UNDERSTANDING THE IMPACT OF DIFFERENTIATION AND UNITIZATION ON THE PERCEPTUAL FEATURES LEARNED DURING CATEGORY TRAINING

Andrew Thomas Hendrickson

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the

requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____

Robert L. Goldstone, Ph.D.

_____

Peter M. Todd, Ph.D.

_____

Thomas A. Busey, Ph.D.

_____

Robert M. Nosofsky, Ph.D.

October 15, 2015

ACKNOWLEDGMENTS

I have been extremely fortunate throughout my life to have the support of enough people to populate a small country and am glad to have the opportunity to acknowledge a few of them here. First, I want to thank my parents, Kathy and Chris Hendrickson, for their unconditional love and support. They have been my role models in every aspect of my life and I cannot thank them enough for everything they have done for me and taught me. I dedicate this thesis to them. I would also like to thank my brothers, Tommy and Peter, who somehow forgave me for sitting on them when they were young and became not only fantastic brothers, but two of my closest friends as well.

If I have been fortunate to have family that have been friends, then I have been equally as fortunate to have friends that have been family. I would like to thank everyone from Pittsburgh who not only tolerated my strange interests, but encouraged me to do what I love: Acadia, Chris, Emily, Jay, Jen, Jenna, Jess, Kevin, Kevin, Lucy, Matthew, Rich, and Vanessa. I would also like thank the great friends I met in Bloomington: David, Ji, Mike, and Winter for showing me the ropes in room 290; Dave, David, Josh, Lisa, Paulo, Rick, Sam, Sarah, Seth, Skyler, Thomas, and Tom for lively debates and thoughtful suggestions in the Goldstone/Todd lab; and Alejandra, Andre, Ashley, Chris, Erin, Francisco, George, Gordon, Hilary, Jared, Jared, Jess, Jette, Kate, Rachel, Richard, Stephen, and Viridiana for making it fun. Not to be forgotten, I would not have finished this thesis without the support and friendship of the Computational Cognitive Science group in Adelaide: Amy, Dan, Emma, Heidi, Jess, Lauren, Keith, Sean, Steve, Valance, and Wai Keen. Though this

iv

list is long, it is certainly not complete and I thank everyone who has supported me through the years.

I first learned that I was interested in cognitive science while studying in Pittsburgh, and I cannot thank John Anderson, Patricia Carpenter, Ken Kotovsky, Marsha Lovett, David Plaut, and Mark Stehlik enough for helping me find my interest in computational modeling of human behavior. Additionally, I owe a deep debt of gratitude to everyone in the psychology and cognitive science programs at Indiana University, which is where I learned how to be a cognitive scientist. Rob Goldstone, thank you for teaching me the craft of seeing empirical questions everywhere. Peter Todd, thank you for teaching me to always be thinking of the big picture. Rob Nosofsky and Tom Busey, thank you for your insightful comments and feedback while serving on my committee and the rigorous computational modeling you taught me in your classes. John Kruschke, thank you for teaching me to love statistics. And thank you to Patricia Crouch who always was there to help put out my fires and did it with a smile.

Finally, I would never have accomplished what I have without the love and support of my wife, Diana. Without her none of it would have been possible.

Andrew Thomas Hendrickson

UNDERSTANDING THE IMPACT OF DIFFERENTIATION AND
UNITIZATION ON THE PERCEPTUAL FEATURES LEARNED DURING
CATEGORY TRAINING

Perceptual representations are a foundational aspect of all cognitive processes that involve input from the external environment. Yet there is ample evidence that these perceptual representations are altered by experience in systematic ways. This work focuses on understanding how perceptual representations are modified through two perceptual learning processes, differentiation and unitization, in the context of category learning. First, we review the empirical evidence for perceptual learning with a focus on the evidence for unitization and differentiation processes in the context of category learning. This section also includes a discussion of the role of differentiation and unitization learning processes in four computational models of perceptual learning. Second, we present a series of four experiments that measure the change in perceptual representations after learning category structures designed to promote differentiation and unitization in perceptual learning. Third, we investigate the impact of these category structures on the features inferred by a model that incorporates both differentiation and unitization perceptual learning processes. Fourth, we develop a modeling framework to directly compare the fit of computational models that assume different perceptual representations to the empirical results. Finally, we conclude by considering the implications and limits of these results.

Robert L. Goldstone, Ph.D.

Peter M. Todd, Ph.D.

Thomas A. Busey, Ph.D.

Robert M. Nosofsky, Ph.D.

CONTENTS

CHAPTER 1

Literature Review

1.1 Overview

This chapter is divided into four sections. In the first section, we review how perceptual learning has been historically conceptualized as a critical aspect of learning. Then, we review the empirical evidence for three processes of perceptual learning: differentiation, unitization, and imprinting. In the third section, we review four computational models of perceptual learning and discuss how those models implement the processes of differentiation and unitization. Finally, we discuss the interaction between differentiation and unitization learning within the context of a single learning environment as a motivation for the empirical work in the next chapter.

1.2 Perceptual Learning

A common intuition is that the perceptual system is stable across time. We believe that how we see, hear, touch, taste, and smell a particular object today is how we will perceive it tomorrow, next month, or next year because if the object does not change then our perception of that object is always the same. Yet our experience is not stable across time because our perceptual system is shaped by learning.

Helmholtz (1910/1924) was one of the first psychologists to point out the importance

of learning to perception. He proposed that the "unconscious inferences" performed by the visual system to process input, including binocular disparity between the eyes, could be learned over the course of development. Helmholtz's early theoretical work on binocular vision that led to his theories of perceptual learning were subsequently supported by adaptation studies 60 years later (Nelson, 1977).

The debate over what mechanisms underlie perceptual learning through experience have existed since before the cognitive revolution. J. J. Gibson and Gibson (1955) transformed the debate by clearly identifying two theoretical frameworks for how perception might be shaped by experience. The first proposes that perception is a "creative" process in which initially meager representations are enriched through additional experience. This framework anticipates many of the exemplar-based models of learning in categorization (Nosofsky, 1984, 1986; Medin & Schaffer, 1978) and memory (Shiffrin & Steyvers, 1997; Raaijmakers & Shiffrin, 1981). These models are enrichment-based because improved processing of new items is based entirely on enriching a stored set of exemplars through additional experience. The enrichment framework is characterized by experience leading to an improvement in inference but not processing (J. J. Gibson & Gibson, 1955).

The second theoretical framework is founded on improving perceptual processing by differentiating aspects of the perceptual representation. J. J. Gibson and Gibson (1955) believe this learning framework would include learning to perceive dimensions of variation and properties of stimuli not previously processed and increase the discriminability of existing stimulus features. The enriching and differentiating frameworks are not mutually exclusive. ALCOVE (Kruschke, 1992), a successful and prominent exemplar-based categorization model, includes mechanisms to enrich the set of exemplars as well as connection weights that are adjusted to minimize error which often increases discriminability.

In this work we will mostly constrain ourselves to discussing empirical results and com-

putational modeling through the framework of differentiating. Enrichment processes alone do not seem adequate to explain the systematic changes across all sensory domains including visual acuity (Fine & Jacobs, 2002), auditory processing (Puel, Bonfils, & Pujol, 1988; Girard, Collet, Bouchet, & Pernier, 1994), tone detection (Metherate & Weinberger, 1990; Weinberger, Javid, & Lepan, 1993), phoneme detection (Liberman, Harris, Kinney, & Lane, 1961), haptic localization (Weinstein, 1968), taste (Curtis, Stevens, & Lawless, 1984; Green et al., 1996), and smell (Blanes-Vidal et al., 2009; Overbosch, 1986). Instead, we will focus on understanding two critical perceptual learning processes, differentiation and unitization, and how they improve the utility of perceptual representations (E. J. Gibson, 1969) in the context of category learning.

1.3 The processes of perceptual learning

Perceptual learning is not characterized by one pattern of change; it consists of multiple processes that alter perceptual processing (Goldstone, 1998). In this section we will review the evidence for three perceptual learning processes: differentiation, unitization, and imprinting. These processes vary in how quickly they influence perceptual features (Kami & Sagi, 1993) as well as their effect on the set of perceptual features. Differentiation and unitization are both processes that alter representations slowly but have opposite effects on representations: differentiation separates perceptual feature values and dimensions making them easier to discriminate from each other while unitization combines configurations of existing perceptual features into new features. Both processes occur across a wide array of stimuli and learning tasks but evidence for them mostly is found when they change the perceptual representations to be more useful for solving the current task of the learner (E. J. Gibson, 1969). Imprinting is similar to the process of unitization but occurs much quicker, often on the order of a few presentations of a stimulus or feature (Kami & Sagi,

1993). All three perceptual learning processes occur across a wide array of stimuli, learning tasks, and measures of perceptual learning.

1.3.1 Differentiation

The most commonly studied pattern of perceptual learning is Differentiation in which existing perceptual features are adjusted based on experience. This pattern is characterized by a slow change in discriminability that takes place over the course over many trials but the set of perceptual features does not change.

Improvements in perceptual processing that arise from prolonged exposure to discrimination tasks have been found in many low-level visual tasks (Fine & Jacobs, 2002). Perceptual tasks that show improvement include the detection or discrimination of visual gratings (De Valois, 1977; Fiorentini & Berardi, 1980, 1981; Mayer, 1983), stimulus orientation judgment (Ahissar & Hochstein, 1996; Dosher & Lu, 1998; Shiu & Pashler, 1992; Vogels & Orban, 1985), motion direction discrimination (Ball & Sekuler, 1982, 1987; Sekuler & Machamer, 1983), texture discrimination (Ahissar & Hochstein, 1996; Karni & Sagi, 1991; Kami & Sagi, 1993), time to perceive random dot stereograms (Ramachandran & Braddick, 1973), stereoacuity (Fendick & Westheimer, 1983), Vernier discrimination tasks (Beard, Levi, & Reich, 1995; Bennefl & Westheimer, 1991; Fahle & Edelman, 1993; Kumar & Glaser, 1993; McKee & Westhe, 1978; Saarinen & Levi, 1995), and object recognition (Furmanski & Engel, 2000). These improvements in processing due to perceptual learning can occur even for a signal that is task-irrelevant and below the threshold of conscious awareness if it is co-located with the cover task (Watanabe, Náñez, & Sasaki, 2001).

A traditional hallmark of visual perceptual learning effects has been that they are highly specific to stimulus properties including stimulus size, orientation, direction, and many others. More importantly to some models of perceptual learning, the improved processing

due to learning does not transfer to other retinal locations (Shiu & Pashler, 1992; Karni & Sagi, 1991; Kami & Sagi, 1993; Ahissar & Hochstein, 1997; Fahle, 1994). This has led many researchers to claim that perceptual learning is the result of improved processing of low-level perceptual information that is specific to retinal locations (Zhaoping, Herzog, & Dayan, 2003; Karni & Sagi, 1991; Kami & Sagi, 1993; Adini, Sagi, & Tsodyks, 2002; Teich & Qian, 2003) despite the evidence against changes in neural sensitivity of cells located in low-level visual cortex (Schoups, Vogels, Qian, & Orban, 2001; Ghose, Yang, & Maunsell, 2002).

The empirical evidence for the location-specificity of perceptual learning was found in tasks where training is limited to a single retinal location and transfer is assessed in a novel location (Mollon & Danilova, 1996). Xiao et al. (2008) theorized that novel locations might receive fewer perceptual and cognitive resources than locations that had received any form of training and this might lead to the lack of transfer. To test this they developed a new experimental paradigm with contrast discrimination training in one location and Vernier discrimination training in another location. This training task resulted in near perfect transfer of learning to trained locations but no transfer to novel locations (Xiao et al., 2008). Transfer of learning across locations has also been found when both tasks are orientation discrimination tasks (T. Zhang, Xiao, Klein, Levi, & Yu, 2010; J.-Y. Zhang et al., 2010; Dosher, Jeter, Liu, & Lu, 2013) that previously did not show transfer to untrained locations (Shiu & Pashler, 1992; Ahissar & Hochstein, 1997).

Differentiation and category learning

Perhaps the most striking example of learning categories shaping perception is language learning. Language perception depends on the process of perceptual differentiation for children to learn the basic phonemes of their native language. Most languages consist of

approximately 40 phonemes, the smallest units of speech that change the meaning of a word, drawn from a possible set of over 800 consonant and vowel sounds (Ladefoged & Disner, 2012). Infants as young as one-month old are able to discriminate between many of these phonemic groups from what will become their native language (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) but also between phonemes that do not exist in their native language (Lasky, Syrdal-Lasky, & Klein, 1975; Werker & Lalonde, 1988). Adults improve the discrimination between phonemes in their own language but lose the distinctiveness across phonemic boundaries their language does not possess (Miyawaki et al., 1975). Differentiation of phoneme boundaries for native language perception in adults is so strong that many observers do not make above chance discriminations within a phonemic category but make near perfect between category judgments (Liberman, Harris, Hoffman, & Griffith, 1957). The differentiation process that changes perception from weakly discriminating all phonemic differences to a native-language specific set of strong phonemes begins to occur within the first year of language experience (Werker & Tees, 1984; Werker & Lalonde, 1988). However, even in adulthood there is evidence the differentiation learning process for language is still at work because people can learn to discriminate new phonemic categories with extensive experience (Pisoni, Aslin, Perey, & Hennessy, 1982).

Differentiation induced by category learning can be specific to a stimulus dimension. Goldstone (1994) demonstrated differentiation that was specific to stimulus dimensions by using two-dimensional stimuli and category structures that highlighted one dimension or another. The stimuli consisted of 16 color squares that varied across two dimensions, size and brightness, shown in Figure 1.1. The two critical conditions for assessing perceptual learning due to category training were the groups that learned a single-dimension category structure. Participants in these conditions received category training in which only one stimulus dimension was relevant for categorization. For one condition the size value was

relevant for category membership (the 8 stimuli with size values less then the midpoint of the size dimension were assigned to be in one category), and the other half were trained on a category structure with brightness as the category relevant dimension. The largest change in perceptual discriminability was an increase in discriminability between pairs of stimuli that were assigned to different categories. Critically, these pairs of stimuli were not the same in the size-relevant and brightness-relevant category conditions. For the size-relevant condition, the largest increase occurred between pairs of stimuli that came from the second and third columns of Figure 1.1, and the brightness-relevant condition had a large increase in pairs of stimuli from the second and third rows of the figure.

This advantage for discriminations between two stimuli from different categories relative to discriminations between stimuli from the same category has been found across a wide range of perceptual domains (Harnad, 1987; Goldstone & Hendrickson, 2010; Newell & Bülthoff, 2002). Furthermore, this advantage emerges as a result of category training (Livingston, Andrews, & Harnad, 1998; Özgen & Davies, 2002; Goldstone, Lippa, & Shiffrin, 2001; Goldstone, 1994). Yet the degree to which it can be attributed to changes in perceptual representations is still under debate and perhaps the most notable example of this has been the perception of color. People do not discriminate evenly across variation in hue, people are worse at discriminating between two shades of red then same variation in hue if it spans the border between red and orange (Wright, 1947; Roberson & Davidoff, 2000). This has led to the claim that category learning has shifted our perceptual representations of hue such that hues that are in the same category are more similar to each other than hues in different categories (Harnad, 1987) and learning a new color category induces this change in representation (Özgen & Davies, 2002; Özgen, 2004). Bornstein and Korda (1984) proposed that this advantage for processing between-category discriminations might be due to the presence of labels and not the result of changes in perceptual process-

ing. This view is strengthened by results that suggest verbal interference may reduce or eliminate the advantage of between-category discriminations of color (Roberson & Davidoff, 2000; Winawer et al., 2007; Hanley & Roberson, 2011). However, experience does shape perceptual representations of color because extensive training leads to massive increases in speed and discriminability but eliminates the between-category advantage of discrimination for novices (Witzel & Gegenfurtner, 2013, 2015).

Differentiation is not limited to separating stimulus values along a single dimension because category learning also leads to differentiating two stimulus dimensions. Faces are generally perceived as a composite percept (Gauthier, Curran, Curby, & Collins, 2003) but Goldstone and Steyvers (2001) found category training of special faces could result in learning to extract feature dimensions from faces. These special faces were constructed by fusing four faces in different proportions (Figure 1.2) to create a two-dimensional space of face stimuli. Training on categories with only one relevant dimension, horizontal or vertical category boundaries, produced learning of that feature dimension but also the other feature dimension. Participants transferring to a new categorization task were faster to learn when the new boundary was orthogonal to the original boundary than a diagonal boundary. This result is particularly surprising because a diagonal boundary would preserve some category relevance for the originally relevant feature dimension, potentially improving performance on the transfer condition. However, the results are instead consistent with the category training resulting in the perceptual system learning a perceptual representation in which each stimulus consists of two feature dimensions: one dimension consisting of the category relevant variation in faces and the other dimension capturing the remaining variation. Given that perceptual representation, a transfer categorization task in which the second dimension becomes category relevant might be easier to learn than a task that requires a combination of both dimensions, since learning a category structure that only requires a single dimension

is faster then a structure that requires two dimensions (Shepard, Hovland, & Jenkins, 1961).

1.3.2 Unitization

Unitization is the perceptual learning process that is the opposite of differentiation; unitization results in learning perceptual features that are composed of existing features instead of learning to separate perceptual features or feature values. Unitized features are processed by the perceptual system as a new individual feature rather than merely a conjunction of individual features that are separately processed.

Shiffrin and Lightfoot (1997) show evidence that a simple visual search task can lead to learning perceptual features that combine information from multiple independently-varying component parts (also reported in (Czerwinski, Lightfoot, & Shiffrin, 1992)). The learning of new perceptual features was characterized by a dramatic decrease in the amount of time to respond per distractor using stimuli that were constructed by arranging three segments inside a rectangle such that no single segment would uniquely identify a stimulus (Figure 1.3). Transfer tasks revealed that this improvement was specific to the stimuli being trained but not unique to the groupings of targets and distractors. After 50 sessions of training, a number of transfer tasks were assessed including shuffling the targets and distractors and introducing new sets of targets and distractors created from the same segments. In the novel stimuli condition (experiment 2), participants were transferred to a search task with novel stimuli that were created from the same component segments. Search speed per distractor reverted to the same speed as the first session of training. In the new conjunction condition, participants were transferred to a search task with familiar target and distractor stimuli, but a new pairing of targets and distractors such that a conjunction of three segments was necessary to identify the target item (as in the initial training). Search speed per distractor in the new conjunction condition did not show any difference from the end of the initial

training (experiments 2b and 3). In the new single segment condition, participants were transferred to a search task with familiar target and distractor stimuli, but a new pairing of targets and distractors such that a single segment was sufficient to identify the target item. Search speed per distractor in the new single segment condition did not show any difference from the end of the initial training and was identical to the transfer in the new conjunction condition (experiment 3). These results suggest that during the initial training participants were learning a feature representation in which each whole stimulus was represented as an individual perceptual feature. The improvement in perceptual processing of the target stimuli was not specific to specific distractor stimuli and the gains were not lost when the distractors were changed.

Not all visual search tasks with these stimuli led to learning new perceptual features. When the initial training consisted of target stimuli that could be uniquely identified by a single feature, initial search speed was much faster and improved quicker. However, when the sets of targets and distractors were changed to require a conjunctive search (experiment 5), all search speed gains were lost and participants did not learn any faster than the initial conjunctive search training from previous experiments. This complete lack of transfer did not occur when switching from conjunctive search to single segment search (or a different conjunctive search), suggesting that only conjunctive search led to forming new perceptual features that map onto whole stimuli in this task.

Perceptual unitization does not require structured task instructions that highlight the conjunction of components, exposure to stimuli containing structure across components can produce unitization. In a passive viewing task with stimuli that contained some components that reliably co-occurred and others that varied independently, Fiser and Aslin (2001) found participants rated test configurations containing the co-occurring components as more familiar than those containing independently-varying components. A similar pref-

erence for looking at test configurations containing co-occurring components was found in nine-month-old infants (Fiser & Aslin, 2002).

Developing visual expertise often requires learning to recognize and quickly process complex configurations of features as an overall perceptual unit. One domain of visual expertise that has been extensively studied is face processing. It has been hypothesized that the primary perceptual representation of faces are a single perceptual feature that encompasses the entire configuration of the face (Bradshaw & Wallacei, 1971) because inverted faces (Yin, 1969; Barton, Keenan, & Bass, 2001; Freire, Lee, & Symons, 2000) and components of faces (Tanaka & Farah, 1993; Farah, Wilson, Drain, & Tanaka, 1998) are processed much less efficiently than faces relative to other non-face stimuli. Yet this advantage does not seem to be limited only to faces. Other domains of visual expertise also show strong advantages for processing upright and intact stimuli but only for experts in those domains: fingerprint experts are much better at processing complete prints than individual components and novices are poor at both (Busey & Vanderkolk, 2005), dog experts are much better at processing upright dog images relative to inverted dog images (Diamond & Carey, 1986), and extensive training with artificial stimuli leads to better processing of intact stimuli (Gauthier & Tarr, 1997; Gauthier, Williams, Tarr, & Tanaka, 1998; Gauthier & Tarr, 2002). Over the course of development, the processing of faces as a perceptual feature is learned through experience (Carey, Diamond, & Woods, 1980).

Neuroscience evidence also indicate neurological signals that were considered indicative of processing faces but not other stimuli (Kanwisher, McDermott, & Chun, 1997; Bentin, Allison, Puce, Perez, & McCarthy, 1996) are involved in processing complex visual stimuli for experts but not novices (Bukach, Gauthier, & Tarr, 2006). This has been shown for fingerprint experts (Busey & Vanderkolk, 2005), car experts (Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier et al., 2003), bird experts (Gauthier et al., 2000; Tanaka &

Curran, 2001; Gauthier et al., 2003) dog experts (Tanaka & Curran, 2001) and expertise developed from laboratory training (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier & Tarr, 2002; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002).

Unitization and category learning

As with differentiation, perceptual unitization can result from learning categories and different perceptual features can be learned from different category structures defined across the same stimuli. Pevtzow and Goldstone (1994) found people were much faster to detect a combination of line segments that when present together were diagnostic of category membership (Figure 1.4) than a combination of line segments presented equally often but not diagnostic of the category. As in Shiffrin and Lightfoot (1997), none of the individual line segments was diagnostic by itself, all three needed to be present to correctly categorize each stimulus. By using two orthogonal category structures – the horizontal and vertical categories shown in Figure 1.4 – the category relevant and irrelevant clusters were the same clusters but changed their relevancy based on the category structure. Across both category training conditions, responses were fastest when the whole and probe stimuli matched and both contained a category relevant part. If either the whole or probe stimulus contained a familiar, but category irrelevant part then judgements were slower. The familiar but category irrelevant parts were still faster than judgments containing novel parts. This suggests that though each prototype was shown equally often in the horizontal and vertical category training conditions, different features were learned because different parts were relevant for categorization.

The complexity of the unitized feature has a direct impact on how long it takes to learn. Goldstone (2000) assessed this learning by comparing the improvement in reaction time to make categorization judgments across category structures that varied the complexity of a

All		One	
Category 1	Category 2	Category 1	Category 2
ABCDE (5x)	ABCDZ	ABCDE	ABCYE
	ABCYE		
	ABXCE		
	AWCDE		
	VBCDE		

Table 1.1: The two category structures from Experiment 5 from Goldstone (2000). The All category structure encourages learning the unitized feature ABCDE that is composed of five independent segments. The One category structure encourages learning a single feature that corresponds to either the D or Y segment.

perceptual feature that would be relevant for categorization. These category structures, shown in Table 1.1, were defined across stimuli that were line drawings that resembled the outline of half an eggshell and each stimulus was composed of five connected line segments with the endpoints joined by a curve as shown in Figure 1.5. The *ALL* category structure required all five segments to be present together to correctly assign a stimulus to category 1. This structure was hypothesized to promote learning a relatively large perceptual feature that combines the information from all five segments (ABCDE) to create a one-to-one mapping between the feature and category membership. This is in contrast to the *ONE* category structure in which the information from one segment is required to make a correct categorization. Goldstone (2000) found that the speed of categorization responses to the ABCDE stimulus were slower in the *ALL* condition but improved much more than in the *ONE* condition. The responses in the final session of categorization were faster in the *ALL* condition than would be expected by processing all five segments at the same rate as the *ONE* condition which suggests in the *ALL* condition segments are not processed independently. This co-active architecture of segment processing is consistent with the

process of learning a unitized perceptual feature.

