two RB categories can be distinguished by a boundary along only one of the two sensory dimensions.
Figure 3. Input distributions for the category learning model. Red and blue dots reflect members of different categories.

The category distributions were created by sampling a bivariate Gaussian distribution using the \texttt{mvnorm} function in the \textit{MASS} package in R. I sampled for a single category using normalized coordinates (0-1) and then manipulated and rotated that distribution to create all other categories (Figure 3). Thus, each category type—and each individual category—has the identical information in terms of relative variance, covariance, and the relationship between the two categories in terms of overlap. These distributions were initially created in a normalized, arbitrary stimulus space that varied from 0 to 1 along both dimensions and then normalized to
a .3 to .7 space to better fit within the space of the four training distributions (Independent, Types A, B, C, and D). Table 2 shows the means, variance, and covariance of these distributions.

To create the categories, first a single category was sampled in this arbitrary space with high variance along the x-axis (RB-Y, Category A). This category distribution was then mirrored across the x-axis to create the other category (RB-Y, Category B). This first category was rotated 45 and 135 degrees around the normalized .5, .5 midpoint to create the II-Negative-A and II-Positive-A categories, respectively. Both of these categories were then mirrored across the space with a slope matching the optimal boundary between the categories. The slope was set at -1 for II-Negative and +1 for II-Positive. The stimulus coordinates were also converted into an unnormalized Gaussian across ten units for input into the model, using the same method as for the training stimuli.

Table 2. Category Distribution Information for Category Learning Stimuli

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean (X, Y)</th>
<th>Variance (X, Y)</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-Positive: Category A</td>
<td>(0.57, 0.45)</td>
<td>(0.0059, 0.0053)</td>
<td>0.0030</td>
</tr>
<tr>
<td>II-Positive: Category B</td>
<td>(0.45, 0.57)</td>
<td>(0.0053, 0.0059)</td>
<td>0.0030</td>
</tr>
<tr>
<td>II-Negative: Category A</td>
<td>(0.55, 0.57)</td>
<td>(0.0059, 0.0053)</td>
<td>-0.0030</td>
</tr>
<tr>
<td>II-Negative: Category B</td>
<td>(0.43, 0.45)</td>
<td>(0.0059, 0.0053)</td>
<td>-0.0030</td>
</tr>
<tr>
<td>RB-X: Category A</td>
<td>(0.58, 0.51)</td>
<td>(0.0026, 0.0086)</td>
<td>0.00028</td>
</tr>
<tr>
<td>RB-X: Category B</td>
<td>(0.42, 0.51)</td>
<td>(0.0026, 0.0086)</td>
<td>0.00028</td>
</tr>
<tr>
<td>RB-Y: Category A</td>
<td>(0.49, 0.58)</td>
<td>(0.0086, 0.0026)</td>
<td>-0.00028</td>
</tr>
<tr>
<td>RB-Y: Category B</td>
<td>(0.49, 0.42)</td>
<td>(0.0086, 0.0026)</td>
<td>0.00028</td>
</tr>
</tbody>
</table>

Model architecture. To have the category learning network make use of the internal hidden representations that I trained via the autoencoder in the representation learning phase, the autoencoder network is a part of the category learning network. This entire architecture can be seen in Figure 2. As during the autoencoder training, the 20-unit sensory input layer is connected to a 10-unit perceptual representation hidden layer and there is a small amount of noise in the encoding of the stimuli in the input layer (range = 0.1). To test how each of the training distributions impacted learning, the individual connection weights between the input layer and
the hidden layer were separately loaded in and frozen for each training distribution. This perceptual representation hidden layer is then connected to a category decision output layer, which has two units. The target activation of these two units reflects the category identity of the stimuli. Category A stimuli had a target response layer activation of “1 0” and category B stimuli had a target response layer activation of “0 1.” This response layer reflects a category decision made by a motor response (button push) during human category learning tasks.

**Training procedure.** I trained the category learning network using two distinct training paradigms. The first training paradigm (batched learning), like the representation learning paradigm, is conservative and stable in order to most clearly illustrate effects of representation on learning. This batched learning paradigm is meant to be a more abstracted version of the model’s behavior to better understand the constraints of existing representations on the learnability of categories. This model is not meant to perfectly reflect human behavior or the way in which humans update their representations during learning. The second paradigm (online learning) is a closer approximation to experience in actual human category learning experiments, as the network updates its weights after each stimulus presentation. For all four category types (II-Negative, II-Positive, RB-X, and RB-Y), the presentation of the exemplars occurred in a permuted order, meaning that they were randomly presented without replacement. Both models were trained and tested on all 200 stimuli from each category learning environment (100 stimuli per category). The same set of stimuli were used for training and testing the model.