1.3.3   Imprinting

Not all perceptual learning processes occur slowly over many presentations, the imprinting perceptual learning process produces radical changes to perceptual features after a few presentations. The result of imprinting is similar to the unitization process and changes the set of perceptual features by adding new features to the perceptual representation, but unlike the process of unitization, imprinting results in very fast learning because it occurs in environments where features are relatively easy to parse.

Mere exposure to a correlation between components across stimuli can lead people to imprint perceptual features that include all correlated components (Austerweil & Griffiths, 2011). Given instructions to freely "investigate" stimuli printed on index cards for up to five minutes and then select which test stimuli belong with the inspection set, participants who inspected stimuli with correlated components only selected test stimuli with correlated features while participants who inspected stimuli with independently varying components selected both correlated and independent stimuli at test. However, changing the stimuli structure is not necessary to change the features encoded by imprinting. Lin and Murphy (1997) found that given a single presentation of a set of stimuli, the speed and accuracy of detecting changes to a component of those stimuli could be easily manipulated by presenting a cover story that highlighted the functional aspect of components. Across two cover stories that drew attention to different components of the same stimuli, changes to components that were relevant to the cover story were more quickly and accurately detected than irrelevant components.

Imprinting can change perceptual features very early in development. Three month old infants prefer to look at stimuli that violate the Gestalt principle of good continuation

(Quinn & Bhatt, 2005; Quinn, Burke, & Rush, 1993). By merely exposing them to 12 stimuli that share a component that violates good continuation (Figure 1.6), the looking time preference is reversed and infants prefer to look at a test stimulus with good continuation rather than the familiar component (Quinn et al., 2006; Quinn & Schyns, 2003). This reversal in looking time preference suggests that viewing the 12 stimuli containing the component led infants to form a feature that corresponds to the component. This would lead to faster processing of the component and infants are less interested in inspecting it in the test phase.

Category structures can also lead to imprinting and the final set of perceptual features is highly sensitive to the order in which categories are learned. The perceptual features learned earlier influence the subsequent features that are learned. Schyns and Murphy (1994) found that participants who learned individual component features (A and B) first were more likely to represent a subsequently learned composite feature (AB) as an explicit combination of the components (A+B) than participants who learned the composite feature first. Two component-first conditions were trained by exposing participants in the first training stage to 3D geometric stimuli containing a component part (either A or B) and in the second stage to stimuli with a composite part (AB). The composite part consisted of both component parts, A and B, that were adjacent and always in the same spatial arrangement. The composite-first condition did not have a first stage of training and consisted only of exposure to stimuli that contained the composite part (AB). When asked to circle the "parts" of novel stimuli containing the composite AB part after all training, participants from the component-first conditions were more likely to circle the A and B parts separately than participants in the composite-first condition. The participants in the composite-first condition most often drew a single circle around both A and B parts.

However, features learned later in training are not necessarily only defined in terms of

Category training			Test
1	2	3	X-Y as XY
X	Y	XY	88%
XY	X	Y	19%

Table 1.2: The category structure for the two conditions (rows) and test phase results of Experiment 2 from Schyns and Rodet (1997). All category training phases consisted of learning to discriminate between stimuli that contained a part (or two parts) and stimuli with no parts present. The category structure columns indicate which part was present during the training phase for each condition. During the XY category training, the X and Y parts were always presented adjacent and in the same spatial orientation. The test phase column indicates the percent of responses that select the XY category label responses to the test stimulus in which X and Y are present but not adjacent.

the earlier features. In a second experiment, Schyns and Murphy (1994) found that when learning the composite AB feature in the first training session, participants were able to learn the component features in subsequent training. After first being exposed to stimuli containing the composite AB part, participants who later were exposed to stimuli containing only one component part (either A or B) were more likely to circle a component part at test than the composite AB feature (though some participants circled both component and composite features). A control group of participants who were exposed to the composite AB part and then an unrelated part C circled the composite AB feature at test.

Schyns and Rodet (1997) expanded on this result and demonstrated that category learning can lead to different perceptual features even when the same set of category structures are eventually learned across all conditions. This result is most clear in Experiment 2 which consisted of four phases: three training phases where participants learned a category structure (with two different orders) and a test phase that was identical across conditions.

During each category training phase participants were asked to learn to categorize grey cell-like stimuli that consisted of one or two dark colored parts (X and Y) embedded in low-frequency visual noise. In each training phase one category consisted of stimuli with no parts present and the other category was composed of stimuli containing the parts shown in Table 1.2. Critically, all three category structures were identical across the two conditions (indicated by rows in the table) but the order the categories were presented in was different.

In the components-first condition (top row) participants learned the component parts (X and Y) before the two parts were presented together (as XY). Participants in this condition learned that X and Y were separate parts and when they saw X and Y in the same stimulus (but not adjacent) they categorized the test stimulus into the XY category. In the composite-first condition (bottom row) participants learned the composite parts first (XY) and later learned the component parts (X and Y). This led to learning a different set of parts in which the composite XY was considered to be a single part along with the component parts and did not consider the test stimulus containing a non-adjacent X and Y parts to be in the XY category.

An asymmetry of the effect of order between component and composite features is seen across both Schyns and Murphy (1994) and Schyns and Rodet (1997). In these short category learning experiments, training orders that induce component feature learning first impede the subsequent learning of composite features if they contain the component features. The reverse is not true: learning the composite features early in training does not seem to impede learning the component features later.

The degree to which imprinting is a fundamentally different process than unitization is unclear. The empirical evidence for imprinting is highly similar to the evidence for unitization, including learning features containing multiple co-occurring (and correlated) components (Fiser & Aslin, 2001, 2002) and category relevant feature learning (Pevtzow &

Goldstone, 1994; Goldstone, 2000). Yet these two processes seem to operate on different time scales with imprinting leading to adding fully formed features and unitization resulting in a slow emergence of new perceptual features across many presentations. Though interesting, this issue is not developed further in this work though it is an avenue for future study.

1.4 Computational models of perceptual learning

In the following section we review four models of perceptual learning. The focus of each review is on a few critical questions: what does the model learn, how does the model represent features, does the model account for the role of category learning in perceptual learning, and can the model account for the differentiation and unitization perceptual learning effects.

1.4.1 The augmented Hebbian reweighting model and the integrated reweighting theory

Petrov, Dosher, and Lu (2005) outline the augmented Hebbian reweighting model (AHRM), a biologically-inspired computational model of perceptual learning that relies on updating weights to account for perceptual learning effects. The model is a three-layer neural network with an input layer that corresponds to the current input stimulus, an intermediate layer of units, and an output layer that has units that correspond to the possible responses. The connection between the input layer and the intermediate layer is fixed with connections that do not update with learning. These connections involve a series of complex, nonlinear transformations that take as input the activation of many input units to compute the activation of 35 units in the intermediate layer. If more than one region of interest is being processed in the input layer, a set of 35 intermediate layer units is added for each region. The same non-linear transformations are applied to each region of interest

and weights are not shared across regions. The connection weights between units in the intermediate layer and output layer are updated using the Hebbian reweighting rule (Hebb, 1949). This learning rule increases the connection weight between units when they are both active. The activation at the output layer is derived by computing the weighted sum across all intermediate units. A defining characteristic of AHRM is that learning in one region of interest does not transfer to any other regions, a property that was considered a defining characteristic of visual perceptual learning (Ahissar & Hochstein, 1997; Karni & Sagi, 1991). However, recent evidence has emerged that some amount of perceptual learning does transfer across locations (T. Zhang et al., 2010; J.-Y. Zhang et al., 2010; Xiao et al., 2008). AHRM cannot account for these transfer effects and has been extended to create a new model that does transfer learning across regions.

Dosher et al. (2013) outline the integrated reweighting theory (IRT) of perceptual learning, a computational model based on AHRM that includes a mechanism to capture the transfer of perceptual learning across locations. In the case of a task with only one region of interest in the stimulus, the IRT model is identical to AHRM. When more than one region of interest is present in the stimulus representation, each region has a region-specific set of intermediate units with connection weights that are updated based only on the activation in that region. However, the intermediate layer of the IRT model also contains a set of region-independent units whose activation is driven by the input to all regions. The connection weights of the region-independent units are also updated by Hebbian learning rules but these weights are influenced by all regions. The activation at the output layer is derived by computing the weighted sum across both region-independent and region-specific intermediate units. The IRT model was able to account for the results of an experiment that showed transfer of perceptual learning across visual locations (Dosher et al., 2013).

Both of these models based on updating the connections between perceptual units and

output units are capable of accounting for the differentiation process of perceptual learning. Discriminating the orientation of Gabor patches is a common experimental paradigm that shows differentiation learning (Shiu & Pashler, 1992) and was the first paradigm to show transfer of perceptual learning across retinal regions (T. Zhang et al., 2010). The AHRM was able to account for the improvement in orientation discrimination performance with feedback (Petrov et al., 2005) and without feedback (Petrov, Dosher, & Lu, 2006) and the IRT was able to account for the transfer of improvement in orientation discrimination performance to new retinal locations (Dosher et al., 2013). Even though these models focus on processing orientation, they are capable of accounting for a wide range of differentiation processes. If there exists any intermediate units that have different activations to the two stimuli the model is learning to differentiate, the Hebbian learning rule will strengthen the connection weights between those intermediate units and the response options. This updating of connections will eventually lead to those intermediate units having more influence in the response of the model and leading to improved discrimination (Hebb, 1949), a hallmark of the differentiation perceptual learning process.

These reweighting models can only account for the process of unitization in some experimental paradigms but not others. A network with a single layer of Hebbian connection weights can learn a category structure or discriminate between two groups of stimuli where a conjunction of units are necessary to make a correct response (Rosenblatt, 1958). Therefore, assuming that the AHRM had an intermediate unit that would differentially activate for each component, the model could learn to discriminate the target stimulus from Shiffrin and Lightfoot (1997) (Figure 1.3) and the category structure in Experiment 5 from Goldstone (2000) (Table 1.1), which are both defined by conjunctions of components. However, a single-layer of Hebbian connections cannot learn an exclusive-OR structure (Minsky & Seymour, 1969), and therefore these models could not learn the category structure in

Experiment 1 from Goldstone (2000) (Table 1.5, Figure 1.5) or the category structure from Pevtzow and Goldstone (1994) (Figure 1.4). A second layer and an additional set of learned connections would be necessary for the model to account for learning an exclusive-OR category structure (McClelland, Rumelhart, & Group, 1986). Furthermore, the reweighting models are not sensitive to correlations or relationships between intermediate units and thus would not be sensitive to the difference between stimuli constructed from independently varying components and those with correlations between components (Austerweil & Griffiths, 2011; Fiser & Aslin, 2001, 2002).

1.4.2   Bayesian chunk learner

The second computational model of perceptual learning we consider is the Bayesian chunk learner (Orbán, Fiser, Aslin, & Lengyel, 2006, 2008, BCL). The BCL is a generative Bayesian model that infers a set of perceptual features from visual stimuli composed of discrete parts. The BCL assumes stimuli are generated by sampling from a hierarchical structure: each stimulus is composed of one or more perceptual features, and each perceptual feature is composed of one or more stimulus parts. The parts are pre-specified primitives of the model that can occur in different locations of a stimulus and a part can be associated with more than one latent chunk. Perceptual features are assumed to be latent chunks or clusters that are associated with one or more parts.

The BCL is a Bayesian model and therefore can be characterized by a likelihood function and prior distributions across the parameters of the model. The BCL uses a rather simple likelihood function that assumes each latent chunk is present in each stimulus independent of other chunks and the parts are also independent of each other, conditional on which chunks are present. The strength of the model is in defining the relationship between parts and latent chunks and the prior on chunks. For each latent chunk i there is a non-zero

chance that chunk will activate a part p in a given stimulus. That probability is specified by $Sigmoid(w_{ip})$ where w_{ip} is the influence of chunk i and part p and is estimated for each pair of i and p and comes from a exponential distribution. The prior distribution of the number of latent chunks across all stimuli is the geometric distribution. This distribution does not pre-specify the number of features, instead it allows for a potentially infinite number of features. This aspect of the BCL separates it from the IRT model which has exactly 35 perceptual features for each region, regardless of the complexity of the stimuli. The geometric prior distribution on the number of features is biased to prefer representations with fewer features but allows the number of features to increase if they are more likely given the stimuli.

The BCL can account for some unitization patterns but none in experimental paradigms that require category information. Features in the BCL are simply clusters of stimulus parts that are statistically related, and therefore the model is good at learning features when the parts are correlated or share a co-occurrence structure across stimuli. Orbán et al. (2008) fit the BCL to the data from experiments with features constructed from chunks of parts (Fiser & Aslin, 2001, 2002) and found it accounted well for the pattern of generalization. It is likely the BCL would be able to infer the correlation structure between parts of other results including the difference features learned in the correlated and independent components conditions of Austerweil and Griffiths (2011) and the correlation of segments across stimuli in Shiffrin and Lightfoot (1997) (Figure 1.3). Because the model treats trials as exchangeable the BCL would have trouble accounting for order effects that produce different patterns of feature learning (Schyns & Rodet, 1997; Schyns & Murphy, 1994) though this issue is solved in a similar Bayesian model (Austerweil & Griffiths, 2013).

The BCL cannot account for the process of differentiation because the component parts of the model are pre-defined based on the stimuli. Furthermore, the similarity between parts

are treated as either identical or completely mismatching, with no continuum between the two values. This characterization completely eliminates any adjustment of similarity or discriminability of parts without significantly altering the model.

1.4.3 Conceptual and perceptual learning by unitization and segmentation model (CPLUS)

The CPLUS model (Gerganov, Grinberg, Quinn, & Goldstone, 2007; Goldstone, 2003; Goldstone, Gerganov, Landy, & Roberts, 2009) is a computational model that unifies two perceptual learning processes, unitization and differentiation, into a single perceptual learning framework that is modulated by category information. The model consists of a connectionist network with three layers: an input layer, a hidden layer, and an output layer. The input and hidden layers are connected with only feed-forward weights and the hidden and output layers are connected by a set of weights that project activation in both directions. The input and output layers are standard connectionist representations: the input layer consists of one node for each pixel in the stimulus and the output layer consists of nodes that map onto categorical responses. The hidden layer consists of a set of feature-detector units and the number of units is pre-determined before training the model (Austerweil & Griffiths, 2011). A feature-detector unit consists of the same number of nodes as the input layer and each node in the feature-detector unit has a weighted connection to exactly one node in the input layer. The connection weights from one input node to the corresponding nodes in different feature-detector units are updated independently and not shared. To categorize a new stimulus the activation of that stimulus at the input layer is propagated to all nodes in the hidden layer via weighted connections, and then the activation of the hidden nodes is propagated to the output nodes via weighted connections.

Learning the connection weights in the CPLUS model is driven by two processes: a

category learning process and a segmentation learning process. The category learning process is a supervised learning rule that updates all connections between layers. First, the activation of each node is determined by a combination of the weighted connections to the input layer and the weighted connection to the category label. Both sources of information drive the activation of the nodes in the hidden layer. Next, the weight connections between the hidden and output layers are updated by the delta learning rule (Widrow & Hoff, 1960) with a pre-defined learning rate. Then, the connections between the input and hidden layers are updated using the competitive learning rule with leaky learning (Rumelhart & Zipser, 1985). This learning rule updates the connection weight to change the activation of a hidden node to be more similar to the activation of the node in the input layer it is connected to. Not all connections are updated with the same learning rate. The connections to the nodes in the hidden unit whose activation is most similar to the input layer are updated with the highest learning rate and connections to all nodes in other hidden units are also updated but with a smaller learning rate. This results in all nodes in the hidden layer becoming more similar to the input node they are connected to, but the nodes in the most-similar feature-detecting unit update more. Since the activation of the hidden nodes was partially influenced by the activation of and connections to the output layer, the category label contributes to the activation of the hidden units and indirectly influences the learning of the connections between the input layer and hidden units. This occurs despite not including any back-propagated error-driven learning signals (Rumelhart, Hinton, & Williams, 1988). The category learning process updates all connection weights after each stimulus is presented and weights are updated in parallel.

The second learning process in CPLUS is the segmentation learning process. This is an unsupervised learning rule that updates the connections from the input layer to the nodes in the hidden layer but not the connections to the output layer. Unlike in the category

learning process where feature-detecting units compete with each other and all nodes in a unit are updated with the same learning rate, in the segmentation learning process, the nodes from each feature-detector that are connected to the same input node compete with each other. For each input node, the connection weight of the hidden node with the most similar activation is updated to increase the similarity between them and all other weights are not changed. This learning rule works, but learns features that are not psychologically plausible (Goldstone, 2003). In order to account for perceptual constraints, the similarity between an input node and a hidden node is influenced by the similarity of nearby nodes. The result of combining these two learning rules into a single model of perceptual learning is a model that learns to segment stimuli into perceptual features with a bias to learn features that are relevant for categorization.

The CPLUS model can account for both unitization and differentiation processes when category feedback is present. Goldstone (2003) shows CPLUS reliably learns different sets of unitized features that correspond to the category relevant combinations of line segments when trained on the two category structures from Pevtzow and Goldstone (1994) (Figure 1.4). These results suggest the strong influence of category information on connection weight learning and the competition between feature-detector units for each category would lead CPLUS to learn appropriate unitized features in other experimental contexts with category feedback (Goldstone, 2000; Schyns & Rodet, 1997, e.g.). Furthermore, though it has not been explicitly trained on such a stimulus set, CPLUS can account for the category-induced differentiation learning effects because it is designed to learn features that discriminate between two groups of stimuli that vary, even if the groups are difficult to discriminate. However, CPLUS makes a strong representational commitment to explain differentiation processes because CPLUS accounts for perceptual learning by extracting discrete features. This implies the model does not represent feature dimensions, though the connectionist

architecture of CPLUS allows it to make smooth generalizations between features that may approximate feature dimensions if enough discrete features are present.

It is unclear if CPLUS would reliably account for either unitization or differentiation processes without category feedback. There are at least two issues that hinder CPLUS from accounting for either perceptual learning process without categorization. First, the competitive learning rule in the segmentation learning process is not leaky, which can lead to a suboptimal distribution of weights (Rumelhart & Zipser, 1985). Second, the segmentation learning process by itself has no pressure to group pixels that occur in separate locations in the same stimulus to the same feature unit. Both of these issues are not inherent to the framework of CPLUS, but it suggests that CPLUS might rely heavily on the perceptual constraints of similarity to learn the appropriate features in the correlated segments condition from Austerweil and Griffiths (2011) and Quinn et al. (2006) and be unable to learn conjunctions of non-adjacent components that become unitized features (Schyns & Murphy, 1994; Fiser & Aslin, 2001, 2002, e.g.).

1.4.4 Non-parametric Bayesian framework for learning flexible feature representations (NBFF)

Austerweil and Griffiths (2011) outline a framework for Bayesian computational models that infer perceptual features from pixel representations of stimuli as well as an explicit mapping of which features are in which stimuli. Being Bayesian models, the NBFF models have the same general structure as the BCL model (Orbán et al., 2008), where the model specifies a generative process for producing stimuli and category labels and the model infers likely parameter values, including the perceptual features, given a set of stimuli and labels. The mathematics of a NBFF model are reviewed in depth in the Learning Stimulus Features chapter where it is applied to the stimuli in Experiments 1 through 4 from the next chapter,

but a high-level overview is presented here for comparison with the other perceptual learning models.

The NBFF framework defines the relationships between stimuli, which are composed of features, and features, which are composed of stimulus pixels. Stimuli are generated from a potentially infinite set of features, though a finite number of features are inferred for any given set of stimuli. The set of features present in each stimulus is modeled as a matrix of binary indicators with a row for each stimulus and a potentially infinite number of columns indicating features. Austerweil and Griffiths (2011) use the Indian Buffet Process (Ghahramani & Griffiths, 2005; Griffiths & Ghahramani, 2011) to define a prior distribution across matrices that contain an unknown number of columns. The IBP generates matrices that have a potentially infinite number of columns but a finite number of columns that contain at least one non-zero element, ensuring that a finite number of features are inferred for a given set of stimuli (Griffiths & Ghahramani, 2011; Austerweil & Griffiths, 2013). Features are composed of a finite set of pixels that are very likely to be present in the stimulus if the feature is present. The construction of features depends on the prior beliefs about the relationship between pixels within a feature. The most basic assumption, which defines a class of NBFF models that make this assumption, is that all pixels are present in each feature independent of all other pixels, but that assumption can lead to inferring psychologically implausible features (Goldstone, 2003; Goldstone et al., 2009). This is the assumption made by the BCL model (Orbán et al., 2008). However, the equation that defines the relationship between pixels in NBFF models can be modified to include spatial constraints between pixels that lead to inferring plausible psychological features.

The NBFF framework for models of perceptual learning can account for unitization processes both with and without category information. Austerweil and Griffiths (2011) show NBFF models extract the appropriate unitized features when presented with the

stimuli from a number of experimental paradigms that show unitization processes without category information (Shiffrin & Lightfoot, 1997; Czerwinski et al., 1992; Austerweil & Griffiths, 2011). Austerweil and Griffiths (2013) extend the NBFF framework to include category information and category order effects. This expanded model treats category labels as additional pixels of each stimulus, a method used previously in other computational models of category learning (Love, Medin, & Gureckis, 2004)[1]. This extension of the NBFF model successfully inferred the category-specific perceptual features found in Pevtzow and Goldstone (1994) However, NBFF models do not account for all unitization process effects. As we discuss in depth in the Learning Stimulus Features chapter, the NBFF models do not successfully infer a complete set of features for exclusive-OR category structures (Goldstone, 2000).

The original NBFF framework (Austerweil & Griffiths, 2011) cannot account for order effects that lead to perceptual unitization (Schyns & Rodet, 1997; Schyns & Murphy, 1994) because it assumed all trials were exchangeable and thus did not preserve order information. Austerweil and Griffiths (2013) address this issue by switching from using a Gibbs sampler (Geman & Geman, 1984), which assumes trials are exchangeable, to a particle filter (Gordon, Salmond, & Smith, 1993) to perform inference in the model. Particle filters preserve order information because they do not assume trials are exchangeable and have been used in other Bayesian models of category learning to account for order effects (Sanborn, Griffiths, & Navarro, 2010). An NBFF model using particle filters to perform inference but not the standard Gibbs sampler was able to infer two different sets of features that match the features people circled in the two category order conditions of Schyns and Rodet (1997) (Austerweil & Griffiths, 2013). Austerweil and Griffiths (2013) also expanded the NBFF framework to infer a limited set of psychologically plausible transformations of features in

---

[1]An alternative modeling framework that infers different features for each category is also presented in Austerweil and Griffiths (2013) but not discussed here.

stimuli. Some of these transformations, such as the translation of features, are consistent with low-level perceptual learning transfer effects (Xiao et al., 2008; T. Zhang et al., 2010; J.-Y. Zhang et al., 2010; Dosher et al., 2013) but many of the other transformations, such as scaling, rotation (orientation), and reflection, might be inconsistent with the specificity of perceptual learning (Zhaoping et al., 2003; Karni & Sagi, 1991; Kami & Sagi, 1993; Adini et al., 2002; Teich & Qian, 2003) and more consistent with features at higher levels of cognition (Hahn, Chater, & Richardson, 2003; Rips, 1989).

The relationship between the NBFF framework and the differentiation process is similar to the CPLUS model (Goldstone, 2003). NBFF models make the strong claim that features are discrete units and not dimensions of variation. Therefore any differentiation processes that are accounted for by the NBFF are due to learning features that differentiate between the groups and not from differentiating values along a continuum.

1.5   Differentiation and Unitization

Of the four models outlined in the previous section, one model only accounts for the differentiation processes (ICAN), one model only accounts for the unitization process (BCL), but the remaining two models include both differentiation and unitization processes. Both of these models, the CPLUS and the NBFF framework, do not have individual mechanisms for differentiation and unitization perceptual learning processes that operate in different contexts. Instead, both processes emerge from a single perceptual learning mechanism that produces behavior consistent with unitization or differentiation depending on the stimuli and category structures. These models suggest a novel empirical prediction: evidence for differentiation and unitization can emerge from the same stimuli given two category structures that encourage learning different features.

Three studies suggest differentiation and unitization processes can occur across the

29

same set of stimuli. First, Shiffrin and Lightfoot (1997) found evidence of unitization when searching for a conjunction of three segments and they also found some improvement when searching for a target identified by a single segment. Though this improvement suggests a differentiation process because discrimination improved, Shiffrin and Lightfoot (1997) argue these improvements are more consistent with search strategy improvements since the segments were highly discriminable. Second, Goldstone (2000) found evidence of unitization processes in the *ALL* condition, which required all five segments to make a category judgment, but learning in the *ONE* condition, which required only one segment to make a judgment, suggests a differentiation learning process. The improvement in the reaction time for the *ONE* condition is consistent with differentiation in one sense because the stimuli containing the Y segment become easier to discriminate from the stimuli containing the D segment (Table 1.1) but it is also consistent with learning perceptual features that correspond to segments Y and D. The degree to which it is differentiation or unitization depends on if participants viewed the eggshell stimuli more like a face, a single stimulus to be differentiated (Goldstone & Steyvers, 2001), or a configuration of line segments to be unitized (Shiffrin & Lightfoot, 1997). Finally, across two conditions with different training order, Schyns and Rodet (1997) found evidence that people could learn features that correspond to component parts (X and Y) as well as a composite of those parts (XY). However, learning occurred in fewer than 40 training trials, suggesting that both component features and the composite feature were easy to extract. This may be due to imprinting features and not differentiation or unitization. In the next chapter we test these predictions in a series of four experiments.
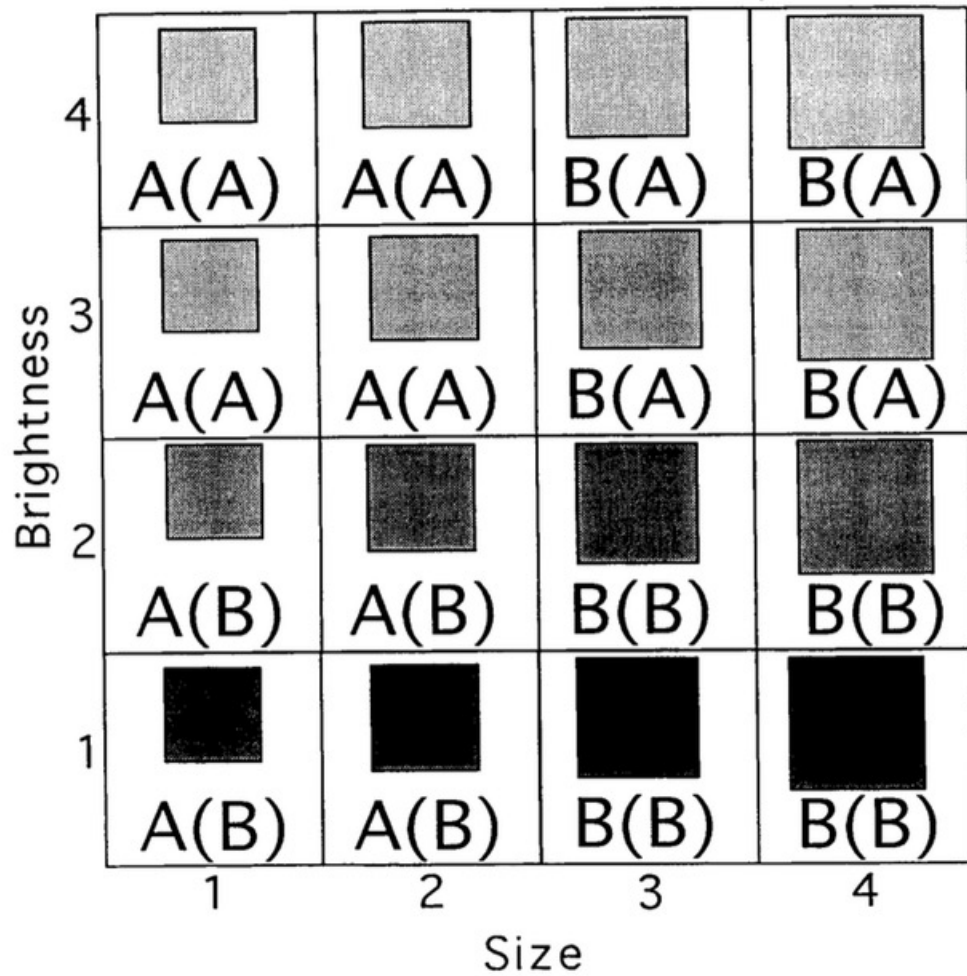
Figure 1.1: The 16 stimuli from Experiment 2 in Goldstone (1994). The stimuli varied across two dimensions: the size and the brightness of the square. The first letter in each cell indicates the category membership of that stimulus when categorizing stimuli by size and the second letter indicates the category membership when categorizing by brightness.
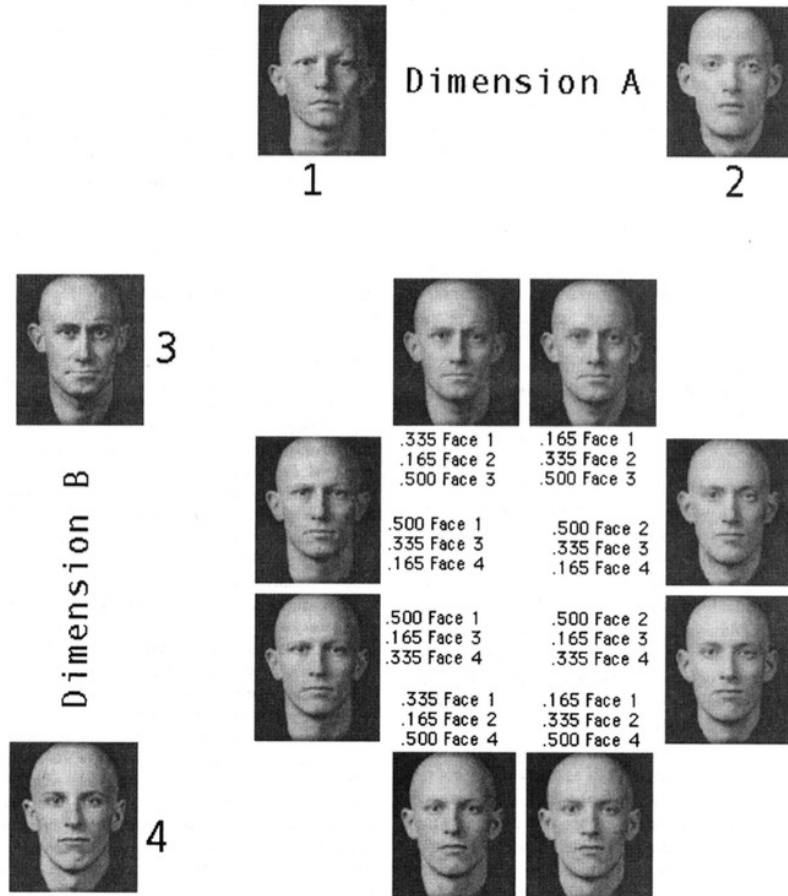
Figure 1.2: The stimuli from Experiment 1 in Goldstone and Steyvers (2001) were created by morphing four faces (labeled 1, 2, 3, and 4) to create 8 stimuli that varied in two dimensions (labeled A and B). Across the columns faces vary along dimension A and faces further left are more similar to face 1 than faces to the right. Across the rows faces vary along dimension B and faces in higher rows are more similar to face 3 than faces in lower rows.
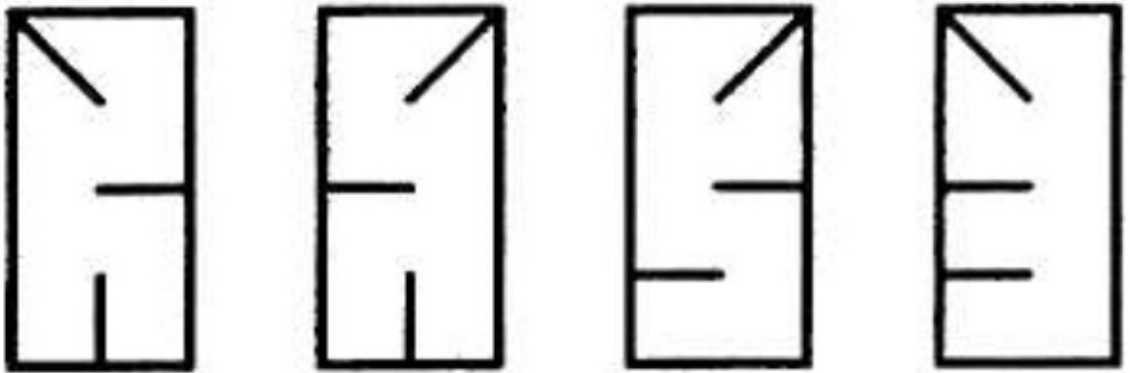
Figure 1.3: Stimuli from Shiffrin and Lightfoot (1997). One stimulus was randomly selected to be the target items and the other three were distractor items. All stimuli are constructed from three segments inside a rectangle and none of the four stimuli has a unique segment that is not shared with at least one other stimulus. Across multiple training sessions using these stimuli, visual search response times were consistent with learning perceptual features that corresponded to whole stimuli.
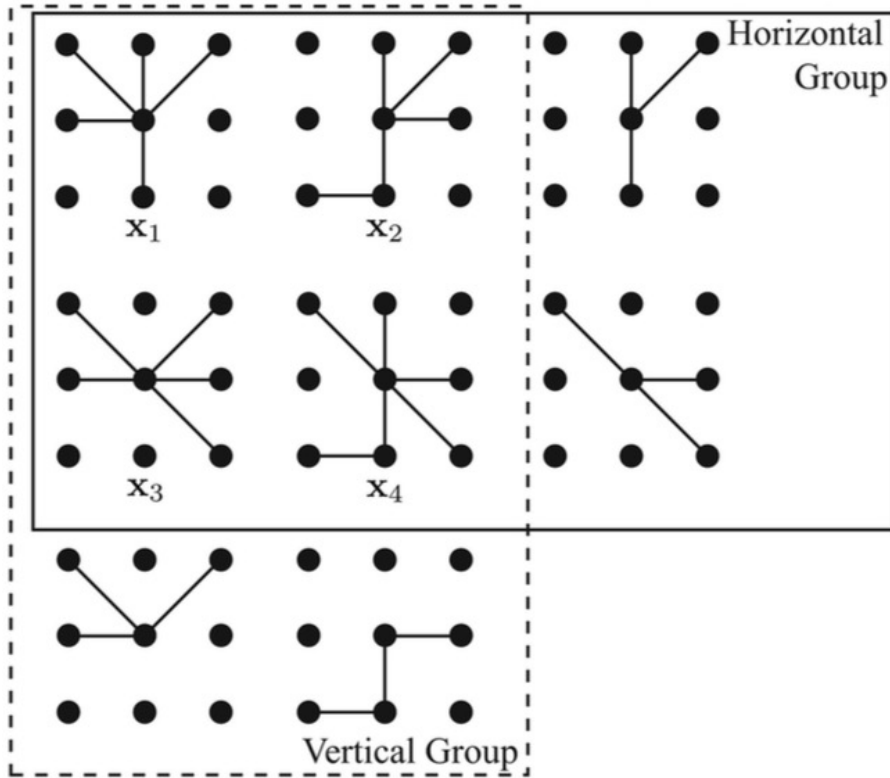
Figure 1.4: The four prototype stimuli (labeled x_1 through X_4) and the component parts used to create them in Pevtzow and Goldstone (1994). When learning the horizontal category structure the two top stimuli were in the same category and the two bottom were in another. The component parts inside the Horizontal Group box were the relevant parts for this category structure. When learning the vertical category structure the two left stimuli were in the same category and the two right were in another. The component parts inside the Vertical Group box were the relevant parts for this category structure. This figure is reproduced from Austerweil and Griffiths (2013).
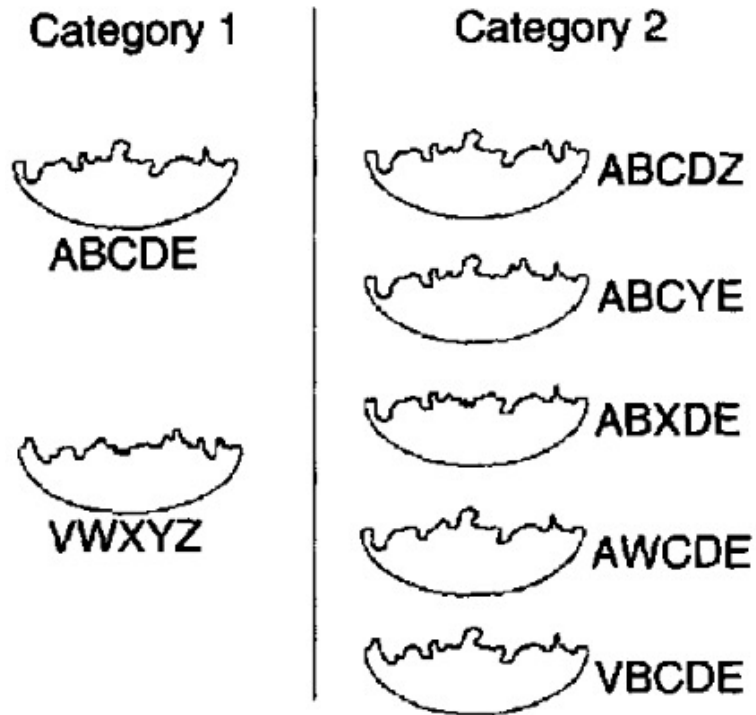
Figure 1.5: Stimuli and category structure for Experiment 1 in Goldstone (2000). Each stimulus was created by combining five line segments with a curved bottom segment. Category 1 contained the stimulus that contained segments A, B, C, D, and E and the stimulus that contained none of these segments. All stimuli in Category 2 contained four of those segments.

Familiarization Trials
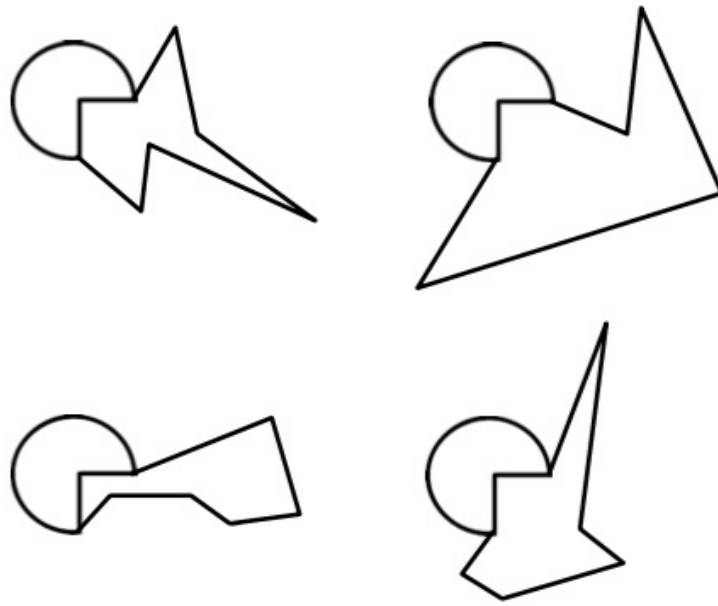
Test Trials

Figure 1.6: Familiarization and test stimuli from Quinn et al. (2006). The familiarization stimuli all share the circle-with-notch component. At test infants decreased their looking time to the circle-with-notch component after viewing the familiarization trials suggesting they learned a feature that corresponds to the circle-with-notch component.

CHAPTER 2

Experiments 1 through 4

In the following experiments we measure the effect of learning category structures on perceptual representations. The role of category learning in shaping changes in perceptual processing has been critiqued by those arguing any change in processing is due to task demands (Firestone & Scholl, 2014, 2015). To avoid task demands, we measure changes in perceptual processing with a whole-part perceptual discrimination task where participants are asked to make fast judgments to determine if a part stimulus matches a whole stimulus (Tanaka & Farah, 1993). The whole-part discrimination task avoids task demands because participants are motivated to be as fast and accurate as possible on all discrimination trials, regardless of the category training. Furthermore, whole-part judgments minimize the role of language as a mediator of perceptual judgments (Witzel & Gegenfurtner, 2013; Hanley & Roberson, 2011) because many part stimuli do not contain enough information to make conclusive category judgments and activate a category label. Whole-part discriminations are also advantageous because they are simple judgments amenable to computational modeling (Nosofsky, 1986).

The category structures and stimuli for these experiments are designed to promote differentiation and unitization perceptual learning processes. The stimuli and *Unitization* category structure are inspired by the eggshell stimuli and category structure from (Goldstone, 2000) (Table 1.5 and Figure 1.5) that show evidence of unitization. In order to use the same

stimuli and have all the stimulus segments that are relevant for the *Unitization* category structure continue to be relevant for the differentiation structure, a new *Differentiation* category structure was developed. The *Differentiation* category structure consists of three separate category structures each containing two categories. Each category structure is defined by the presence or absence of a single segment, and across the three category structures, all of the category relevant segments from the *Unitization* category structure are relevant for a differentiation structure. The stimuli are designed to be difficult to parse into their independently-varying components (Braunstein, Hoffman, & Saidpour, 1989) to ensure that learning is a slow process (Kami & Sagi, 1993). This decreases the likelihood that the features will be learned through feature imprinting (Schyns & Rodet, 1997; Schyns & Murphy, 1994; Austerweil & Griffiths, 2011; Quinn et al., 2006; Quinn & Schyns, 2003).

The first three experiments compare the effect on perceptual discriminations of learning category structures that are consistent with differentiation and unitization learning processes across the same set of stimuli. The fourth experiment addresses the difficulty of separating differentiation and unitization processes by including a category order manipulation (Schyns & Rodet, 1997; Schyns & Murphy, 1994) in a task designed to elicit slower perceptual learning tasks. This experiment tests if the order of differentiation and unitization perceptual learning has a dramatic effect on the set of features people learn.

2.1 Experiment 1: *Unitization* category structure

Experiment 1 measures the effect on perceptual discrimination judgments of learning a category structure that encourages learning perceptual features through unitization. The category structure is inspired by the category structure from Goldstone (2000) with one category consisting of a small number of stimuli that are identified by a conjunction of stimulus properties and the other category containing all other stimuli. This category

structure has been shown to produce changes in reaction times consistent with learning a unitized perceptual feature Goldstone (2000).

Perceptual discrimination performance is measured by sequentially presenting two stimuli and asking participants to indicate if the two stimuli were the same or different. The number of segments present is manipulated by "occluding" some of the stimulus segments of the second stimulus behind a mask. The second stimulus on all perceptual discrimination trials is referred to as the part stimulus because often only a portion of that stimulus is visible, while the first stimulus is referred to as the whole stimulus (Tanaka & Farah, 1993).

A change in perceptual processing due to learning unitized perceptual features could potentially manifest itself in a number of ways, but two critical tests will be conducted. First, does the category membership of the whole stimulus influence perceptual processing? If unitized features are learned, then whole stimuli that contain these unitized features might be expected to be more accurately processed during perceptual discriminations because more features are available for those stimuli. Second, does the category relevance of the unoccluded stimulus segments in the part stimulus influence perceptual processing? If category relevant features are learned better than category irrelevant features, then perceptual discriminations that contain only category relevant stimulus segments are be expected to be more accurate than those that include category irrelevant segments.

2.1.1   Method

Participants

99 Indiana University undergraduates were recruited to participate and were compensated with course credit. 67 participants completed the experiment within the allotted hour and only data from those participants are included in the analyses.
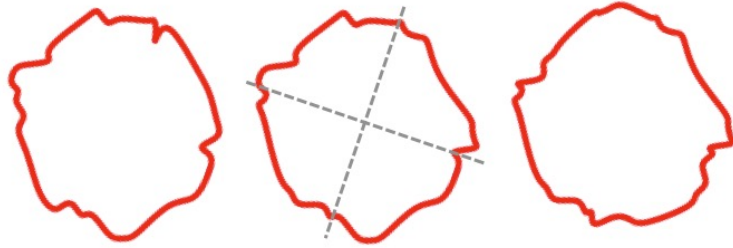
Figure 2.1: Sample stimuli used in all four experiments showing the variation across stimuli. The middle stimulus shares two segments with both of the left and right stimuli. Dashed lines have been added to the middle stimulus to indicate the segment positions and where the segments join together. The angle these segments join was randomized for each participant.

Materials

Stimuli were red contour lines that formed an outline against a white background. Stimuli were formed by placing a randomly selected arc segment in each of four positions so that the segments aligned to form an outline. A set of sample stimuli are shown in Figure 2.1 with an illustration of how segments are aligned into positions. Segments were drawn from a set of 60 segments constructed in Mathematica by fitting a spline through eight randomly perturbed points from a 90-degree arc. The points were randomly selected and they were perturbed no more than 20% of the radius of the arc. Each spline was additionally constrained to pass through two randomly selected control points shared among all segments to ensure segments aligned smoothly to form a connected outline. Stimuli were displayed as 360 by 360 pixel bitmap images against a white background on an LCD monitor in a dark booth. The diameter of each stimulus was approximately 5 centimeters and the viewing distance was approximately 40 centimeters, yielding a viewing angle of approximately 3.5 degrees for each stimulus. The angle at which the segments aligned was fixed for all stimuli for a participant and randomized across participants.

Category structure

	Unitization Category	Differentiation Structures		
Stimulus	Structure	Category 1	Category 2	Category 3
ABCD	Yes	1	1	1
ABCZ	Yes	1	1	1
ABYD	No	1	1	0
ABYZ	No	1	1	0
AXCD	No	1	0	1
AXCZ	No	1	0	1
AXYD	No	1	0	0
AXYZ	No	1	0	0
WBCD	No	0	1	1
WBCZ	No	0	1	1
WBYD	No	0	1	0
WBYZ	No	0	1	0
WXCD	No	0	0	1
WXCZ	No	0	0	1
WXYD	Yes	0	0	0
WXYZ	Yes	0	0	0

Table 2.1: Table of stimuli and which category they belong in. The *Unitization* category structure has Yes and No responses pre-coded since the focus category was always the "Yes" response option. For each of the *Differentiation* category structures "Yes" and "No" responses were randomly assigned to the 1s and 0s. One critical aspect of these category structures is that the final segments, D and Z, are not relevant to any category judgment.

A set of 16 training stimuli were constructed for each participant from eight randomly selected segments which were arbitrarily labeled A, B, C, D, W, X, Y, and Z. Two segments were randomly assigned to each position and only appeared in that position (A and W in the first, B and X in the second, C and Y in the third, and D and Z in the fourth). A

stimulus was produced by selecting one of the two segments for each of the four positions and arranging them to form a contiguous outline. 16 unique stimuli were constructed in this way and each segment appeared in exactly half of the stimuli.

The *Unitization* category structure consisted of two categories: the focus category consisting of the four stimuli that either contained segments A, B and C or contained none of those segments. The non-focus category that contained the remaining 12 stimuli. Table 2.1 shows the *Unitization* category structure across all 16 stimuli and Figure 2.2 shows 16 sample stimuli arranged into the set of focus and non-focus categories.