**Training Paradigm 1: Batched learning.** I trained the category learning network for 10 epochs of batched learning with a learning rate of 0.01 with a learning algorithm of Steepest Descent. By batched learning, I mean that I first presented all 200 stimuli to the model and then the model updated its connection weights. I measured the error at each epoch and test accuracy
after each epoch to evaluate the model’s ability to learn the categories. For each of the training distributions (Independent, Types A, B, C, and D) and each of the categories (II-Positive, II-Negative, RB-X, and RB-Y), I trained 10 simulated subjects to get a sense of the variability in the behavior of the model.

*Training Paradigm 2: Online learning.* I trained the category learning network on non-batched, online learning with a single presentation of each of the category stimuli (200 exemplars total, 100 per category). I trained the network with a learning rate of 0.5 with a learning algorithm of Steepest Descent. By online learning, I mean that I presented each stimulus one by one, with the model updating its connection weights after each stimulus presentation. I measured the error across the 200 exemplars and test accuracy only after all 200 exemplars were seen to evaluate the model’s ability to learn the categories. For each of the training distributions (Independent, Types A, B, C, and D) and each of the categories (II-Positive, II-Negative, RB-X, and RB-Y), I trained 10 simulated subjects to get a sense of the variability in the behavior of the model. This online learning paradigm may better match what happens with human learning, as the weights are updated after each trial during learning. However, I am not comparing the model’s performance directly with the time course of human learning.

**Predictions**

I designed the model architecture, training environments, and category distributions to understand how prior experience with perceptual information can warp representations in such a manner that it will impact category learning. Thus, there are specific predictions for how the model should be able to learn individual categories that is dependent on the nature of the training environment.
**Independent training environment.** This training environment reflects a situation where the two sensory dimensions are independent. I predict that this will result in the model demonstrating equal learnability across the four category learning environments. Because the training environment presents equal variance along the x and y dimensions as well as equal spacing across the positive and negative axes, the category learning model should be able to make any distinction between categories that is required. This training environment serves as a sort of control environment to understand how the model will respond to information that should not result in a bias constraining subsequent category learning.

**Type A and Type B training environments.** These two training environments are similar in that they both involve a correlation between the x and y dimensions in the two-dimensional space. However, Type A has a negative correlation and Type B has a positive correlation between the two dimensions. I predict that this difference in directional relationship will drastically alter the behavior of the model. I predict that these two training environments will demonstrate the largest difference between the learnability of the II-Negative and II-Positive category environments. Because the model will experience a high variability along the negative axis for the Type A training distribution, as a consequence, the model should be better at distinguishing between stimuli that fall along the negative axis. This enhanced ability to discriminate between stimuli along the negative axis should lead to enhanced performance for learning II-Positive categories, because the two categories can be distinguished along the negative axis. In contrast, training with the Type A distribution should be detrimental for learning II-Negative categories, as it requires distinguishing between stimuli along the positive axis, which is not emphasized by a Type A distribution. Thus, for the Type A distribution, I predict that the II-Positive categories will have a higher accuracy and lower error than II-
Negative categories. Because the distribution is exactly the opposite for the Type B distribution, I predict that the II-Negative categories will have a higher accuracy and lower error than II-Positive categories for the Type B distribution.

**Type C and Type D training environments.** In contrast to the Type A and Type B environments, the Type C and Type D training environments do not reflect a correlation between the two dimensions and, instead, have higher variance along one of the input dimensions (x/y) than the other. The rationale for the predictions of the model’s behavior is similar to the other training environments. For Type C, for example, because the variance in the distribution is higher along the y-axis, this should result in the model dedicating more energy and space in the hidden representations to representing this y-axis dimension. As a result, stimuli that fall along the y-axis dimension will be easier to discriminate after training. This should lead to enhanced performance for learning RB-Y categories, which can be distinguished by the y-axis alone. In contrast, training with the Type C distribution should be detrimental for learning RB-X categories, which require distinguishing the categories along the x-axis, which is not emphasized by the Type C distribution. Thus, for the Type C distribution, I predict that the RB-Y categories will have a higher accuracy and lower error than the RB-X categories. Because the Type D distribution has higher variance along the x-axis than the y-axis, for the Type D distribution, I predict that the RB-X categories will have higher accuracy and lower error than the RB-Y categories.