2.1.2 Procedure

The experiment consisted of two phases that had a different mixture of categorization trials and whole-part perceptual discrimination trials. The first phase of the experiment was composed of seven blocks consisting only of categorization training trials. The 16 stimuli were presented exactly once in each of the seven initial blocks, resulting in 112 categorization training trials in the first phase. The second phase consisted of a mixture of categorization trials and whole-part perceptual discrimination trials. There were 736 total trials in the second phase, 416 whole-part discrimination trials and 320 categorization trials.

Categorization trials

A category training trial consisted of a stimulus from the training set being displayed on the screen for 400 ms, followed by a visual mask for 800 ms, and then a prompt asking was displayed until the participant responded by pressing a key. Feedback was presented for 1500 ms on incorrect trials and 1300 ms on correct trials. The feedback included the stimulus, accuracy, and the number of points scored (or missed) on the trial. After feedback a blank screen was shown for 500 ms before the next trial. Stimuli were presented with equal frequency during category training regardless of the category structure being learned.

Correctly categorizing a stimulus into the focus category while training the *Unitization* category structure was rewarded with six points while correctly categorizing a non-focus category stimulus was worth two points.

The *Unitization* category structure had a unique label randomly sampled for each participant from four nonsense labels: beme, kipe, wune, and vade. If the label assigned to the *Unitization* category structure was "beme" then response prompt was "Was that a beme?" and the correct response for all four members of the focus category was "Yes" and for all others the correct response was "No". Before each of the seven blocks in the categorization-only phase a break screen was displayed to help the participants learn which stimuli belonged in the focus category. All four stimuli from the focus category were displayed on the screen at once and labeled as belonging to the focus category. Participants could study the break screen for as long as they wanted before beginning the next block.

Whole-part perceptual judgments

The first phase of a perceptual discrimination trial was identical to a category training trial and consisted of a stimulus (the whole stimulus) being displayed for 400 ms followed by a visual mask for 800 ms. Instead of being asked to categorize the stimulus, a partially occluded stimulus (the part stimulus) was presented and participants were asked "Does what is visible match what you saw before?" No feedback was provided after the participant responded and a blank screen was presented for 500 ms.

The occlusion mask for the part stimulus was selected from the set of occlusions shown in Figure 2.3. The occlusion masks were rotated and aligned with the segments in the whole and part stimuli so that segments were either completely visible or completely occluded. The mask occluded from zero to three segments of the part stimulus. The part stimulus was selected such that the segments came from the same training set as the whole stimulus

and was constrained such that the visible segments matched across all segments or exactly one segment did not match between the whole and part stimuli.

2.1.3 Results

Category learning

The rate of learning over time and the effect of category membership were assessed via mixed effects linear regression. Participants were modeled with a random intercept but the slope of the line as a function of block was considered a fixed effect. During first seven blocks, which consisted of only the categorization trials, there was a significant effect of block $(F(1, 66) = 41.3, p < 0.001)$, category type $(F(1, 66) = 102.8, p < 0.001)$, and a significant interaction between category type and block $(F(1, 66) = 5.0, p = 0.028)$.

Once whole-part perceptual discrimination trials were interspersed with categorization trials (blocks 8 through 52) the main effect of category type continued to be significant $(F(1, 65) = 25.8, p < 0.001)$ but the effect of block was not significant $(F(1, 63) = 1.6, p = 0.2)$[1] and there was no significant interaction $(F(1, 66) < 1, p = 0.39)$. These results suggest that categorization performance reached a plateau when whole-part judgment trials were introduced and did not improve from then on.

Whole-part perceptual judgments

Overall, whole-part discrimination accuracy was not significantly above chance (M = 0.51, $t(66) < 1, p = 0.45$). However, there was a significant effect of the number of segments in the part stimulus on whole-part discrimination accuracy $(F(1, 66) = 37.3, p < 0.001)$. Figure 2.5 suggests this effect is due to accuracy increasing as the number of unoccluded segments increased.

---

[1]The degrees of freedom varies in this analysis because due to random chance there were not focus and non-focus categorization trials in every block for some participants.

When the set of part stimuli was restricted to only those in which the category irrelevant segment could be present or occluded (part stimuli with one to three segments unoccluded), the main effect of the number of unoccluded segments in the part stimulus remained significant ($F(1, 66) = 5.3, p = 0.024$) but there was no significant effect of the irrelevant segment being present or occluded ($F(1, 66) = 2.1, p = 0.15$) or of the category membership of the whole stimulus ($F(1, 66) < 1, p = 0.3$). No interactions between these factors were significant.

2.1.4 Discussion

During the initial training phase participants improved their categorization accuracy with more experience. This categorization accuracy did not continue to improve once whole-part discrimination trials were introduced but leveled off and remained constant (Figure 2.4). Throughout both phases, the focus category stimuli were more accurately categorized than the non-focus stimuli. This asymmetry is indicative of a response bias toward selecting the focus category response and suggests the training phase was not sufficient to learn a set of stable perceptual features that could be used for categorization.

Perhaps because of the limitations of the category training phase, there is little evidence for a shift in feature representations in whole-part discrimination performance. Neither of the two predicted effects of perceptual learning were found: there was no effect of the category membership of the whole item, and there was no effect of the category relevance of the segments in the part stimulus. One straightforward interpretation of both the systematic categorization response bias and lack of whole-part discrimination effects is that the categorization task was not learned well enough for participants to extract stable *unitized* perceptual features. Introducing the whole-part discrimination tasks so quickly after beginning category learning may have prevented learning useful features.

2.2 Experiment 2: *Differentiation* category structure

Experiment 2 measures the effect on perceptual discrimination judgments of learning category structures that encourages learning features that correspond to individual segments instead of one large feature. The same stimulus segments that were relevant for categorization in Experiment 1 were relevant for the *Differentiation* category structures. In Experiment 1, segments were predictive of category membership only in conjunction with each other, in Experiment 2 the same segments were independently predictive of category membership. Due to the difficulty in Experiment 1 of learning the category structure after categorization and whole-part discrimination trials were interspersed, the training regime was modified for Experiment 2 to ensure participants reach an accuracy criterion across all categories before progressing to whole-part discrimination trials.

2.2.1 Method

Participants

71 Indiana University undergraduates were recruited to participate and were compensated with course credit.

Materials

The same stimuli from Experiments 1 were used in this experiment.

Category structures

The set of *Differentiation* category structures consisted of three structures each with two categories. The two categories within each structure were defined by the presence of a single segment. For example, one *Differentiation* category structure would assign all eight stimuli containing the A segment to one category and the remaining eight that contain

the W segment to another category. Each of the three *Differentiation* category structures corresponded to one of the segments relevant to the *Unitization* category structure from Experiment 1. The D and Z segments were not relevant for either the *Unitization* or *Differentiation* category structures.

All *Differentiation* category structures mirrored the *Untitization* category structure in that they were constructed and taught as a "Yes/No" response to a single category label. Unlike the *Unitization* category structure, all *Differentiation* category structures had eight stimuli that belonged to the "Yes" and "No" groups and the mapping of group to responses was randomized across participants.

Procedure

The categorization trials of the *Differentiation* category structures were slightly different than the the *Unitization* category structure from Experiment 1. Instead of having only a single category that could be asked about on a categorization trial, there were three category structures that could be asked about for every stimulus. The labels for each of those category structures was selected from the set of random category labels used in Experiment 1. Categorization trials began with a stimulus being presented for 400 ms followed by a visual mask for 800 ms, and then one of the three possible category structures was queried on each trial. The response prompt was identical to that in Experiment 1, except that participants were exposed to three of the four labels and their associated category structures over the course of the experiment.

Experiment 2 consisted of two phases which were composed of the same categorization and whole-part discrimination trials as Experiment 1. The first phase, though still composed only of categorization trials as in Experiment 1, was not a fixed number of trials but continued through three distinct stages. In the first stage, one of the three *Differentiation*

47

category structures was randomly selected and categorization trials asked only about that one category structure until an accuracy threshold of 80% correct was reached within a block. The second stage of training included trials that asked about two of the three category structures, and all three category structures were included in the final stage of category training. If, for example, the randomly chosen labels were beme, kipe, and wune; one possible training ordering would be if the first stage consisted only of kipe trials, the second stage consisted of trials that asked about the kipe or wune categories, and the final stage consisted of all three category structures. The second phase of Experiment 2 was the same as in experiment 1, consisting of 736 total trials, 416 whole-part discrimination trials and 320 categorization trials.

A break screen was shown to help the participants learn each category structure before every block in the first phase of training. For each of the category structures being shown in the current stage (one in the first, two in the second, and all three in the third stage) a help screen displayed two "Yes" stimuli and two "No" stimuli. If more than one help screen was due to be shown then the order of the screens was randomized. Participants could study each screen for as long as they wanted before moving to the next one or beginning the next block.

2.2.2 Results

Category learning

Participants struggled to learn all three *Differentiation* category structures within the hour allotted for the experiment. Of the 71 total participants, 68 (95%) reached the second stage of training, 40 (56%) reached the third stage of training, and 33 (46%) completed the training phase and proceeded to the second phase. The analysis of the whole-part perceptual discrimination performance includes all participants who reached that stage though most

did not complete the full set of trials in that phase.

The participants who did learn each of the category structures did so efficiently. For all subsequent analyses the number training blocks necessary to reach the accuracy criterion is conditional on reaching the criterion. On average, the participants who completed the first stage did so in 3.4 blocks, the second stage was completed in 6.0 blocks, and the third stage was completed in 6.1 blocks. The participants who reached the second phase learned all three category structures in 14.25 blocks, ranging from 8 to 38 blocks of training.

Once participants reached the second phase, in which whole-part perceptual discrimination judgments were interleaved with categorization trials, categorization performance remained highly accurate. On average, participants got 91.5% of categorization trials correct during the interleaved phase. This was much higher than Experiment 1, suggesting that training until criterion was successful in ensuring participants learned the category structures.

Whole-part perceptual judgments

Whole-part discrimination accuracy was significantly above chance (M $= 0.54$, $t(32) = 5.1, p < 0.001$). There was a significant effect of the the number of unoccluded segments in the part stimulus on accuracy ($F(1, 32) = 44.0, p < 0.001$). Figure 2.6 shows that accuracy increased as the number of unoccluded segments increased.

When the set of part stimuli was restricted to only those in which the category irrelevant segment could be present or occluded (part stimuli with one to three segments unoccluded), the main effect of number of unoccluded segments on accuracy remained significant ($F(1, 32) = 15.5, p < 0.001$) but there was no significant effect of the irrelevant segment being present or occluded ($F(1, 32) = 1.3, p = 0.26$). The interaction between category relevance and number of segments was not significant ($F(1, 32) < 1, p = 0.5$).

2.2.3 Discussion

Participants struggled to learn all three *Differentiation* category structures. Slightly less than half of the participants successfully completed the categorization training within an hour. One surprising result was the lack of effect of category relevance on whole-part discrimination. There was no significant difference between judgments that do or do not include the category irrelevant segment. This suggests participants did not shift their attention to only process segments that were relevant for categorization.

2.3 Experiment 3: Whole-part perceptual discrimination

Experiment 3 measures the pattern of whole-part perceptual discrimination judgment accuracy without any categorization training. This is an important control for Experiments 1 and 2, to compare the whole-part discrimination performance after learning the *Unitization* and *Differentiation* category structures to discrimination performance without any category learning.

2.3.1 Method

Participants

27 Indiana University undergraduates were recruited to participate and were compensated with course credit. All participants completed the experiment within the allotted hour.

Materials

The same stimuli from Experiments 1 and 2 were used in this experiment.

Procedure

The experiment consisted of 416 whole-part perceptual discrimination trials. As in Experiments 1 and 2, this set was constructed by pairing each whole stimulus with each occluder mask twice. The whole and part stimuli matched on all unoccluded segments on approximately half the trials and the order of trials was randomized across participants.

2.3.2 Results

Whole-part discrimination accuracy was significantly above chance (M $= 0.58$, $t(26) = 9.7$, $p < 0.001$). There was a significant effect of the number of unoccluded segments in the part stimulus on discrimination accuracy ($F(1, 26) = 191$, $p < 0.001$). Figure 2.7 shows that accuracy increased as the number of unoccluded segments increased. A discussion of these results in the context of Experiments 1 and 2 is included in the next section.

2.4 Analysis of whole-part discrimination performance across Experiments 1, 2, and 3

Though these experiments were conducted with separate samples of participants, the structure of the whole-part trials were identical across all studies and comparing across them may provide insight into the differences due to category structure.

A repeated-measures ANOVA predicting whole-part discrimination accuracy found a significant main effect of category structure ($F(2, 124) = 20.7$, $p < 0.001$), a significant main effect of the number of unoccluded segments in the part stimulus ($F(1, 124) = 178.2$, $p < 0.001$), and a significant interaction between category structure and number of segments ($F(2, 124) = 6.5$, $p = 0.002$). The accuracy in the *No category* experiment (M $= 0.58$) was significantly higher than the *Differentiation* category structure (M $= 0.44$, Welch's unequal variance t-test $t(57.8) = 3.0$, $p = 0.004$), which was significantly higher than the *Unitization*

category structure (M $= 0.51$, Welch's unequal variance t-test $t(73.9) = 3.4, p = 0.001$).

The slope of the best fitting line relating the number of unoccluded segments to accuracy showed an increase of 5.2% in accuracy for each additional unoccluded segment. The interaction between category structure and the number of unoccluded segments was due to differences in the relationship between the number of unoccluded segments and accuracy between category structures. This can be modeled as the slope of the best fitting line for each category structure. The slope for the *No category* experiment ($\beta = 0.073$) was higher than the *Differentiation* category structure ($\beta = 0.060$), which was higher than the *Unitization* category structure ($\beta = 0.040$).

These results suggest that learning the category structures, or more likely, having categorization trials interspersed with whole-part discrimination trials, decreased accuracy on the discrimination trials. The interspersed categorization trials had more of an effect for the *Unitization* category structure than the differentiation structures. This is interesting because participants were more accurate on the interspersed differentiation categorization trials than the categorization trials. This might be due to differences in training between the two conditions because participants in the differentiation training condition were trained to criterion and only highly accurate participants advanced. This might have created a biased sample toward more attentive participants or participants who had learned the categorization well enough to devote less attention to it.

Did category relevance have a differential effect across the three experiments? To compare the effect of category relevance in Experiments 1 and 2 with Experiment 3, which had no category structure, a position for each participant in Experiment 3 was arbitrarily selected to be category irrelevant. A repeated-measures ANOVA predicting whole-part discrimination accuracy (on part stimuli with one to three segments unoccluded) found a significant main effects of category structure ($F(2, 124) = 17.75, p < 0.001$), number of unoccluded segments

in the part stimulus ($F(1, 124) = 40.0, p < 0.001$), and a significant interaction between category structure and number of segments ($F(2, 124) = 4.3, p = 0.015$). There was not a significant effect of the category irrelevant segment being occluded ($F(1, 124) < 1, p = 0.9$) and category relevance did not interact with any other factors (all F values ¡ 1).

2.5 Discussion

Experiments 1 through 3 do not show strong evidence of a change in perceptual discriminations due to learning either the *Unitization* or *Differentation* category structures. This is made clear by the lack of effect of category membership in Experiment 1 and category relevance in Experiments 1 and 2.

Furthermore, a lack of effect of category relevance in the *Differentiation* category structure condition suggests people were not able to learn to segment these stimuli into the individually varying components within the experiment time. If people learned to represent these stimuli as a combination of the four components, an advantage for processing category relevant components would be predicted by attention shifting models of categorization (Nosofsky, 1986; Kruschke, 1992; Love et al., 2004). These models would shift attention to the components of the stimuli that are relevant for categorization, leading to improved processing. It is very likely, therefore, that the amount of exposure to these stimuli within an hour long session was insufficient for participants to learn any stable feature representation, regardless of category structure. Instead, participants were likely processing the stimuli as whole stimuli and struggling to make accurate judgments when more segments are occluded. This is consistent with the result found in all three experiments that discrimination accuracy increased as the number of unoccluded segments in the part stimulus increased. When the part stimulus contained more unoccluded segments, participants were more accurate in their judgment despite the number of possible mismatching segments being fixed.

In these one-hour experiments participants were not able to learn new features that significantly influenced their perceptual discrimination performance. The pace of feature learning in these experiments is much more consistent with the perceptual learning evidence from visual search (Shiffrin & Schneider, 1977) or psychophysics (Fine & Jacobs, 2002) where learning requires more than one session of training, than studies that find perceptual learning within a session of category training (Goldstone, 2000; Goldstone & Steyvers, 2001; Goldstone, 1994; Pevtzow & Goldstone, 1994). In Experiment 4 this concern is addressed by increasing the amount of category training to more than one session.

2.6   Experiment 4: Multi-session experiment

Experiments 1 and 2 did not find reliable changes in perceptual discrimination performance due to category learning. A major issue with interpreting the results of these studies was the restriction on the amount of category training participants received. Participants in Experiment 1 received a fixed amount of training on the *Unitization* category structure and by the end of the experiment did not approach near-perfect accuracy (Figure 2.4), suggesting their representations of the categories were incomplete. This issue was addressed in Experiment 2 by adding an accuracy threshold to category training, but less than half of the participants (46%) were able to learn the *Differentiation* category structures which may have produced a selection effect. Furthermore, the previous experiments did not account for individual differences in perceptual discriminations by measuring performance before and after learning category structures.

These issues are addressed in Experiment 4 by spreading the category training and whole-part discriminations across multiple sessions. This allows perceptual discrimination performance to be assessed both before and after learning a category structure to assess relative changes due to category learning within a single individual. Furthermore, it allows

whole-part discrimination performance to be assessed without interspersing categorization trials that may differentially interfere with processing. It also allows an individual to learn multiple category structures in sequence. We will compare the perceptual discrimination performance of participants who start with learning the *Unitization* category structure and then learn the *Differentiation* category structures to participants who learn categories in the opposite order.

The perceptual learning literature makes two contrasting predictions about how order should impact the set of perceptual features. Perceptual learning effects due to the fast-acting Imprinting perceptual learning process (Schyns & Murphy, 1994; Schyns & Rodet, 1997; Quinn et al., 2006) predict that different orders of category training would produce different sets of perceptual features if imprinting, differentiation, and unitization processes share a common learning mechanism. Training on the *Unitization* category structure first should result in learning a unitized perceptual feature that will remain in set of features even after differentiation category training. Training on the *Differentiation* category structure first should result in learning a set of differentiated perceptual features and subsequent unitization category training will not add a unitized perceptual feature. However, perceptual learning effects due to the perceptual unitization process with structural relations across clearly separable components (Pevtzow & Goldstone, 1994; Austerweil & Griffiths, 2011), over many sessions of training (Shiffrin & Lightfoot, 1997), or the development of unitized perceptual features through the acquisition of expertise (Busey & Vanderkolk, 2005; Gauthier & Tarr, 1997; Gauthier et al., 1998; Gauthier & Tarr, 2002) predict that given enough training, a unitized perceptual feature can be added to the set of perceptual features even after the components are features.

2.6.1 Method

Participants

12 male Indiana University undergraduates aged 18 to 20 completed seven one-hour sessions and were compensated $8 per session. All participants reported having normal or corrected-to-normal vision and no color blindness.

Materials

Four sets of stimuli were generated for each participant following the same procedure outlined in Experiments 1 through 3. One set of stimuli was randomly selected to be the set used during both *Unitization* and *Differentiation* category training. The other three sets were never presented during training but used to assess whole-part perceptual discrimination of novel stimuli at various stages of training.

Category structures

The *Unitization* category structure was identical to the the structure in Experiment 1 and the *Differentiation* category structures were identical to those in Experiment 2. Table 2.1 shows all four category structures across a set of 16 stimuli. The four category labels from Experiments 1 and 2: beme, kipe, wune, and vade were randomly assigned to the *Unitization* category structure and the three *Differentiation* category structures. As in the previous experiments, the focus category within the *Unitization* category structure, which consisted of only four stimuli, was always assigned to the "Yes" response of the category query. For the *Differentiation* category structures the "Yes" and "No" responses were randomly assigned within each structure.

2.6.2 Procedure

Category training trials

The categorization trials were identical to those in Experiments 1 and 2.

Perceptual discrimination trials

Whole-part perceptual discrimination trials were identical to those in Experiments 1 through 3 except that instead of the segments of the part stimulus matching all segments from the whole stimulus or mismatching exactly one segment, a third condition was added in which none of the segments unoccluded in the part stimulus matched those in the whole stimulus.

Structure across sessions

The experiment consisted of seven sessions. Sessions 1, 4, and 7 contained only whole-part perceptual discrimination trials, sessions 2 and 3 consisted of learning the first category structure, and sessions 5 and 6 consisted of learning the second category structure. Half the participants were randomly assigned to learn the *Unitization* category structure first and the other half learned the *Differentiation* category structures first. Training for each category structure was split across two one-hour sessions. A session consisting of whole-part perceptual discrimination trials without any categorization trials was included before any category learning, and after learning the *Unitization* category structure, as well as after learning the *Differentiation* category structures.

Each whole-part perceptual discrimination session consisted of 832 trials. The 16 training stimuli were presented as the whole stimulus with each of the 13 occlusion patterns shown in Figure 2.3 three times. This produced 624 whole-part discrimination trials with stimuli used in training. The remaining 208 trials presented whole and part stimuli selected from a set of 16 stimuli not used during training. One set of 16 novel stimuli was created

for each whole-part perceptual discrimination session. These stimuli were generated follow-ing the same procedure as the training stimuli and presented with segments at the same rotation angle as the familiar trained stimuli. The 16 novel stimuli were presented once with each of the 13 occlusion patterns. The order of all perceptual discrimination trials was randomized within each session.

The four training sessions devoted to category training followed the same overall pat-tern. All sessions lasted 58 minutes regardless of categorization performance. However, the composition of trials varied depending on categorization accuracy. The category training began exclusively with categorization trials for both the *Unitization* and *Differentiation* category structures. Only categorization trials were presented for both the *Unitization* and *Differentiation* category structures until a high accuracy was achieved. For both category structures, after the threshold was reached, the second phase of category training began in which the trials were a mixture of categorization trials and whole-part perceptual discrimi-nation trials. This phase continued until both sessions of category training were completed. 75% of trials during this mixture phase were perceptual discrimination trials and only 25% were categorization trials. Participants continued to receive feedback when they made in-correct category responses but received no feedback on perceptual discrimination trials.

The procedure of category learning for both the *Differentiation* and *Unitization* category structures closely resembled the procedure in Experiment 2. The training of the *Differen-tiation* category structures began with presenting only one of the *Differentiation* category structures. The second and third structures were added only when participants reached a high accuracy within a block (80%). In order to mirror this procedure when training the *Unitization* category structure, participants were required to reach the same accuracy crite-rion in three blocks before advancing. Break screens, as described in Experiments 1 and 2, were presented every 48 trials during the categorization-only phase of learning. These brake

screens showed examples of each category structure being learned. Once the second phase of category training began and whole-part discrimination trials were presented, the break screens no longer contained any category structure information and instructed participants to rest for a minute.

2.6.3 Category training results

11 of the 12 participants reached the categorization accuracy threshold during the first session of the first category structure, and the one remaining participant reached the threshold early in their second session. All participants reached the accuracy threshold in their first session of their second category structure. A repeated-measures ANOVA predicting the number of blocks required to reach the accuracy thresholds shows a main effect of the within-subject factor of training order (first training phase, M = 20.2; second training phase, M = 9.1; $F(1, 10) = 8.8, p = 0.014$) but no significant effect of the type of category structure ($F(1, 10) = 4.4, p = 0.061$) or significant interaction between category order and category structure ($F(1, 20) = 3.4, p = 0.10$).

Categorization accuracy did not seem to drop off after interspersing whole-part discrimination and the categorization trials. These categorization judgments were 92% correct across all training conditions. An ANOVA predicting accuracy on these interspersed categorization trials found a significant effect of which category was learned first (*Unitization* M = 0.95; *Differentiation* M = 0.89; $F(1, 10) = 10.6, p = 0.009$) but no significant effect of how many category structures had been learned ($F(1, 10) = 4.3, p = 0.064$) nor an interaction between the two factors ($F(1, 10) = 1.9, p = 0.20$). These effects are shown in Figure 2.10.

2.6.4 Whole-part perceptual discrimination results

The data presented here is restricted to the whole-part discrimination trials from sessions 1, 4, and 7 that consist only of discrimination trials.

Stimulus-level effects

First we measure the effect of three factors that are specific to the stimulus and vary across all conditions: the novelty of the stimuli, the number of unoccluded segments in the part stimulus, and the proportion of segments that match between the whole and part stimuli. Differences due to these factors might be entirely due to the properties of the stimuli themselves (except for novelty), but an interaction of these factors with the novelty suggests that category learning is changing the influence of that factor on whole-part discrimination judgments. A three-factor repeated-measures ANOVA was performed with three stimulus-level effects: familiarity, the number of unoccluded segments in the part stimulus, and the match proportion.

There was a main effect of familiarity ($F(1, 11) = 36.1, p < 0.001$). Whole-part trials that consisted of stimuli from training were more accurately processed (M = 0.78) than novel stimuli (M = 0.72).

The number of unoccluded segments in the part stimulus also had a significant effect on whole-part performance ($F(1, 11) = 73, p < 0.001$). Part stimuli with one unoccluded segment were the least accurate (M = 0.71) and accuracy increased as the number of unoccluded segments increased (two segments, M = 0.77; three segments, M = 0.80; four segments, M = 0.83). Post-hoc comparisons, using a Bonferroni corrected alpha value of 0.017 for three comparisons, show all adjacent comparisons were significant different: comparisons with one unoccluded segment were less accurate than two segments ($t(11) = 10.6, p < 0.001$), two segment judgments were less accurate than three ($t(11) = 4.1, p = 0.001$), and three segment judgments were less accurate than those with four segments unoccluded ($t(11) = 3.7, p = 0.003$). This pattern is the reverse of the effect found in the first three experiments but this is likely due to the addition of trials where no segments in the part stimulus match those in the whole stimulus (see Figure 2.11 and the interaction

analysis below).

The proportion of segments that matched between the whole and part stimuli, coded as a three-level categorical predictor, had a significant effect on performance ($F(2,11) = 44.7, p < 0.001$). stimulus pairs in which no segments matched were the most accurate (M = 0.83), followed by pairs with all segments matching (M = 0.77), and the least accurate pairs were those that matched on all but one segment (M = 0.61).[2] Post-hoc comparisons, using a Bonferroni corrected alpha value of 0.017 for three comparisons, show that the one mismatch trials were significantly less accurate than the all matching trials ($t(11) = 6.0, p < 0.001$) or the none matching trials ($t(11) = 14.6, p < 0.001$) and the all matching trials were not significantly different than the none matching trials ($t(11) = 2.1, p = 0.05$).

All three two-way interactions in the repeated-measures ANOVA were significant. The number of segments and the proportion of matching segments showed a significant interaction ($F(2,22) = 65.1, p < 0.001$), the proportion of matching segments and familiarity showed a significant interaction ($F(2,22) = 7.5, p = 0.003$), and the number of segments and familiarity showed a marginally significant interaction ($F(1,11) = 4.9, p = 0.049$). The three-way interaction was not significant ($F(2,22) = 2.3, p = 0.1$).

Figure 2.11 shows the effects and interactions between the stimulus-level factors. Overall, familiar trials are more accurate than novel, more unoccluded segments increases accuracy, and the one mismatch trials are less accurate than the all matching or none matching trials. Furthermore, the effect of having more unoccluded segments in the part stimulus increases accuracy for trials with unoccluded segments that all match or none match, but decreases accuracy for trials with one mismatch. This is reasonable since additional segments in the first two trial types agree with the correct response, but additional segments mislead in the one mismatch condition. Also, the advantage of processing trained stimuli

---

[2]Part segments composed of only one segment that mismatched the whole stimulus were coded as none matching rather than one mismatching but changing this does not change the overall pattern.

over novel stimuli seems to be strongest for stimuli with one mismatching segment (third panel of Figure 2.11).

Session-level effects

Next, we measure the effect of two factors that vary across category training sessions: the number of category structures learned thus far, and which category structure the participant learned first. The two effects are first measured as separate analyses then combined into a five-factor repeated-measures ANOVA with the stimulus-level factors to assess interactions.

Participants significantly improved as they got additional training (repeated-measures ANOVA $F(2, 11) = 10.5, p < 0.001$). Accuracy on whole-part judgments was lowest before any training (M = 0.69), higher after the first training regardless of the category structure being taught (M = 0.78), and highest after both category structure training sessions (M = 0.84). Post-hoc comparisons, using a Bonferroni corrected alpha value of 0.025 for two comparisons, show that accuracy before any training was significantly lower than after the first category training ($t(11) = 10.1, p < 0.001$) and the accuracy after the first training was significantly lower than after the second category training ($t(11) = 6.8, p < 0.001$).

Furthermore, the participants who received the *Unitization* category structure training first were significantly more accurate (M = 0.79) than those who trained on the *Differentiation* category structure first (M = 0.74, $F(1, 10) = 6.2, p = 0.03$, Figure 2.12). Interestingly, there was not a significant interaction between the number of category structures learned and the type of category structure first learned ($F(2, 20) = 1.1, p = 0.3$). This suggests that the advantage in perceptual discrimination for the participants who learned the *Unitization* category structure first was not due to the training itself but something about this random sample of people. This interpretation is strengthened by the significant difference between the two groups in whole-part perceptual discrimination accuracy before any category train-

ing (*Unitization* first, M = 0.70; *Differentiation* first, M = 0.66; Welch two-sample t-test $t(7.6) = 2.8, p = 0.024$).

A five-factor repeated-measures ANOVA found no significant interactions between any combination of the session-level factors and the stimulus-level factors. The interaction between session-level and stimulus-level factors with the largest F value was for the non-significant interaction between familiarity and the category structure first learned ($F(1, 10) = 4.5, p = 0.06$).

The main effect of initial training condition is troubling since participants were randomly assigned to their training order. This effect, combined with the significant difference in performance due to "initial training" before any training, strongly suggests that learning the *Unitization* category structure first did not improve processing for those individuals, they were better at perceptual discrimination before any training. Furthermore, the lack of significant interaction between initial training and the number of category structures learned suggests that category training did not increase or decrease the advantage.

Category-specific effects

Finally, we measure the effect of three factors that are specific to the category being learned. It is hypothesized the *Unitization* category structure training should lead to improved processing of the four stimuli that belong in the focus category (Figure 2.2) by learning the two perceptual features relevant for the *Unitization* category structure (the features composed of segments ABC and WXY in Table 2.1). To test for an advantage in perceptual discriminations for whole stimuli that belong to the focus category, we compare accuracy on whole-part discriminations of familiar stimuli after *Unitization* or *Differentiation* category structure training. A repeated-measures ANOVA predicting accuracy on whole-part discriminations with category structure, whole stimulus category membership, and the num-

ber of unoccluded segments in the part revealed a main effect of the number of unoccluded segments ($F(1, 11) = 15.9, p = 0.002$), no significant main effect of category structure ($F(1, 11) < 1$), and no significant main effect of the focus category membership of the whole stimulus ($F(1, 11) = 1.6, p = 0.23$). However, there was a significant interaction between category structure and focus category membership ($F(1, 11) = 5.6, p = 0.038$) and the remaining interactions were not significant (category structure and unoccluded segments $F(1, 11) < 1$; category membership and unoccluded segments $F(1, 11) = 3.9, p = 0.07$; three way interaction $F(1, 11) = 2.7, p = 0.13$).

Figure 2.13 suggests the interaction between focus category membership and category structure is due to the focus category having an effect only after *Unitization* category training. Post-hoc comparisons show significantly higher accuracy on trials with whole stimuli that belong in the focus category after *Unitization* category training (focus M = 0.85; non-focus M = 0.82, paired $t(11) = 2.8, p = 0.017$) but no effect of focus category after *Differentiation* category training (focus M = 0.84; non-focus M = 0.83, paired $t(11) < 1$). This pattern suggests perceptual discriminations were more accurate for the stimuli in the focus category after learning the *Unitization* category structure than after the *Differentiation* category structures. This improvement does not seem to depend on the number of unoccluded segments in the part stimulus.

The second category-specific effect we consider is the category relevance of the segments in the part stimulus. When restricting part stimuli to those that are familiar and have at least one segment occluded, exactly half of the part stimuli contain a segment that is not relevant to either the *Unitization* or *Differentiation* category structures. Across all conditions, there is not a significant difference in accuracy on whole-part discrimination trials containing the irrelevant segment (M = 0.79) and those occluding that segment (M = 0.77, paired $t(11) = 1.5, p = 0.16$). Yet category relevance should only matter after any

category training. In the two sessions of whole-part discriminations following any category training, accuracy was significantly higher on trials occluding the irrelevant segment (M = 0.85) than those containing that segment (M = 0.81, paired $t(11) = 3.7, p = 0.003$). To understand the relationship between the presence of category irrelevant segment and other factors, a repeated-measures ANOVA was conducted to compare the effects of the presence of the category irrelevant segment, the category structure, and number of unoccluded segments in the part on discrimination accuracy. There were significant effects of the presence of the category irrelevant segment ($F(1, 11) = 13.0, p = 0.004$), unoccluded segments ($F(1, 11) = 87.4, p < 0.001$), and category structure ($F(1, 11) = 35.0, p < 0.001$). The only significant interaction was between category structure and the presence of the category irrelevant segment ($F(1, 11) = 5.1, p = 0.015$). Figure 2.14 suggests this interaction may be due to the lack of effect of category relevance before any category training. Post-hoc tests of this interaction (using a Bonferroni corrected p-value of 0.017) show that the effect of occluding the category irrelevant segment is not significant after no category training ($t(11 = 1.1, p = 0.28$)) or after *Differentiation* category training (present M = 0.81; occluded M = 0.85; $t(11) = 2.4, p = 0.035$), but is significantly different after *Unitization* category training (present M = 0.85; occluded M = 0.80; $t(11) = 3.4, p = 0.006$). This pattern is shown in Figure 2.15. [3]

Finally, we consider the effect of category structure on judgments with different number of unoccluded segments in the part stimulus. Two specific predictions of the differentiation and unitization learning processes are tested. First, the differentiation perceptual learning process is hypothesized to improve processing of individual segments when learning the *Differentiation* category structure because this structure will promote learning category relevant features that map onto the individual segments. Yet there is no difference in ac-

---

[3]This effect is mostly due to a smaller variance in the judgments after unitization rather than a larger difference between the means.

curacy on perceptual judgments of familiar stimuli with one unoccluded category relevant segment after training on the *Differentiation* category structure (M = 0.80) than the *Unitization* category structure (M = 0.78; paired t-test t(11) ¡ 1). Similarly, the *Unitization* category structure is hypothesized to improve processing of conjunctions of three category relevant segments since the category relevant features align with the segments. Yet there is no difference in accuracy on perceptual judgments of familiar stimuli with three unoccluded category relevant segments after training on the *Unitization* category structure (M = 0.86) than the *Differentiation* category structure (M = 0.86; paired t-test t(11) ¡ 1). Figure 2.16 shows the effect of category structure across the number of segments in the part stimulus and suggests there was no differential effect of category structure in any level of the number of segments. These null effects are particularly surprising from the perspective of perceptual learning because they suggest category training was not leading to improvements in making discrimination judgments when the part stimulus corresponded to the diagnostic feature of the category.

2.6.5   Discussion

The results from Experiment 4 show a much stronger effect of learning the *Unitization* and *Differentiation* category structures than Experiments 1 and 2. All participants were able to learn both category structures and then maintain categorization accuracy above 90% when whole-part discrimination trials were interspersed. Furthermore, this learning had a dramatic influence on whole-part perceptual discrimination judgments. After any category training, participants were more accurate on judgments of familiar stimuli and stimuli with the category irrelevant segment occluded. Yet these changes do not point to a differential effect due to learning the *Unitization* or *Differentiation* category structures.

There was a large effect on discrimination accuracy of which category structure was

learned first, but this effect was found even before any training and did not change after learning (Figure 2.12). This result highlights a potential issue with between-subject comparisons with few participants, an unlucky sample can bias results. However, including the pre-training whole-part discrimination session identified this unlucky sample and highlights the importance of comparing pre- and post-training within an individual.

Experiment 4 does contain evidence of differential processing due to learning the *Unitization* or *Differentiation* category structures. First, learning the *Unitization* category structure led to more accurate on judgements when the whole stimulus was a member of the focus category and not in the non-focus category. There is no difference in accuracy for the same stimuli after learning the *Differentiation* category structure (Figure 2.13). Second, the difference between perceptual discriminations containing the category irrelevant segment and those without it is significant after learning the *Unitization* category structure but not after learning the *Differentiation* structure (Figure 2.14).

These effects are consistent with some difference in learning between the two conditions, but are not unique to perceptual learning. Both of these effects are consistent with attention shifting accounts of category learning. First, the improvement for judgments of category relevant part stimuli after *Unitization* training might be due to an additional shift of attention to category relevant segments (Kruschke, 1992; Nosofsky, 1986). Similarly, the improvement for judgments of stimuli from the focus category after *Unitization* training might be due to shifting attention to specific exemplars in memory (Nosofsky, 1988b). Yet the category-specific effects on perceptual discrimination are not inconsistent with perceptual learning. These differences could be accounted for by *Unitization* category training shifting attention to the category relevant segments more than the *Differentiation* category structure. More challenging for the perceptual learning account is the lack of advantage in processing part stimuli that correspond to the diagnostic features of categories (Figure

2.16). These were strong predictions of the perceptual learning account that would be difficult to account for with attention shifting alone. Yet the *Differentiation* category structure did not improve judgments of single segment part stimuli and the *Unitization* category structure did not improve judgments of three segment part stimuli. Overall, these results do not demonstrate strong evidence for differentiation and unitization perceptual processes in the same learning environment.
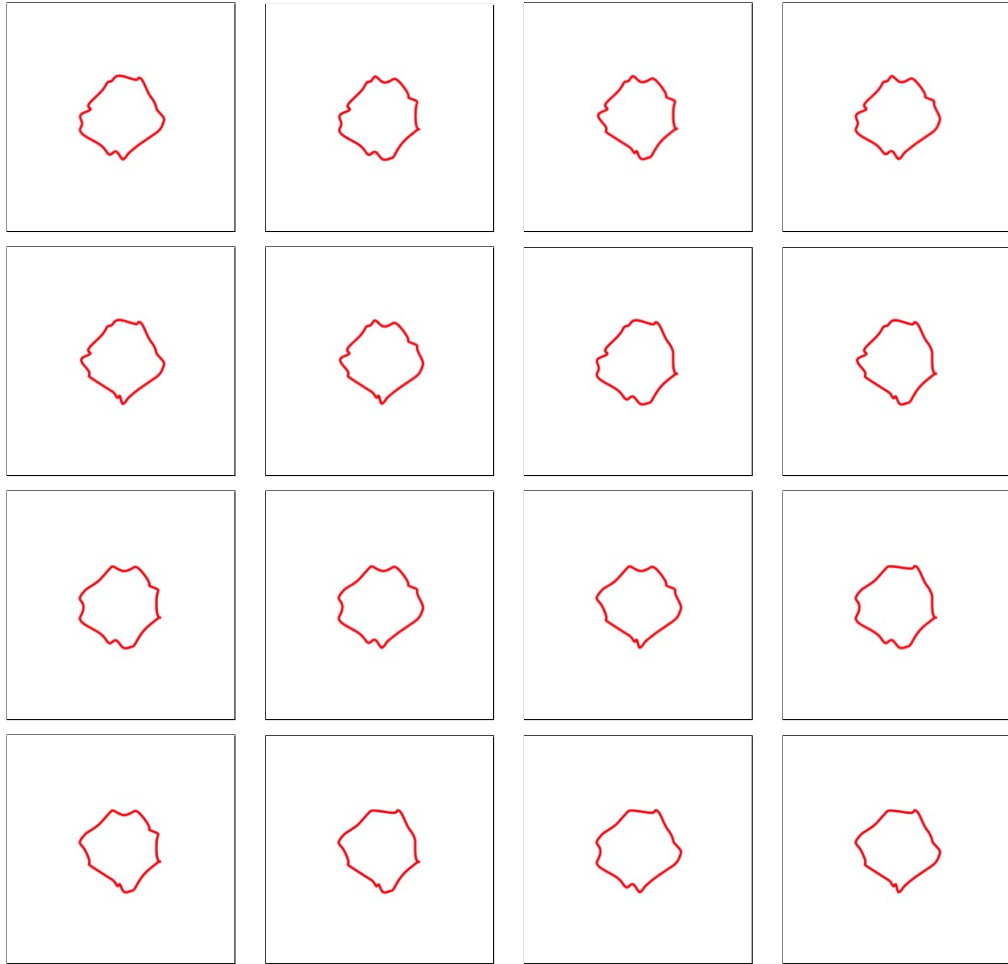
Figure 2.2: 16 sample stimuli showing the *Unitization* category structure. In this example the stimuli in the left column would be in the focus category and the other three columns would be non-focus category stimuli. The focus category members are composed of segments ABCD, ABCZ, WXYD, and WXYZ respectively.
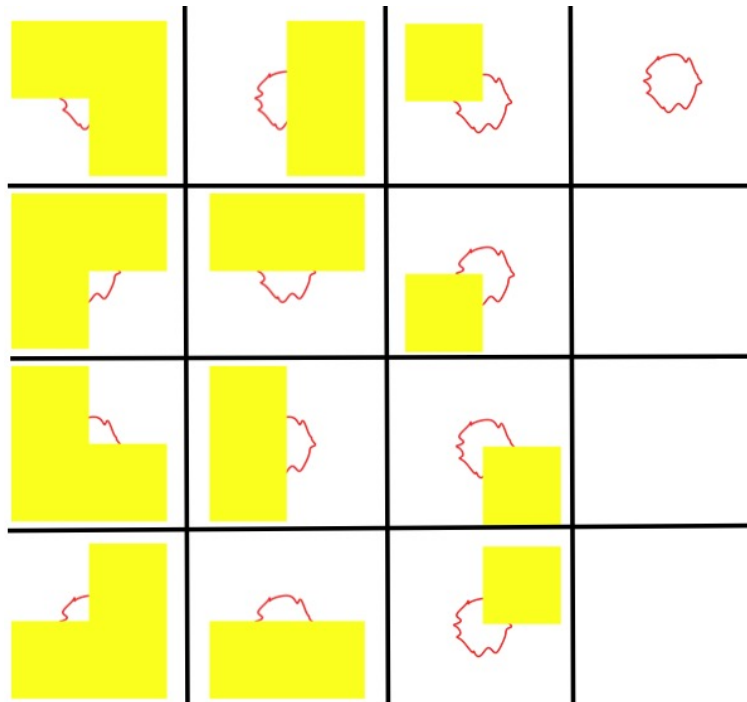
Figure 2.3: The 13 occlusion masks used to obscure segments to create part stimuli for all experiments. Part stimuli with one, two, and three visible segments were equally probable.
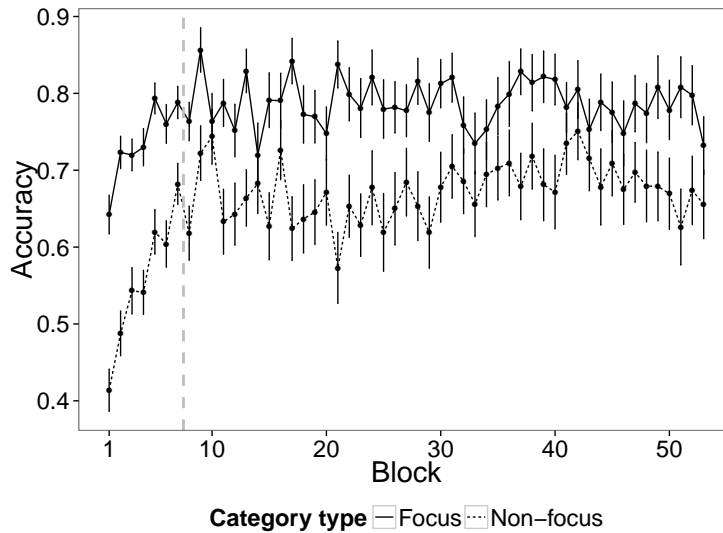
Figure 2.4: Accuracy on categorization judgments when learning the *Unitization* category structure in Experiment 1. Participants were more accurate classifying stimuli from the focus category (solid line) than the non-focus category (dotted line). The transition from the first phase of learning, which consisted of only categorization trials, to second phase that contained a mixture of categorization and whole-part discrimination trials is indicated by the vertical dashed grey line. Accuracy on categorization judgments increased during the first phase but remained stable during the second phase. The error bars indicate standard errors.
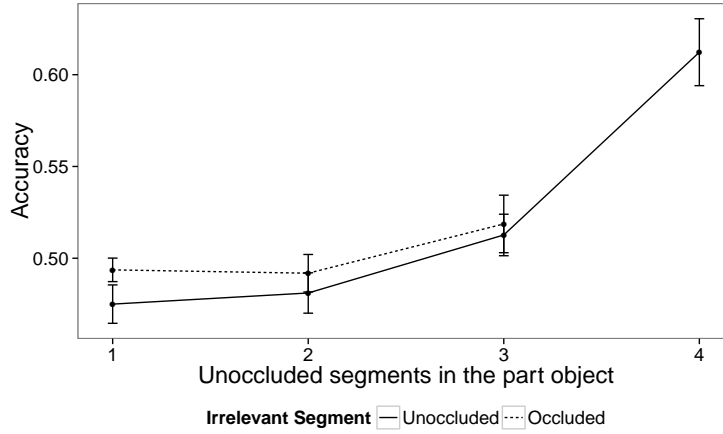
Figure 2.5: Accuracy on whole part judgments when interspersed with *Unitization* category structure trials in Experiment 1. The line type indicates if the category irrelevant segment was occluded or present in the part stimulus. Error bars indicate standard errors.
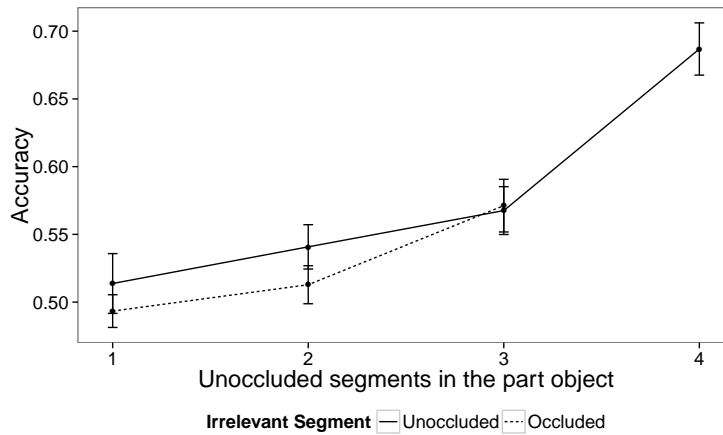