The most important predictions for the model’s category learning behavior involve the two II categories for the Type A and Type B distributions and the two RB categories for the Type C and Type D distributions. I expect to find the largest differences in category learning for these comparisons. However, I will train the model on all four categories for each of the training
distributions to understand how the model approaches category learning challenges that are completely aligned with the training distributions, completely misaligned with the training distributions, or somewhere in between.

Results

Training Paradigm 1: Batched learning

The accuracy of the model across epochs is shown in Figure 4 and the error of the model across epochs is shown in Figure 5. Note that in these and all other results figures shown below, there are error bars reflecting standard error of the mean at a group level on the plot, though they are often too small to be visible. Additionally, performance for each of the 10 simulated subjects is shown as transparent, colored dots and the mean performance across the 10 subjects is shown as a solid black dot.

Accuracy. I quantified accuracy as the percent of category exemplars for which the model met the target activation criterion of 0.5. For the control Independent distribution, all four category types were learned very quickly to a very high degree of accuracy. Importantly, after the first epoch, the accuracy of all four category learning types is the same, consistent with an interpretation that for the Independent distribution, the model is unbiased in category learning.
In line with my predictions for Type A and Type B distributions, one of the II category types was learned very well and the other was not learnable at all (50-60% accuracy throughout all epochs). Critically, the model’s performance depended on which distribution type had been experienced and which category distinction was being learned. For Type A representations, the II-Positive categories were learned very well, the RB-X and RB-Y categories were learned at an intermediate level, and the II-Negative categories were learned relatively poorly. For Type B representations, this pattern is reversed; the II-Negative categories were learned very well, the RB-X and RB-Y categories were learned at an intermediate level, and the II-Positive categories were learned very poorly. It is clear in the case of the Type A and Type B distributions that what
is able to be learned is dependent on the nature of the existing hidden unit representations, which were altered based on experience. The Type A representations preserve the variance that is relevant for the II-Positive category learning type and discard the variance that is relevant for the II-Negative. This thus reflects an alignment between the Type A representations and the II-Positive categories and a misalignment between the Type A representations and the II-Negative categories. They Type B representations, in contrast, preserve the variance that is relevant for the II-Negative categories and discard the variance that is relevant for the II-Positive categories, reflecting an alignment between Type B and II-Negative categories and a misalignment between Type B and II-Positive categories.

For Type C and Type D, there are not quite as clear differences between the category types. However, the broad patterns of accuracy for RB-X and RB-Y categories are in line with my predictions. For Type C representations, RB-Y categories have higher accuracy than RB-X categories starting in the second epoch, with the two II categories showing an intermediate level of accuracy. For Type D representations, the RB-X categories have higher accuracy than the RB-Y categories also around the second epoch. Here the pattern of accuracy for the two II categories is slightly more variable, with the accuracy overlapping more with the two RB categories. The Type C representations are well-aligned with the RB-Y categories, as the representations preserve the variance along the y-axis. In contrast, the Type D representations are well-aligned with the RB-X categories, as the representations preserve the variance along the x-axis.

Error. When examining the error patterns of the model for different types and categories, the results are exactly in line with my predictions. As expected, for the control Independent distribution, all four categories had comparable error patterns and all had steep reduction in error.
For Type A and Type B distributions, the error is highest, and remains high, for the category distinction that conflicts with the trained distributions. For Type A distribution, the II-Negative categories are difficult for the model and have the highest error, whereas the II-Positive categories have the lowest error. For the Type B distribution, the II-Positive categories are difficult for the model and have the highest error, whereas the II-Negative categories have the lowest error.

The Type C and Type D distributions also have results that are in line with my predictions. For Type C, the categories with the lowest error are the RB-Y categories and the
categories with the highest error are the RB-X categories. For Type D, this pattern is reversed and RB-X categories have the lowest error and RB-Y categories have the highest error.

**Training Paradigm 2: Online learning**

For this training paradigm, I examine performance of the model after exposure to all of the 200 exemplars and show the performance of the model after all exemplars, rather than performance across time. This assesses how well the model has learned the categories after it has seen all of the exemplars and reflects a situation where a human has encountered all category exemplars and is tested on them. The accuracy of the model after 200 stimulus presentations is shown in Figure 6 and the cumulative error of the model across the 200 stimulus presentations is shown in Figure 7.