Figure 2.6: Accuracy on whole part discrimination judgments when interspersed with *Differentiation* category structure trials in Experiment 2. The line type indicates if the category irrelevant segment was occluded or present in the part stimulus. Error bars indicate standard errors.
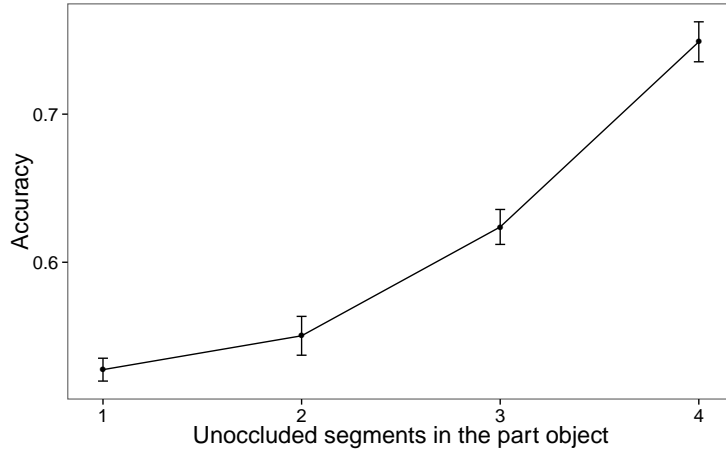
Figure 2.7: Accuracy on whole part discrimination judgments in Experiment 3. Error bars indicate standard errors.
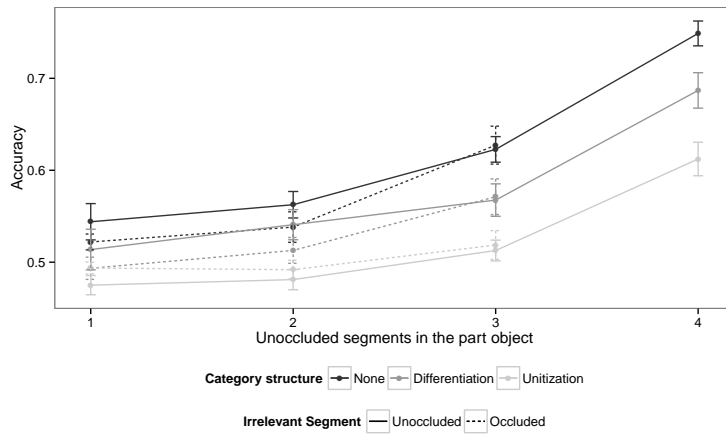

Figure 2.8: Accuracy on whole part judgments across the three category structures from Experiments 1 through 3. The line type indicates if the category irrelevant segment was occluded or present. Error bars indicate standard errors.
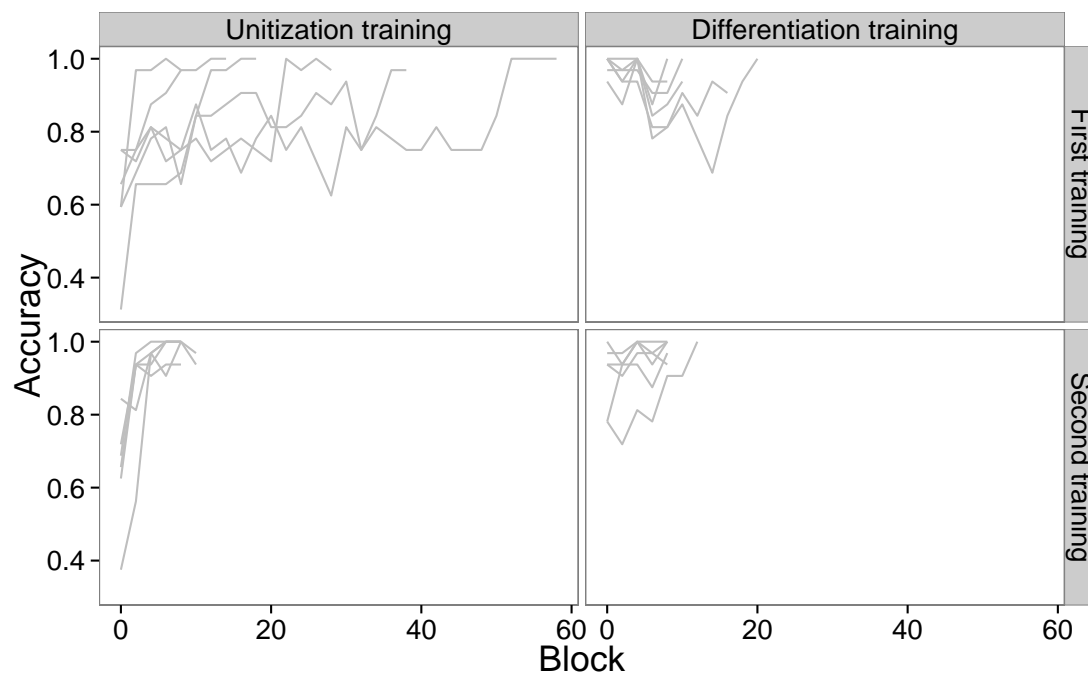
Figure 2.9: Categorization accuracy across training block until the three accuracy criteria were reached in each category learning phase in Experiment 4.
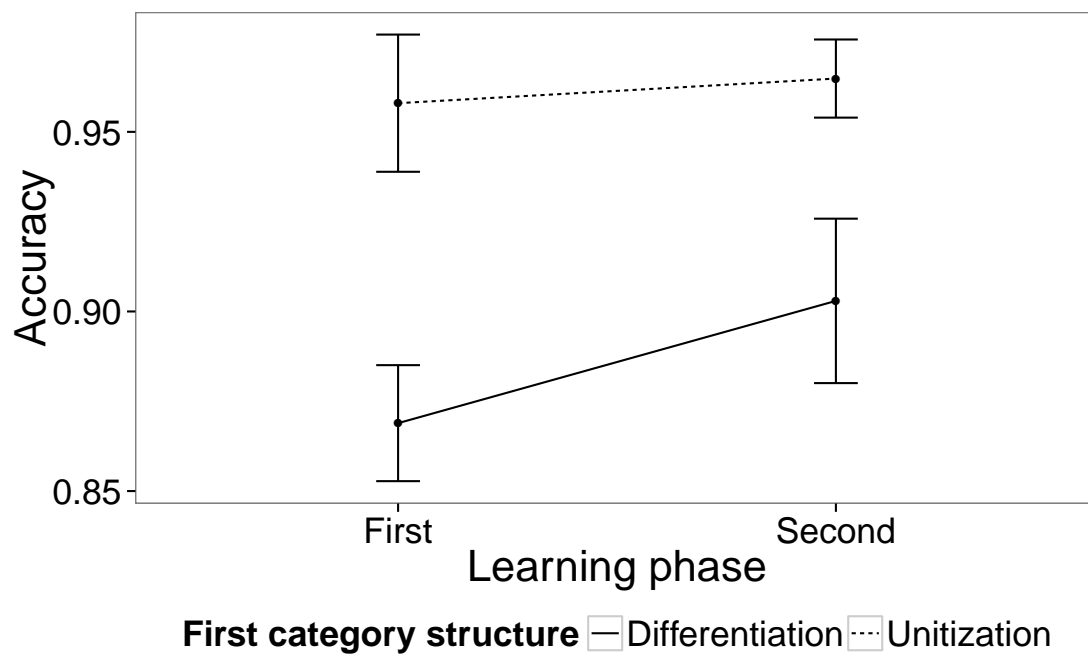
Figure 2.10: Categorization accuracy after the accuracy criteria were reached in each category learning phase in Experiment 4. Error bars indicate standard errors.
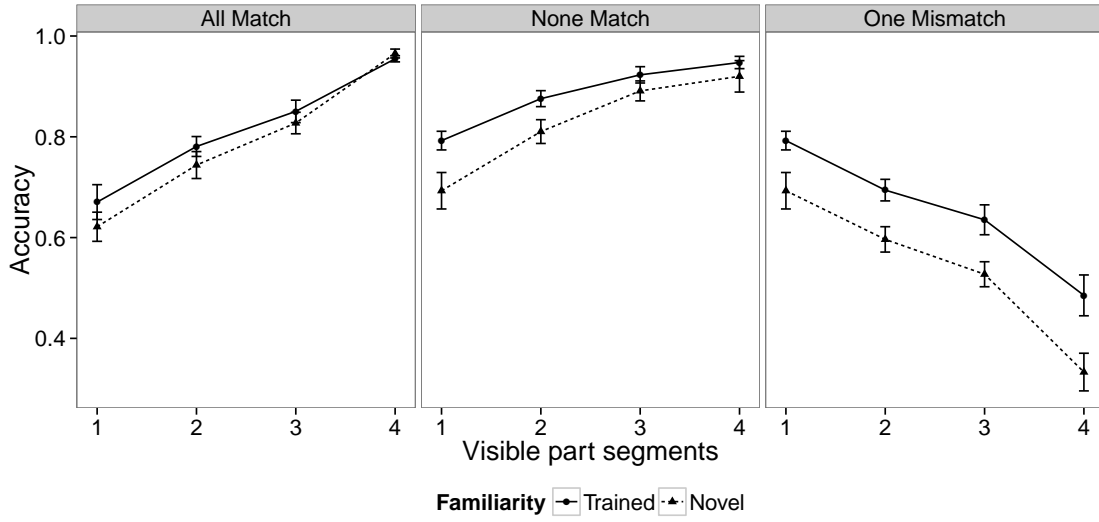
Figure 2.11: The effect of the number of segments in the part stimulus and stimulus familiarity on whole-part judgment accuracy in Experiment 4. The left panel shows trials in which all segments in the part stimulus match those in the whole stimulus. The middle panel shows trials in which no segments in the part stimulus match the whole stimulus. The right panel shows trials in which all segments in the part stimulus match the whole stimulus except for one which does not match. These results are aggregated across category structures and category orders.
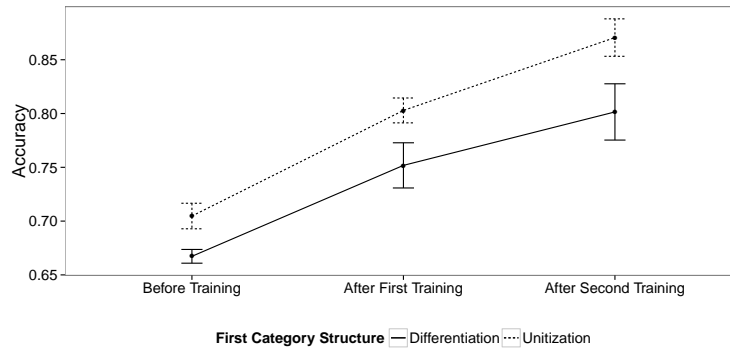
Figure 2.12: The effect on perceptual discrimination accuracy of the first category structure learned across multiple test phases in Experiment 4. The advantage in accuracy for the group of participants who learned the *Unitization* category structure first does not seem to be due to any training phase, in fact they were significantly more accurate before any training occurred. Error bars indicate standard errors.
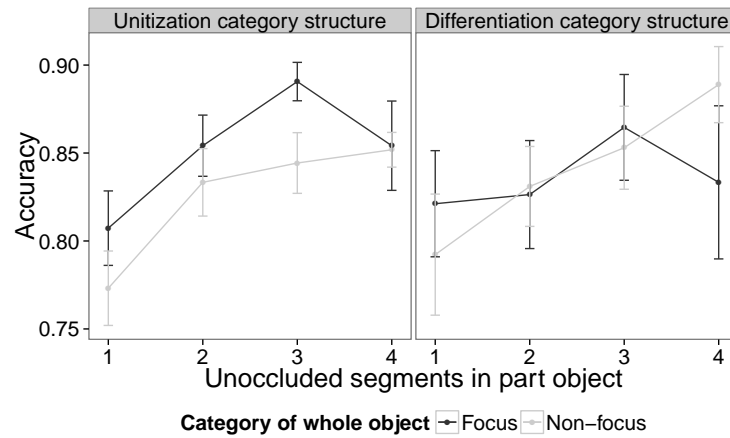

Figure 2.13: The effect on perceptual discrimination accuracy of the category membership of the whole stimulus in Experiment 4. The whole stimulus belonging to the focus category of the *Unitization* category structure improves performance after *Unitization* but not *Differentiation* category training. Error bars indicate standard errors.

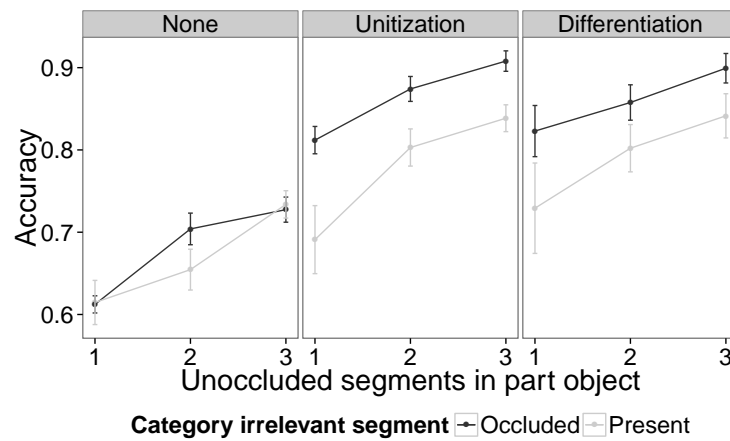Figure 2.14: The effect on perceptual discrimination accuracy of occluding the category irrelevant segment across the number of segments in Experiment 4. Occluding the category irrelevant segment significantly improves performance after *Unitization* category training. The improvement in accuracy after *Differentiation* category training is marginally significant. Error bars indicate standard errors.
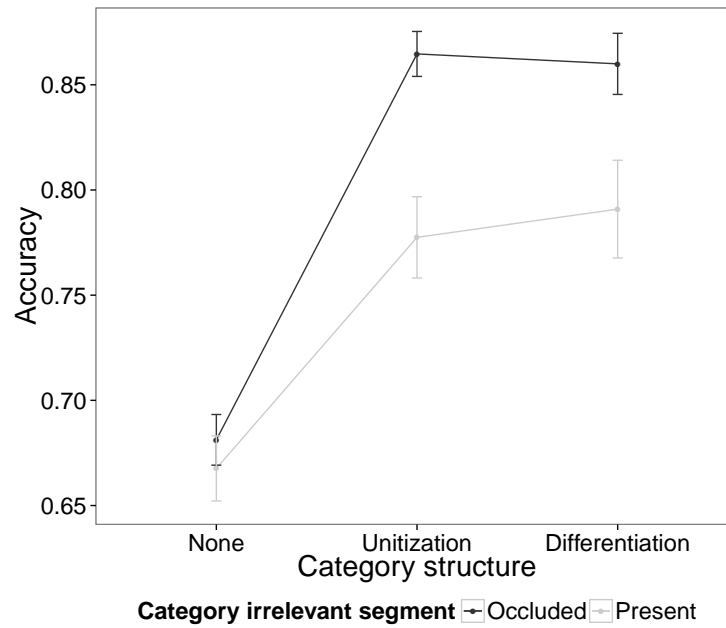
Figure 2.15: The effect on perceptual discrimination accuracy of occluding the category irrelevant segment. There is no effect of occluding the category irrelevant segments before any training, a significant effect after *Unitization* training, and a marginal effect after *Differentiation* training (details in the text). Error bars indicate standard errors.
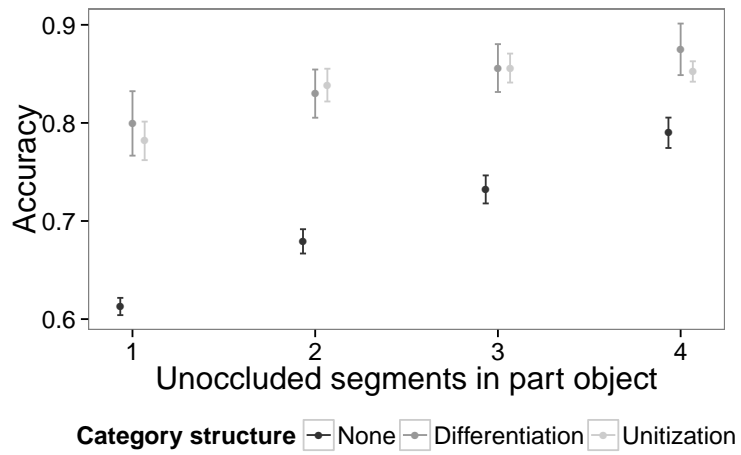
Figure 2.16: The effect of category structure and number of unoccluded segments on perceptual judgment accuracy. This plot only includes judgments of familiar stimuli in which the category irrelevant segment is occluded in the part stimulus. Training in either category structure improves accuracy, but the *Differentiation* and *Unitization* category structures have the same effect on accuracy across all number of segments in the part stimulus. Error bars indicate standard errors.

CHAPTER 3

Learning Stimulus Features

3.1 Overview

In this chapter we present the mathematics of a model from the non-parametric Bayesian framework for learning flexible feature representations (NBFF) and provide an intuition for how it infers a set of perceptual features from a set of stimuli (Austerweil & Griffiths, 2011, 2013). Particular care is taken to review the intuition for how the Indian Buffet Process (Ghahramani & Griffiths, 2005; Griffiths & Ghahramani, 2011) allows the model to infer a flexible set of features where the number of features is not predetermined. Next we apply this model to the stimuli and category structures from the four experimental paradigms presented in the previous chapter and compare the stimuli inferred by the model to the features the category structures were designed to elicit. Finally, we conclude by discussing the limitations of this model to account for all category structures with a focus on what the results indicate for models of perceptual unitization.

3.2 Model definition

Learning a set of features from a given set of stimuli requires answering two mutually-constraining questions: what do the features look like and which features are in each stimulus. At one extreme, every unique stimulus could be a single feature. This simplifies the

learning process because the mapping of features to stimuli is trivial. At the other extreme, every feature could simply be a single pixel within each stimulus. This also simplifies the learning process by making the features trivially easy to learn. Neither of these extreme solutions captures the true benefits of extracting useful features: useful features reduce the dimensionality of the stimuli representations by moving from pixels to features, and useful features describe stimuli at the level that is useful for further cognitive processing (J. J. Gibson & Gibson, 1955; E. J. Gibson, 1969). Inferring features that provide these benefits is non-trivial for computational models and in the following section we describe a model that does so and apply it to the stimuli and category structures from the previous experiments.

An instructive way to think about the problem of feature learning comes from the machine learning literature where it has been reformulated as a special case of non-negative matrix decomposition (Doshi-Velez & Ghahramani, 2009; Ghahramani & Griffiths, 2005; Wood, Griffiths, & Ghahramani, 2006; Austerweil & Griffiths, 2011, 2013). In this formulation, the goal of feature learning is to infer two matrices $\mathbf{Z}$ and $\mathbf{Y}$ such that

$$\mathbf{X} = \mathbf{Z} * \mathbf{Y} \tag{3.1}$$

where $\mathbf{X}$ is a NxD matrix in which each row corresponds to a stimulus and the nth contains the stimulus dimensions d that are present in stimulus n. $\mathbf{Y}$ is a NxK matrix of binary factors that indicate if stimulus n contains feature k. $\mathbf{Z}$ is a KxD matrix in which the kth row shows which stimulus dimensions d are present in feature k.

A random matrix decomposition in this framework would be akin to finding any random set of features and assignment of features to stimuli. There is no assurance a random decomposition would be a useful set of features for describing the stimuli. What is needed is a prior that favors parsimonious sets of features without limiting the possible set of

features that can be inferred. Austerweil and Griffiths (2011, 2013) present the NBFF framework for specifying non-parametric Bayesian models of feature learning. Using Bayes theorem, the goal of finding $\mathbf{Z}$ and $\mathbf{Y}$ can be transformed into:

$$P(\mathbf{Z}, \mathbf{Y} | \mathbf{X}) \propto P(\mathbf{X} | \mathbf{Z}, \mathbf{Y}) P(\mathbf{Z}) P(\mathbf{Y}) \tag{3.2}$$

where priors $P(\mathbf{Z})$ and $P(\mathbf{Y})$ must be specified for the matrices $\mathbf{Z}$ and $\mathbf{Y}$ and a likelihood function $P(\mathbf{X} | \mathbf{Z}, \mathbf{Y})$ must be defined.

The most basic prior across matrix $\mathbf{Y}$ is to assume each entry in the matrix (k, d) is independent of all other entries. If the probability of any pixel being on is ϕ then the prior of matrix $\mathbf{Y}$ is given by:

$$P(\mathbf{Y} | \phi) = \prod_{k,d} \phi^{y_{k,d}} (1 - \phi)^{1 - y_{k,d}} \tag{3.3}$$

Defining a prior across feature ownership matrix $\mathbf{Z}$ is more complex. The NBFF the Indian Buffet Process (IBP) as a prior on $\mathbf{Z}$ (Ghahramani & Griffiths, 2005). The IBP is a non-parametric probability distribution across binary matrices that have a fixed number of rows (N stimuli) and a potentially unbounded number of columns (K features where K is unknown). The IBP has two aspects that make it useful as a prior across feature matrices. The number of potential features K is not fixed and can grow to whatever is necessary; yet the IBP can be parameterized such that it is biased toward representations with fewer features (Ghahramani & Griffiths, 2005; Griffiths & Ghahramani, 2011; Austerweil & Griffiths, 2011, 2013). This combination in the IBP of preferring fewer features but having the flexibility to expand given the complexity of the stimuli make it a natural prior for learning new features. The full derivation of the IBP distribution is presented by Ghahramani and Griffiths (2005), but the resulting equation for the probability of a particular matrix $\mathbf{Z}$

depends only on one parameter α:

$$P(\mathbf{Z}|\alpha) = \frac{exp(-\alpha H_N)a^{K_+}}{\prod_{h=1}^{2^{N-1}} K_h!} \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!} \tag{3.4}$$

where N is the number of stimuli, H_N is the Nth harmonic number ($H_i = \sum_{j=1}^{i} j^{-1}$), K_+ is the number of unique features in $\mathbf{Z}$, and K_h is the number of features in $\mathbf{Z}$ that appear in exactly the same set of stimuli.

Defining priors for $\mathbf{Z}$ and $\mathbf{Y}$ is not sufficient to fully specify all of the terms needed from Equation 3.2. A function is required, $P(\mathbf{X}|\mathbf{Z}, \mathbf{Y})$, that defines the likelihood of the matrix of stimuli $\mathbf{X}$ given specific matrices $\mathbf{Z}$ and $\mathbf{Y}$. Wood et al. (2006) show that the noisy-OR function is an appropriate likelihood function when the stimulus dimensions of both stimuli and features are binary properties. The noisy-OR function defines the probability that a given stimulus dimension d is present in stimulus n as:

$$P(X_{n,d} = 1|\lambda, \epsilon, \mathbf{Y}, \mathbf{Z}) = 1 - (1-\lambda)^q(1-\epsilon) \tag{3.5}$$

where ϵ is a parameter indicating the probability a stimulus dimension is present without regard to features, λ is a parameter indicating the probability a stimulus dimension is present in an item given a feature indicates it should be present, and q is the inner product of the nth row of $\mathbf{Z}$ and the dth column of $\mathbf{Y}$ (Austerweil & Griffiths, 2011, 2013). Assuming all $X_{n,d}$ are independent, the full likelihood is given by Equation 3.6.

$$P(\mathbf{X}|\lambda, \epsilon, \mathbf{Y}, \mathbf{Z}) = \prod_{n,d} P(X_{n,d}|\lambda, \epsilon, \mathbf{Y}, \mathbf{Z}) \tag{3.6}$$

Equations 3.3, 3.4, and 3.6 define the feature learning model proposed by Austerweil and Griffiths (2011). This model was subsequently expanded to include category labels as part

of the stimuli representation (Austerweil & Griffiths, 2013). Category labels are treated as additional binary stimulus dimensions that increase the number of columns in the $\mathbf{X}$ matrix. An inferred feature f in the matrix $\mathbf{Z}$ is associated with or predictive of category membership if the binary indicator in row f is ON in the stimulus dimension columns that correspond to that category label.

3.3   Learning features

In this section we apply the NHFF model defined earlier to the stimuli and category structures from the experiments in the previous chapter.

3.3.1   Stimuli

16 stimuli were constructed by sampling eight random segments (A, B, C, D, W, X, Y, and Z), assigning two segments to each of the four positions (A and W were paired, B and X, C and Y, and finally D and Z), and constructing all 16 unique pairings of those features shown in the first column of Table 3.1.

A stimulus was originally constructed as a 360 by 360 matrix of binary pixels but these matrices were transformed to vectors of length 129,600. The number of pixels in each vector was reduced to 2,373 by removing more than 99% of the pixels that did not contain an *on* pixel in any stimulus. A small proportion of pixels in the reduced stimuli were randomly flipped, $p(flip) = 1/75$, to produce the actual stimuli used in these analyses. These stimuli are shown in Figure 3.1.[1] The 16 stimuli were combined to form an $\mathbf{X}$ matrix with 16 rows and 2,373 columns. Roughly half of the pixels in each row of $\mathbf{X}$ were *on*.

---

[1]The pixels that matched across all stimuli are added back into to these visualizations to create interpretable figures. The redundant stimulus dimensions were removed solely to decrease the duration of simulations which are heavily dependent on the number of stimuli and stimulus dimensions. Removing these redundant dimensions did not materially alter the features the model inferred for each category structure.

3.3.2 Category information

Three separate simulations were run with different category information added to the stimulus matrix **X**. 954 columns of category information were added to **X** in each simulation. These columns constituted approximately 40% of the original number of columns in **X**.

The category information added to the *No category* structure simulation did not contain any *on* pixels. This set of stimuli and category information corresponded to the stimulus information available in Experiment 3 and Experiment 4 in the testing session before any category training. The *No category* simulation reflects the features that the NHFF model extracts from the stimuli without any category information to bias the feature inference process.

The category information added to **X** for the *Differentiation* category structure simulation consisted of six unique columns that were each repeated 159 times. These columns of category information are shown in Table 3.1 in the *Differentiation* category information columns. The first two of these category information columns were perfectly correlated with the A and W segments being present in the stimulus. The pixels in the first column were *on* if the stimulus contained the W segment, and the pixels in the second column were *on* if the stimulus contained the A segment. The second set of two category information columns were correlated with the B and X segments, and the third set of two category information columns were correlated with the C and Y segments. These category information columns mirror the *Differentiation* category structure in Experiments 2 and 4. For example, the category information columns sensitive to segments A and W were perfectly predictive of the *Differentiation* category structure defined by those segments.

The category information added to **X** for the *Unitization* category simulations consisted of two unique columns that were each repeated 477 times. One column corresponded to the focus category of the *Unitization* category structure in Experiments 1 and 4. The pixels in

the focus category information column were *on* when either all three segments A, B, and C were present in the stimulus or when W, X, and Y were present. The pixels in the non-focus category information column followed the exact opposite pattern (Figure 3.1).

3.3.3 Simulation details

Each simulation was run using a modified version of the Gibbs sampler written for simulating the IBP with a noisy-OR likelihood function and distributed as part of the supplemental materials of Austerweil and Griffiths (2013). This implementation was based on code from Wood et al. (2006). The simulations were run for 4,000 samples with the first 500 samples discarded to minimize the effect of the initial conditions. The parameter values for all simulations were set to: $\alpha = 2$, $\lambda = 0.99$, $\epsilon = 0.01$, and $\phi = 0.5$. Additional simulations were run with small changes to these parameter values and they did not produce noticeable changes in the features inferred in each category structure.

3.4 Results

The features inferred from the *No category* stimuli are shown in Figure 3.2. These features roughly correspond to the eight independently-varying segments that were used to construct the set of stimuli but all of the features are missing some pixels that are *on* in the stimuli. The "missing" pixels from the features are present in both segments that occur in that position and the model did not include those pixels in any feature. (Austerweil & Griffiths, 2011) explain that this type of behavior is caused by the assumption in the prior on **Y** that pixels are *on* independent of their neighbors. This issue can be addressed by augmenting the model to add spatial constraints and neighbor effects into the prior on **Y**.

The features inferred from the simulation with the *Differentiation* category information are shown in Figure 3.3. The features in the first three columns were each associated with

a single category label from the category information added to the stimuli. Adding the *Differentiation* category information did not lead to inferring features noticeably different from the NO CATEGORY features. However, the model did infer reasonable associations between features and category labels.

Adding the *Differentiation* category information pixels did not prevent the model from learning two features that did not map onto any category information but did correspond to the two segments not predictive of any category information. These category irrelevant features were also inferred by the model. However, the empirical results show that after learning the *Differentiation* category structure participants were more accurate on judgments that did not contain the category irrelevant segment (Figure 2.14). This suggests that the NHFF model alone cannot account for the learning results in the *Differentiation* condition. In the next chapter we outline a model that uses these perceptual features but learns attention weights for each feature to account for these results.

The features inferred from the simulation with the *Unitization* category information are shown in Figure 3.4. The two features in the first row were associated with the focus category. The first feature in this row roughly corresponds to a combination of the A, B, and C segments that as a conjunction indicate the stimulus belongs in the focus category. Yet the second feature in the top row does not correspond to the other combination of segments (W, X, and Y) that also predict the focus category. This second feature contains approximately 15% of the pixels from each of the W, X, and Y segments, and is more than 90% composed of pixels from those features, but is not a strong representation of a conjunction of the W, X, and Y segments into a single feature. The three features in the second row were associated with the non-focus category. These three features correspond to the A, B, and C segments, respectively. These segments were present in stimuli within the non-focus category, but were not diagnostic of category membership by themselves. The

five features in the third row were not associated with either category because they did not contain any *on* category information pixels. The first two features in the third row correspond to the D and W features that were never diagnostic of category membership. These features match the non-diagnostic features inferred from the *Differentiation* category structure that were not associated with a category. The remaining three features in the third row correspond to the W, X, and Y segments and it is not immediately clear why they were not associated with a category.

The inability of the NHPP model to extract two unitized features was not limited to a particular set of parameters for the model. A set of simulations that vary these parameters, especially the ϕ parameter that is the prior belief that a pixel is *on* in a feature, did not produce one feature representation that included two features that correspond to the conjunctions of segments ABC and WXY.

3.5   Discussion

A computational model based on the non-parametric Bayesian framework for learning flexible features (Austerweil & Griffiths, 2013) learned different sets of perceptual features across the three category structures. The simulations that included *No category* information and the *Differentiation* category structure both inferred perceptual features that matched the individual segments of the stimuli. This suggests that category training in the *Differentiation* might not produce qualitatively different perceptual features than training that merely exposes people to the stimuli.

The simulations that included the *Unitization* category structure information did not infer the same set of features as the other category structures. In addition to inferring the eight features that match the stimulus segments, this simulation inferred two additional features that correspond to more than one segment. One of these features corresponds to the

three category relevant segments in a conjunction that is predictive of category membership, but the other feature only roughly resembles the other category predictive conjunction.

What prevented the model from inferring both unitized features that correspond to the conjunction of the ABC and WXY segments in the *Unitization* category structure? This is particularly challenging because Austerweil and Griffiths (2011) report the same model was able to infer features that are composed of conjunctions of segments from their own experimental results as well as the conjunctive features from the visual search task from (Shiffrin & Lightfoot, 1997). Furthermore, Austerweil and Griffiths (2013) infer conjunctive features that are informed by category structures due to a quick imprinting process (Schyns & Rodet, 1997), as well as a more slowly developing unitization process (Pevtzow & Goldstone, 1994). The *Unitization* category structure is similar to the structures from those experiments, yet it is more complex in one key aspect: the *Unitization* category structure cannot be correctly classified by learning a single conjunction of segments, it is an exclusive-OR structure and exclusive-OR category structures are harder for people to learn (Shepard et al., 1961) and require more complex representations to learn than conjunctions (Minsky & Seymour, 1969). These simulation results suggest the NHFF framework struggles to learn unitized perceptual features from exclusive-OR category structures.

It is not clear from these results if people learned unitized perceptual features when training on the *Unitization* category structure and the model cannot capture that learning, or if the model indicates people struggled to learn features in the *Unitization* training condition. However, the NHFF model does not extract the same set of perceptual features across all category structures, and would predict differences in subsequent perceptual processing due to the different feature vocabularies (Austerweil & Griffiths, 2011). One possible interpretation of these results is that the perceptual features learned during the *Unitization* category training may not perfectly align with the composite features that perfectly predict

category membership.

| | Category Information Columns Added | |
Stimulus	Differentiation	Unitization
ABCD	0 1 0 1 0 1	1 0
ABCZ	0 1 0 1 0 1	1 0
ABYD	0 1 0 1 1 0	0 1
ABYZ	0 1 0 1 1 0	0 1
AXCD	0 1 1 0 0 1	0 1
AXCZ	0 1 1 0 0 1	0 1
AXYD	0 1 1 0 1 0	0 1
AXYZ	0 1 1 0 1 0	0 1
WBCD	1 0 0 1 0 1	0 1
WBCZ	1 0 0 1 0 1	0 1
WBYD	1 0 0 1 1 0	0 1
WBYZ	1 0 0 1 1 0	0 1
WXCD	1 0 1 0 0 1	0 1
WXCZ	1 0 1 0 0 1	0 1
WXYD	1 0 1 0 1 0	1 0
WXYZ	1 0 1 0 1 0	1 0

Table 3.1: Category labels added to each stimulus. In this table the actual segments of the stimuli are represented with binary features. The Differentiation Category features added to the stimuli map one-to-one onto the first three segments of the stimuli and the Unitization Category features are only on for stimuli with all 0s or all 1s in the first three segments.
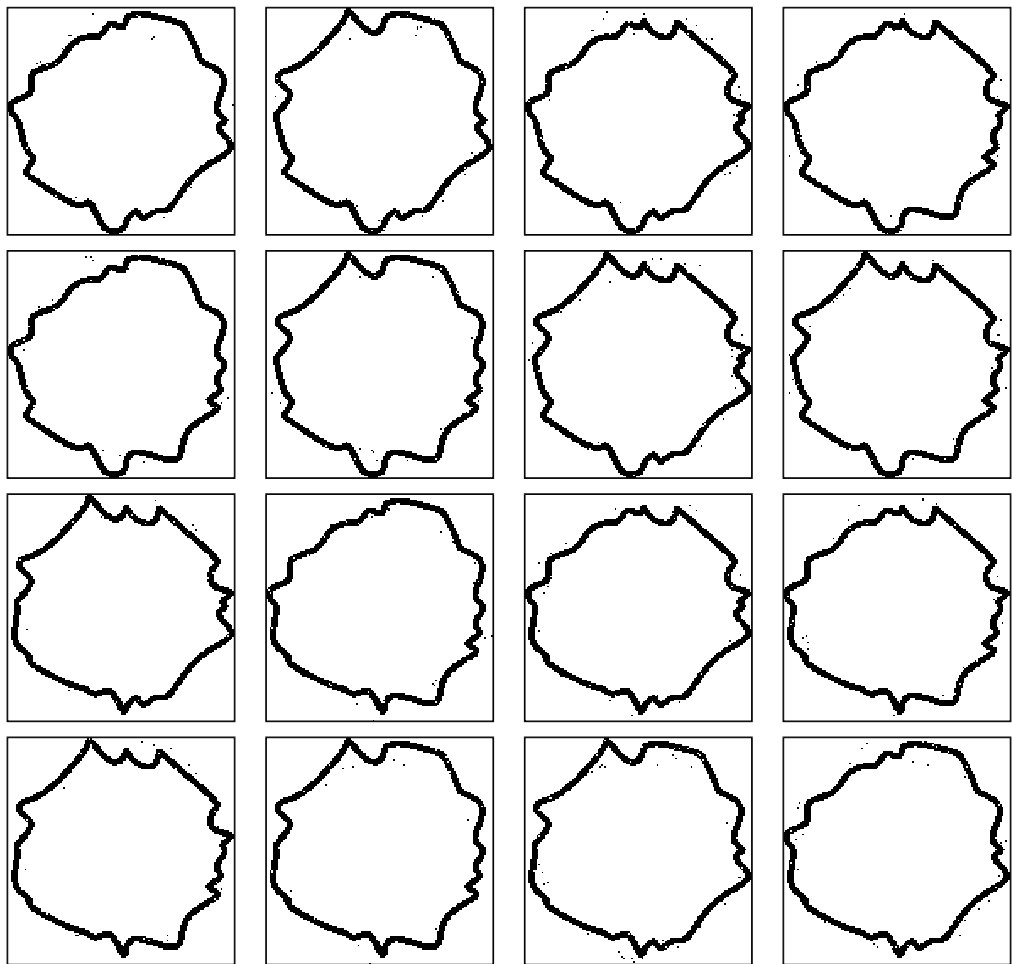
Figure 3.1: The 16 stimuli randomly selected for training the NBFF model with a small amount of noise added to each stimulus. The stimuli in the left column are the four stimuli in the focus category, the top two stimuli share the same segments (A, B, and C) in all positions except the lower right, and the bottom two stimuli share a different set of segments in the same positions (W, X, and Y).
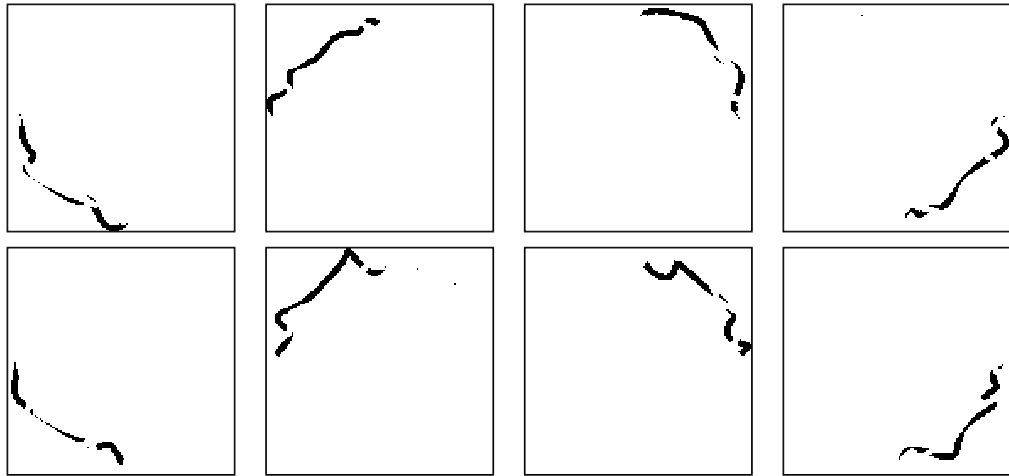
Figure 3.2: The features inferred by the NHFF model when no category information is added to the stimulus representation. These features closely match the eight segments that all segments are composed of.
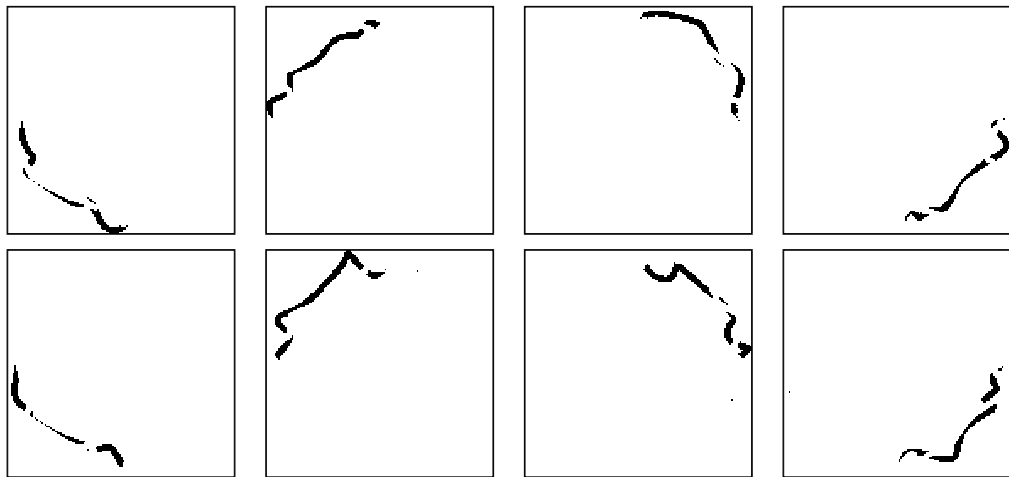

Figure 3.3: The features inferred by the NHFF model when category information consistent with the *Differentiation* category structure was included in the stimulus representation. The features in the first three columns of the figure were all associated with one of the six category information pixels. These features closely match the six segments that are predictive of category membership in Table 3.1. The features in the fourth column did not contain any category information but correspond to the segments that were not predictive of any category label.
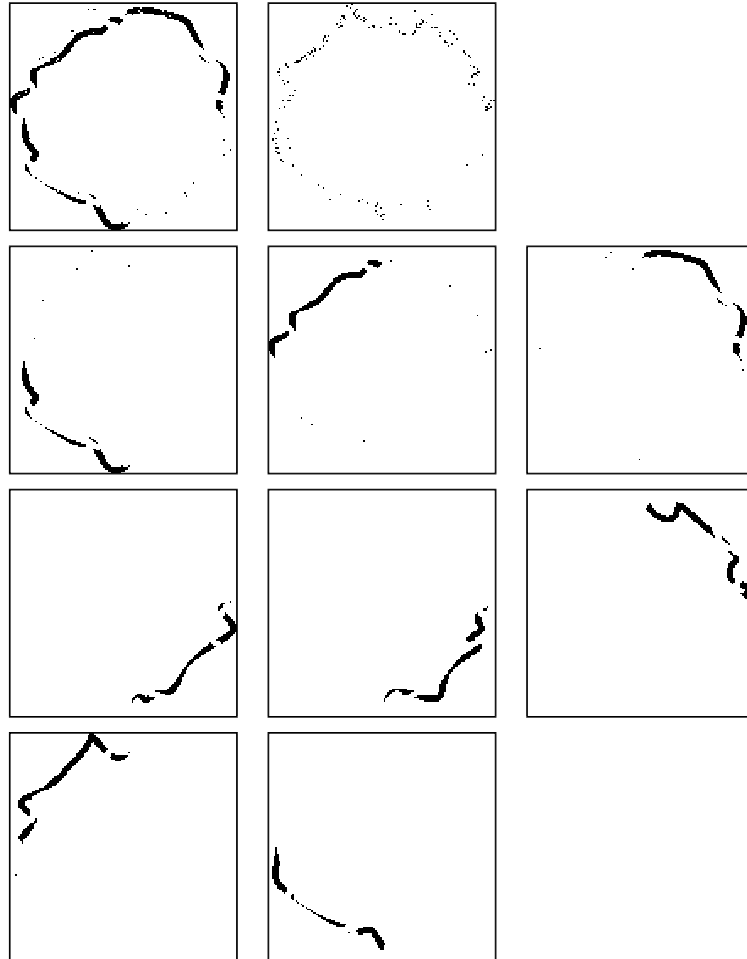
Figure 3.4: The features inferred by the NHFF model when category information consistent with the *Unitization* category structure was included in the stimulus representation (Table 3.1). The features in the first row of the figure contained stimulus pixels that indicated they were associated with the focus category, The features in the second row contained stimulus pixels that indicated they were associated with the non-focus category, and the features in the third and fourth rows did not contain any category information pixels.

CHAPTER 4

Modeling feature vocabularies

4.1 Overview

In this section we develop a modeling framework to test which perceptual features partic-
ipants were using when making the perceptual discrimination judgments before and after
category training. Perceptual features, in this context, do not necessarily have a one-to-one
correspondence with a single segment. Instead, these features detect sets of segments that
are processed quickly and automatically as a group (Goldstone, 2000). The set of segments
we consider range from encompassing a single segment, producing a one-to-one mapping
between that feature and a segment, to features that encompass to all four segments of a
stimulus.

The remainder of this section specifies a modeling framework to compare the fit of
computational models using different vocabularies of features to data from the experimental
results. We begin with a review of the role of similarity in computational models of choice
and explain how models of categorization (Medin & Schaffer, 1978; Nosofsky, 1986) and
identification (Luce, 1963; Shepard, 1957; Nosofsky, 1984, 1986) have been extended to
account for performance in old-new recognition memory and other single category judgment
tasks (Nosofsky, 1988a, 1991). This old-new recognition memory GCM framework will be
adapted to the task-specific structure of predicting same-different judgments in the whole-

part perceptual discrimination task. The best-fitting parameters and overall performance of models assuming different sets of perceptual features will be compared.

4.2 Similarity across features

The similarity between objects has been a fundamental aspect of computational models of choice behavior since the earliest cognitive models (Luce, 1963; Shepard, 1957). Early accounts of similarity have focused on the importance of weighting evidence across many stimulus features or dimensions to compute the similarity between two objects (Tversky, 1977; Shepard, 1964). These principles were foundational to exemplar-based models of categorization in which choice behavior was predicted by the relative similarity of a target item to all examples from multiple categories (Medin & Schaffer, 1978; Nosofsky, 1984, 1986). The GCM framework was later extended to account for choice behavior in old-new recognition memory tasks (Nosofsky, 1988a, 1991; Shin & Nosofsky, 1992; Zaki & Nosofsky, 2001; Knapp, Nosofsky, & Busey, 2006). This formulation of the GCM is particularly relevant to modeling the whole-part perceptual judgment task because the similarity of a test item to a set of exemplars is not compared to another similarity to produce a response (Luce, 1963; Shepard, 1957; Tversky, 1977; Medin & Schaffer, 1978; Nosofsky, 1986) but probabilistically compared to a similarity threshold to determine a choice response.

Shin and Nosofsky (1992) present a clear formulation for how this framework predicts old-new decisions using the similarity derived by summing across weighted feature dimensions. The probability of responding "old" to stimulus x_i is a function of response bias b and S_i

$$P(old|S_i) = \frac{S_i}{S_i + b} \tag{4.1}$$

$$S_i = \sum_j S_{ij} \tag{4.2}$$

$$S_{ij} = exp(-cD_{ij}) \tag{4.3}$$

where S_i is a function of the sensitivity parameter c and the distance D_i. D_i is defined to be the sum of the distances d_{ij} between stimulus x_i and all previous stimuli j in the set previous stimuli in memory J such that

$$D_i = \sum_{j \in J} d_{ij} \tag{4.4}$$

$$d_{ij} = \sum_{k \in K} w_k |x_{ik} - x_{jk}| \tag{4.5}$$

where d_{ij} is the weighted sum of the distance between x_i and x_j across all features or feature dimensions k in the set of all feature dimensions K. The free parameters of the model are b, c, and the weights w_k. All parameters are constrained to be positive and the the weights are further constrained such that $\sum_{kinK} w_k = 1$, and thus $0 < w_k < 1$ for all k.[1]

To predict same-different judgments in the whole-part perceptual judgment task, a special case of the recognition GCM is used. Generally the recognition memory GCM is applied to situations when many exemplars are stored in memory and an overall similarity between those items and a test item is used for recognition (Nosofsky, 1988a). In this task the whole stimulus W is only relevant exemplar to compare the part stimulus P to. Therefore the similarity S_i in the full model reduces to the similarity between the whole and part stim-

---

[1]This formulation assumes a Minkowski r-metric of 1 and a γ value of 1. This has been done for simplicity since those values were left as constants across all models.

uli S_{WP}. A similar formulation was developed by Cohen and Nosofsky (2000) to predict reaction times in same-different judgments.

$$P(same|S_{WP}) = \frac{S_{WP}}{S_{WP} + b} \tag{4.6}$$

$$S_{WP} = exp(-cD_{WP}) \tag{4.7}$$

$$D_{WP} = \sum_{f \in F} w_f |x_{Wf} - x_{Pf}| \, \mathbb{1}_f(P) \tag{4.8}$$

The x_{if} values are defined to be 1 if feature f is present in stimulus i and 0 otherwise. Therefore the term $|x_{Wf} - x_{Pf}|$ has the value 0 if both the part or whole stimulus or neither stimulus contains feature f and 1 otherwise. However, because some segments are occluded in the part stimulus, an indicator function $\mathbb{1}_f(P)$ has been added to Equation 4.8 that does not exist in Equation 4.5. This function has the value 0 if feature f is occluded in part P and 1 otherwise. With this term included, features that are occluded or partially occluded in the part stimulus do not contribute to the perceptual distance between the whole and part stimuli.