**Accuracy.** I quantified accuracy as the percent of category exemplars for which the model met the target activation criterion of 0.5. As with the batched learning paradigm, using the hidden unit weight connections based on the Independent distribution led to very high accuracy among the four category distribution types. There were no large differences in accuracy among the category types.

As with the batched training paradigm, the Type A and Type B distributions had the clearest pattern of accuracy that demonstrates that the nature of experience impacts how categories will be learned. For Type A, the accuracy was highest for II-Positive, lowest for II-Negative, and at an intermediate level for RB-X and RB-Y. For Type B, the accuracy was highest for II-Negative, lowest for II-Positive, and at an intermediate level for RB-X and RB-Y.

Again with Type C and Type D, the pattern is slightly murkier. Figure 7 demonstrates that Type C and Type D category learning is, on average, slightly more successful but more variable than Type A and Type B. For Type A and Type B, there was some incredibly difficultly
to learn one of the categories, whereas for Type C and Type D, all of the categories were learned to a relatively high accuracy. As before, I focus on the relative differences between the category conditions. These relative patterns are in line with my predictions. For Type C, the ultimate accuracy of this training paradigm was higher for RB-Y categories than RB-X categories, with II-Negative and II-Positive closer to the RB-X level. For Type D, the RB-X categories have higher accuracy than RB-Y, with II-Negative and II-Positive closer to the RB-Y levels.

Figure 6. Model accuracy after the single pass through all 200 category exemplars for training paradigm 2 for the five distribution types and four category learning environments. Individual runs of the model are shown as colored dots, the mean performance is shown as a black dot, and the error bars reflect SEM.

Error. As with the batched training paradigm, the error of the model after the 200 weight updates is broadly in line with what I predicted. The control, Independent distribution had very
low and invariable error patterns across the four category learning types. For Type A, the error is highest for II-Negative and lowest for II-Positive. For Type B, this pattern is reversed. For Type C, the RB-X has higher error than RB-Y and for Type D, this pattern is reversed. The pattern of error is also much more variable than during batched training.

![Figure 7](image)

**Figure 7.** Model error after the single pass through all 200 category exemplars for training paradigm 2 for the four distribution types and four category learning environments. Individual runs of the model are shown as colored dots, the mean performance is shown as a black dot, and the error bars reflect SEM.

**Discussion**

The current set of simulations demonstrate that the nature of experience in a sensory environment can shape the representations of this information in a way that drastically impacts category learning behavior. Depending on the nature of the representations, which were shaped
by experience, some category learning problems become easily learnable, whereas others are completely unlearnable. This network model demonstrates that consideration of perceptual processing and the acknowledgement of the constraints that the perceptual system and existing representations place on the learning system are critical for understanding the mechanisms at play during perceptual category learning. The nature of the learning problem may vary substantially depending on the perceptual representations of the sensory information.

These simulations can provide insight into how the sensory world is maintained in perceptual representations and influenced by long-term statistical regularities of the natural world. This model can apply directly to studies of human perceptual learning, including from chapters 4 and 6.

In Chapter 4 (Roark & Holt, 2019), we demonstrated that human participants were quickly able to learn II-Positive categories based on center frequency and modulation frequency acoustic dimensions. However, they had poorer performance while learning II-Negative categories based on these same dimensions. Though the statistical sampling of these categories is identical, just rotated differently in the space, the problem was much more difficult for the II-Negative categories than the II-Positive categories. This pattern of human learning was not predicted by an existing literature that has focused more on whether categories require one or multiple dimensions and could be classified as ‘rule-based’ or ‘information-integration’. However, this pattern of results is in line with the current simulated behavior with the Type A training distribution (Figure 8). Even the pattern of learning for the two RB category types is similar between the Type A distributions and chapter 4 (Roark & Holt, 2019). This is compelling evidence that the difference in behavior observed in chapter 4 may have been due to a difference
in the perceptual representation, which might be driven by a negative correlation relationship in long-term statistical experience between center frequency and modulation frequency.

Figure 8. Pattern of accuracy of the model with the Type A training distribution (top) compared with the humans from Chapter 4 (Roark & Holt, 2019; bottom).
In Chapter 5, we found that in human participants’ similarity-based representations of spectral modulation and temporal modulation, there was no one-to-one relationship between the perceptual dimensions and these two acoustic sensory dimensions. That is, the way these two dimensions are represented likely reflects some combination of both dimensions, rather than a single dimension alone. In Chapter 6, we observed that this interaction in perceptual representations likely impacted the ability to learn the II-Positive and II-Negative categories. We observed a general pattern wherein the II-Positive categories were easier to learn than the II-Negative categories, regardless of short-term structured experience. This pattern is also in line with the Type A distribution’s behavior in the three training paradigms in this chapter (Figure 9). However, this pattern is also consistent with the model’s performance with the Type C or Type D distributions. In the simulated data, the differences between the two II conditions were small, but the general direction was II-Positive had higher accuracy than II-Negative for both Type C and Type D distributions.

We did not examine human RB category learning in the spectral modulation-temporal modulation acoustic space. The main factor of not including these category learning conditions was time. In future experiments, it would be interesting to examine the two RB category learning conditions in this space to better understand how to map the current simulation results to human behavior.
Figure 9. Pattern of performance of the model with the different training distributions on the II-Negative and II-Positive categories (top) compared with the humans in the II-Negative and II-Positive naive conditions from Chapter 5 (bottom).
The current model used relatively simple training spaces that are clearly much more abstract than the way sensory information is presented in the real world. While there was a small amount of noise in the input to the model to reflect a small amount of perceptual noise in the encoding process, there was no noise in the actual distributions. Future expansion of this model should include a simulation of the kind of variability and noise that exists in real-world sensory environments. Additionally, different kinds of relationships in the space should also be tested to make clear predictions about how the many different kinds of relationships (rather than just independence or a perfect correlation) can be represented by the model to affect behavior. Finally, this model is restricted to a two-dimensional space. The world beyond simple experiments has many more dimensions, some of which are relevant, others irrelevant, some present and varying, some rarely present and stable. A future iteration of this model should seek to understand how multiple dimensions may be represented independently and, in conjunction, what the effects on higher-level cognition might be.

The goal of the current neural network model simulation experiment was to illuminate the nature of perceptual representations using a theoretical perspective. This formulation of the model allows us to make predictions about how perceptual representations of sensory information influence category learning mechanisms. What is clear from this investigation is that the nature of learning problem can vary drastically based on the network’s existing hidden unit representations. The same is likely true with human learning. While much of the human perceptual category learning research has used simple, verbalizable dimensions that are likely represented independently both neurally and in mental representations, it is a much more difficult and interesting problem to understand what happens when perception is not so straightforward. The current model demonstrates that the influence of perception on cognition can be drastic.
References


Chapter 8: Conclusion

This dissertation has examined the relationship between sensory information, perceptual representations, and cognitive processes during category learning. Chapter 2 demonstrated that the distributions of perceptual categories and the task used to learn the categories affect how that learning proceeds and the subsequent learned category representations. Chapter 3 demonstrated that the way in which variability is experienced within a trial can affect learning for rule-based, but not information-integration category distributions. Chapter 4 showed that sensory dimensions are not necessarily equivalent to perceptual dimensions and there may be perceptual biases which influence cognitive processing based on existing representations. Chapter 5 demonstrated that prior exposure to highly structured acoustic regularities has limited effects on similarity-based perceptual representations. Chapter 6 expanded this investigation by demonstrating that passive exposure to structured information may have subtle, short-lived effects on how people approach categorization problems, but does not impact overall learning. Chapter 7 introduced a proof-of-concept neural network model that demonstrated that entrenched existing representations that reflect some sort of regularity induce perceptual biases that have specific effects on category learning.

In sum, this body of work makes several conclusions about the nature of perception and cognition. First, the nature of the stimulus environment and task during training can drastically alter the course of category learning. Second, existing perceptual representations can have a strong impact on the sensitivity to short-term statistical regularities. Third, existing perceptual representations seem to have a stronger effect on the learnability of categories than any short-term bias induced by passive exposure to statistical structure in sensory information.
This work demonstrates that it is necessary to bring perception into the picture to understand the mechanisms of perceptual category learning. Historically, researchers have made an implicit assumption about the relation between sensory information in the external world and internal, perceptual representations of that world. Specifically, there is an assumed orthogonal, one-to-one relationship between the sensory world and perceptual representations.

This research opens doors for future research regarding how humans are able to construct stable, yet flexible, representations of the sensory environment based on regularities in long-term experience and how they engage with novel environments that may not contain the same statistical regularities. Overall, this research demonstrates that pre-existing representations, built up through a life-time of experience, can have profound effects on how people structure new knowledge and how they interact with information in new environments.