One interesting property of this model is that when the whole and part match across all unoccluded features then $D_{WP} = 0$ and $S_{WP} = 1$. This occurs regardless of how many features are included in the feature vocabulary or the number of features present in the part stimulus. To account for less than perfectly accurate discrimination on these trials, the model adjusts the value of the b parameter to adjust the background noise of responses (Equation 4.7). Having a single SHARED b parameter across all trials might be an unreasonable constraint, particularly if people adjust their response biases based on the amount of information available in the part stimulus. A more flexible method of parameterizing the b (and c) parameters is to allow them both to freely vary based on the number of unoccluded segments in the part stimulus. This parameterization is characterized as the FLEXIBLE response strategy. Both the FLEXIBLE and a SHARED parameterization was

fit for all models. An alternative modeling framework that could be considered instead of allowing the decision parameters to change is the Feature Contrast Model (Tversky, 1977). The probability of responding "same" in the Feature Contrast Model increases from a baseline response level as more features match between the part and whole. Both Zaki and Nosofsky (2001) and Knapp et al. (2006) present a discussion of the relative merits of these approaches in the context of old-new recognition judgments.

The whole-part GCM is closely related to the Context Theory Model (Medin & Schaffer, 1978). As noted by Nosofsky (1986, p. 42), when features are binary and c is fixed in the GCM, then Equations 4.7 and 4.8 are functionally equivalent to the definition of similarity, S_{xy} between two items x and y, in the Context Theory Model. This similarity is computed across all features f in the set F:

$$S_{xy} = \prod_{f \in F} s_{fxy} \tag{4.9}$$

where $s_{fxy} = 1$ if x and y share the feature f and $0 < s_{fxy} < 1$ otherwise. In the whole-part GCM, features are binary but c is allowed to vary across stimuli, resulting in a slightly different definition of similarity than the Context Theory Model.

4.3 Feature vocabularies

One critical aspect of the whole-part GCM is that it is agnostic to the set of features that comprise the feature vocabulary. This allows it to be fit using different vocabularies of features and the relative quality of the vocabularies can be assessed using model selection techniques. The remainder of this section defines three possible feature vocabularies and compares how well each feature vocabulary accounts for the perceptual discrimination judgment data from Experiment 4. The three feature vocabularies are illustrated in Table 4.1.

Feature sets for whole stimulus *ABCD* and part stimulus *ABY-*

Analytic			Unitized category relevant			Powerset		
Whole	Part	Match	Whole	Part	Match	Whole	Part	Match
A	A	match	A	A	match	A	A	match
B	B	match	B	B	match	B	B	match
C	Y	mismatch	C	Y	mismatch	C	Y	mismatch
D	-	occluded	D	-	occluded	D	-	occluded
			ABC	ABY	mismatch	AB	AB	match
						AD	–	occluded
						BC	BY	mismatch
						CD	–	occluded
						ABC	ABY	mismatch
						ABD	—	occluded
						ACD	—	occluded
						BCD	—	occluded
						ABCD	—-	occluded

Table 4.1: A table of the features used to make similarity judgments for each feature vocabulary for an example whole and part stimulus. The whole stimulus is composed of four segments, ABCD, and the part stimulus consists of three segments, ABY, and the fourth position is occluded. The whole stimulus is represented as four features when assuming the ANALYTIC feature vocabulary and the part stimulus has three features. Two of those features match, one mismatches, and one is occluded in the part stimulus and thus does not contribute to similarity judgments. When assuming the UNITIZED CATEGORY RELEVANT feature vocabulary the whole stimulus is composed of five features, the same four as in the ANALYTIC vocabulary plus the three-segment unitized feature. The POWERSET feature vocabulary produces a representation of the whole stimulus that consists of 14 features and the part stimulus has 6 features (with 8 features occluded in the part stimulus representation).

The perceptual features people use could potentially vary from a single pixel to a full stimulus from Experiments 1 through 4 that were composed of 129,600 pixels. However, all stimuli were constructed by combining four segments, and in the following analyses we assume that perceptual features are combinations of those segments.[2] Perceptual features cannot contain a portion of the pixels that "belong to" a segment; every segment is either completely within a feature or not. We will compare three sets of features: the ANALYTIC, the UNITIZED CATEGORY RELEVANT, and the POWERSET feature vocabularies.

The first set of features we consider is the ANALYTIC feature vocabulary. It consists of four psychological features that have a one-to-one mapping with the true independent segments in the part and whole stimuli. The first two columns of Table 4.1 indicate how these features are defined for a sample whole and part stimuli. The whole stimulus, composed of the segments A, B, C, and D, is represented by four features: A, B, C, and D because each segment maps onto a unique feature in the ANALYTIC feature vocabulary. The part stimulus, composed of segments A, B, Y, and no segment in the occluded fourth position, is represented by three features: A, B, and Y. The first two features match between the part and whole stimuli, third feature mismatches, and the fourth feature is occluded in the part stimulus. The distance between the two stimuli would be computing the weighted sum across all of these features. The ANALYTIC feature vocabulary has four total feature weights, one for each feature, and three free parameters for the weights since the weights must sum to one.

Second, the UNITIZED CATEGORY RELEVANT feature vocabulary contains the four features from the ANALYTIC vocabulary and adds one additional feature. This new feature is composed of the three segments that are relevant for both the *Unitization* and *Differen-*

---

[2]This constraint is relaxed in the previous chapter where the Austerweil and Griffiths (2013) nonparametric Bayesian framework for flexible feature discovery is applied to the pixel-level representation of these stimuli.

tiation category structures. In the example shown in Table 4.1 the whole stimulus would contain the unitized feature ABC and the part stimulus would contain the unitized feature ABY. These two unitized features (ABC and ABY) do not match and would be considered a mismatch in the whole-part GCM. The UNITIZED CATEGORY RELEVANT feature vocabulary has five total feature weights, four of which are free parameters. We consider two rules for how the three-segment unitized feature combined with the three category relevant one-segment features in the UNITIZED CATEGORY RELEVANT feature vocabulary. First, the ADD rule allows the unitized feature and the analytic features to both contribute to similarity simultaneously: the presence of the unitized feature does not influence how the analytic features contributed to similarity. Second, the REPLACE rule gates the contribution of the component one-segment features based on the presence of the three-segment feature. If the unitized feature is present in the part stimulus then the three one-segment category relevant features do not contribute to the similarity for that judgment. The UNITIZED CATEGORY RELEVANT feature vocabulary corresponds to an idealized version of the perceptual features inferred from the *Unitization* category structure by the Bayesian model (Austerweil & Griffiths, 2013) in the previous chapter.

Finally, the POWERSET feature vocabulary is the largest possible set of psychological features for these stimuli. This feature vocabulary consists of 13 total features: the four features from the ANALYTIC model that correspond to individual segments, the four features composed of all adjacent pairs of segments, the four features composed of all sets of three adjacent segments, and the feature composed of all four segments. The POWERSET feature vocabulary has 14 feature weights, 13 of which are free parameters. Table 4.1 shows the 14 features for a sample whole and part stimuli.

4.3.1 Modeling details

For each feature vocabulary, the whole-part GCM was fit with two parameterizations of the b and c parameters. The SHARED parameterization consisted of finding the single best fitting b and c parameters across all whole and part stimuli. This parameterization had two free parameters and reflects a responding strategy that is insensitive to the number of occluded segments in the part stimulus. The FLEXIBLE parameterization was also fit with a freely varying b and c parameter for each unique number of unoccluded segments in the part stimulus. This parameterization required six additional free parameters relative to the SHARED parameterization. Both the SHARED and FLEXIBLE parameterizations were included for all feature vocabulary models because they reflect differences in decision strategies that are independent of the feature vocabulary. The SHARED parameterization will provide the best account for whole-part discriminations if the threshold for making responses does not depend on the number of unoccluded segments in the part stimulus and the FLEXIBLE parameterization will be a better account if participants adjust their response criterion based on the number of unoccluded segments in the part stimulus. Table 4.2 shows a summary of all computational models fit to the data including the number of free parameters in each model.

4.4 Parameter estimation and model assessment

A search was conducted for the free parameters of each model that provided a maximum-likelihood fit to the part-whole judgment data for each testing session of each participant. This search was done by maximizing the log-likelihood of the data given the model

$$ln(L) = \sum_i ln \left(\binom{N_i}{z_i} p_i^{z_i} (1 - p_i)^{N_i - z_i} \right) \tag{4.10}$$

where i is the ith trial type, N_i is the frequency of i, z_i is the frequency of responding "same" to i, and p_i is the estimated probability of responding "same" to stimulus i and

| Feature vocabulary | Free parameter Count | | |
| | | *b* and *c* | |
	Feature weights	Flexible	Shared
Analytic Features	3	8	2
Unitized category relevant (replace)	4	8	2
Unitized category relevant (additive)	4	8	2
Powerset	13	8	2

Table 4.2: The parameter count for each combination of feature vocabulary and decision rule. The flexible *b* and *c* parameter models had four independently-varying *b* and *c* parameters that were used based on how many segments were present in the part stimulus.

is a function of the model parameters. This likelihood function assumes that "same" and "different" responses are made independently of each other and binomially distributed. The search through parameter space was executed via particle swarm optimization using the PSO package in R. The optimization algorithm was run with the default package settings of 5000 search iterations and with the number of particles in the swarm proportional to the number of free parameters $(10 + 2 * parameters)$.

In order to compare the performance of models with different numbers of parameters the AIC (Akaike, 1974) and BIC (Schwarz, 1978) were computed for each model. The equation for these information criteria are

$$AIC = -2ln(L) + 2P \qquad (4.11)$$

$$BIC = -2ln(L) + Pln(N) \qquad (4.12)$$

where $ln(L)$ is the log-likelihood of the model (Equation 4.10), P is the number of free parameters of the model (see Table 4.2), and N is the total number of test trials. As a reference point for model comparison, the AIC and BIC were also computed for a saturated model which had a free parameter for every trial type. Both information criterion were

used because they tend to favor different models, the AIC generally supports models with more parameters and the BIC prefers models with fewer parameters (Burnham & Anderson, 2004).

For each dataset a ΔAIC and ΔBIC score was computed for each model. These values were computed by subtracting the AIC or BIC score of the best model from the score of all models (Burnham & Anderson, 2004). A ΔBIC or ΔAIC score of 0 indicates the best performing model for a particular dataset and larger values indicate increasingly worse model performance. The ΔBIC and ΔAIC scores were normalized by computing AIC weights and BIC weights. AIC and BIC weights were computed following the formulation:

$$wBIC_i = \frac{exp(-0.5 * \Delta BIC_i)}{\sum_{j \in M} exp(-0.5 * \Delta BIC_j)} \tag{4.13}$$

$$wAIC_i = \frac{exp(-0.5 * \Delta AIC_i)}{\sum_{j \in M} exp(-0.5 * \Delta AIC_j)} \tag{4.14}$$

The weighted AIC and BIC values provide the most straightforward interpretation of relative model performance because all BIC and AIC weight scores range from 0 to 1. Larger values indicating better model performance and a model that outperforms all competitor models has an AIC or BIC weight value near 1 and a model that has a single close competitor will have a weight near 0.5.

4.5   Datasets

The data from each participant in Experiment 4 consists of whole-part perceptual discrimination judgments from three whole-part testing sessions: before any training, after the first category structure training (either *Unitization* or *Differentiation* category structures), and after the second category training. Within each test phase, 624 trials were composed of familiar whole and part stimuli that were presented during category training and 208
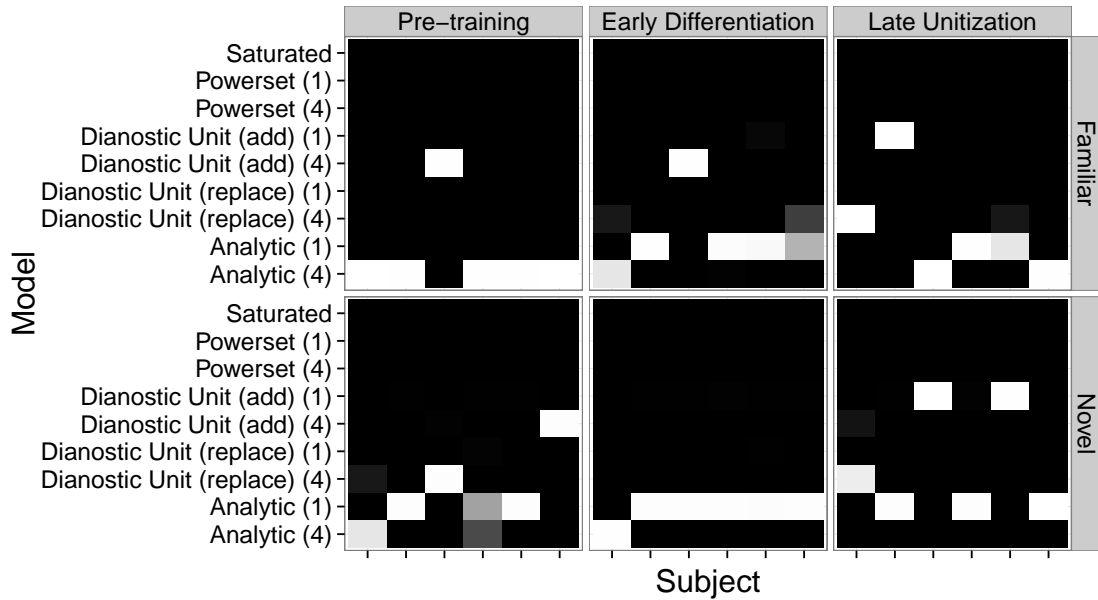
Figure 4.1: The weighted BIC of all models for the participants who learned the *Differentiation* category structures first. Dark cells indicate low weighted BIC values and models that are not preferred according to BIC. Light cells indicate high weighted BIC values. Models followed by (1) indicate fits that have a single the SHARED b and c parameters across all stimuli. The FLEXIBLE b and c parameters are indicated with a (4), indicating there were four levels of each b and c parameter.

trials with stimuli that were novel and had never been seen before. The familiar and novel trials for each of the 12 participants within each testing phase were separated into different datasets, producing 144 datasets. These trials were grouped into 56 trial types that were the unique combinations of segments matching, mismatching, and occluded in each segment position across the all trials. Eight unique models, as outlined in Table 4.2, were fit independently to each dataset. These models were fit via particle swarm optimization to find the maximum likelihood estimate for the model parameters via the methods described earlier.
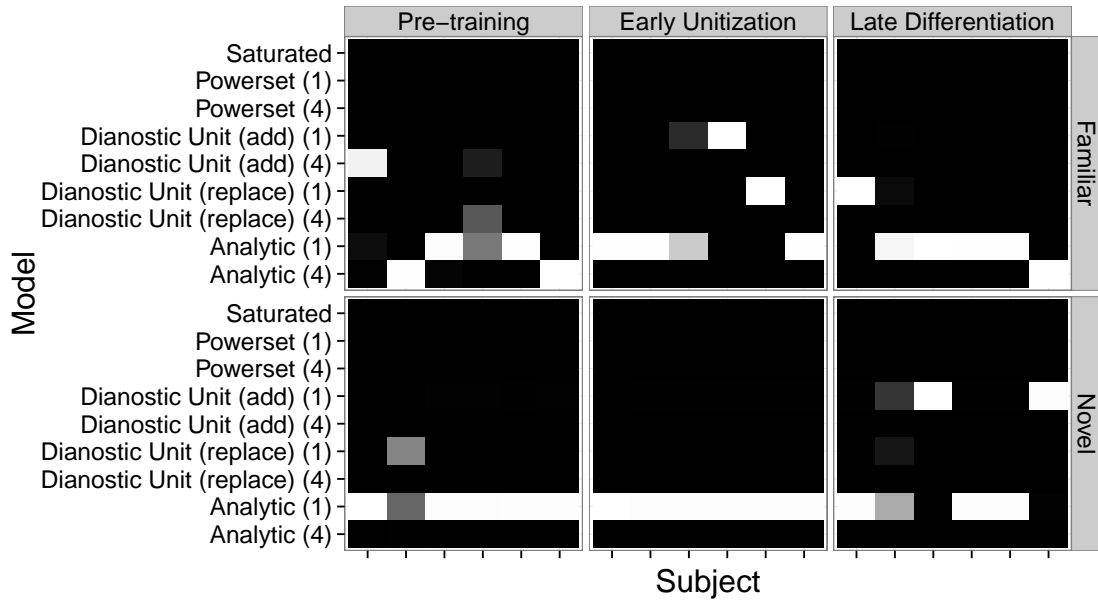
Figure 4.2: The weighted BIC of all models for the participants who learned the *Unitization* category structure first. Dark cells indicate low weighted BIC values and models that are not preferred according to BIC. Light cells indicate high weighted BIC values. Models followed by (1) indicate fits that have a single the SHARED b and c parameters across all stimuli. The FLEXIBLE b and c parameters are indicated with a (4), indicating there were four levels of each b and c parameter.

4.6   Results

4.6.1   BIC weights

The BIC weights for the participants who learned the *Differentiation* category structure first are shown in Figure 4.1. In both familiar (14 of 18) and novel (14 of 18) stimuli, the BIC weights favor the ANALYTIC feature vocabularies for most datasets. This does not seem to be due entirely to the number of free parameters being lower for the ANALYTIC models. Despite the FLEXIBLE ANALYTIC having more free parameters than the FIXED DIAGNOSTIC UNIT models, it is the model that best accounts for the 5 of 6 participants
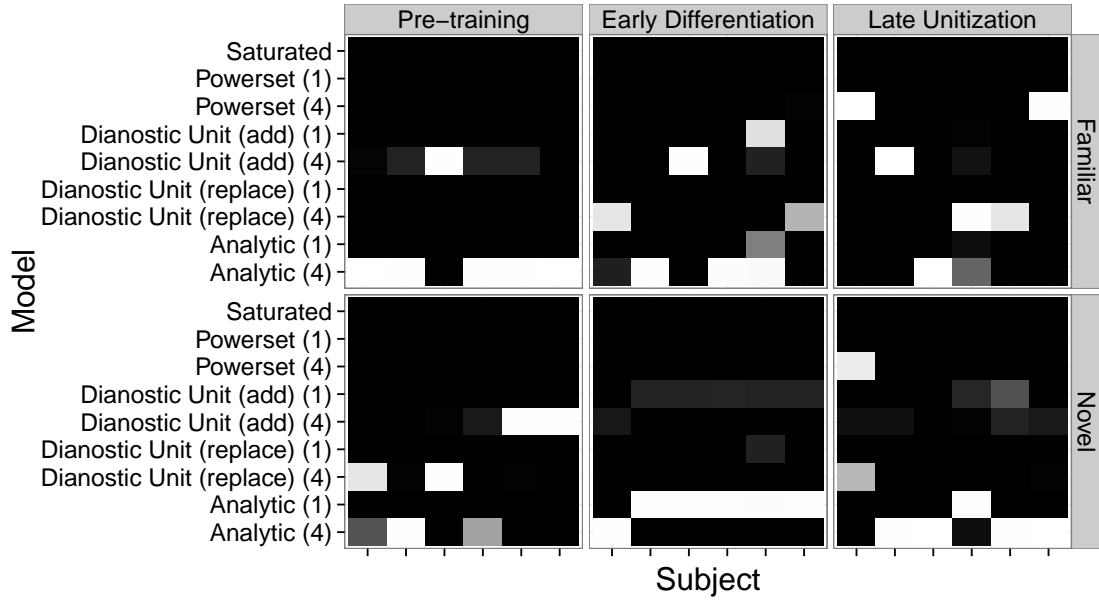
Figure 4.3: The weighted AIC of all models for the participants who learned the *Differentiation* category structures first. Dark cells indicate low weighted AIC values and models that are not preferred according to AIC. Light cells indicate high weighted AIC values. Models followed by (1) indicate fits that have a single the SHARED b and c parameters across all stimuli. The FLEXIBLE b and c parameters are indicated with a (4), indicating there were four levels of each b and c parameter.

before any training.

The BIC weights for the participants who learned the *Unitization* category structure first are shown in Figure 4.2. As in the participants from the other category ordering, the BIC weights favor the ANALYTIC feature vocabularies in 14 of 18 datasets of familiar stimuli and 15 of 18 for novel stimuli. For these participants the model with the fewest parameters, the FIXED ANALYTIC model, is the most frequent high scoring model (26 of 36 datasets).

Across both category training orders, the BIC weights indicate the feature vocabulary that is the best account for judgments of both novel and familiar stimuli is the ANALYTIC feature vocabulary. The UNITIZED CATEGORY RELEVANT feature vocabulary was occasion-
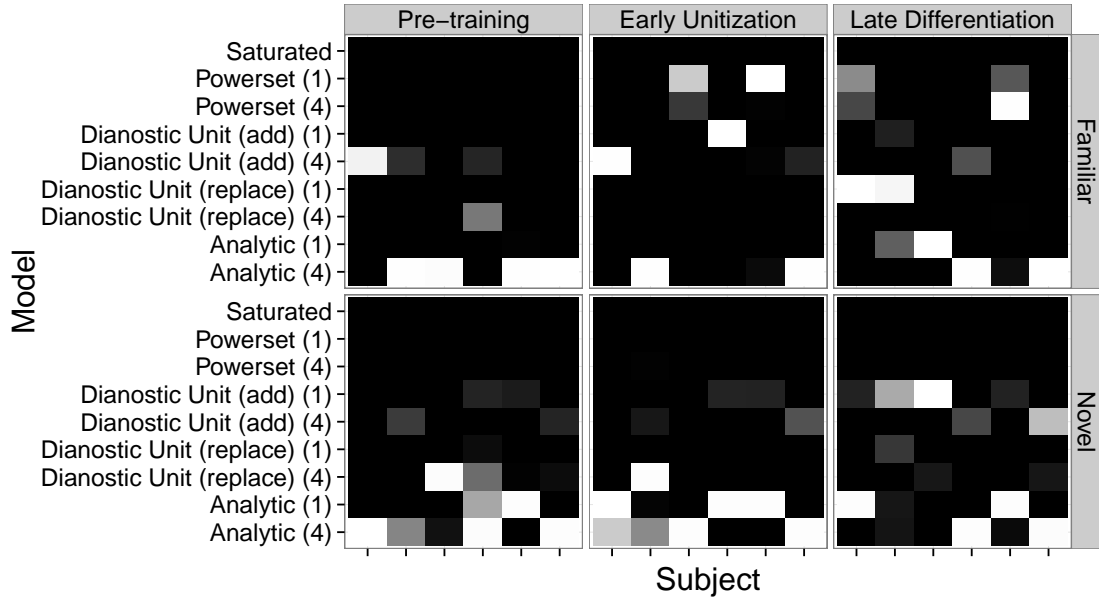
Figure 4.4: The weighted AIC of all models for the participants who learned the *Unitization* category structure first. Dark cells indicate low weighted AIC values and models that are not preferred according to AIC. Light cells indicate high weighted AIC values. Models followed by (1) indicate fits that have a single the SHARED b and c parameters across all stimuli. The FLEXIBLE b and c parameters are indicated with a (4), indicating there were four levels of each b and c parameter.

ally the preferred model for familiar stimuli (2 of 12 before training, 4 of 12 after *Unitization* category training, and 3 of 12 after *Differentiation* category training) but the POWERSET feature vocabulary models and the SATURATED model were never the preferred model. There was no evidence that learning category structures changed the preferred model or that training order had an effect either. The FLEXIBLE response choice models were preferred overall (53 out of 72; $\chi^2(1) = 16.1, p < 0.0005$), suggesting participants were changing their response thresholds based on the number of unoccluded segments in the part stimulus.

4.6.2 AIC weights

The AIC weights suggest a different interpretation of the model comparison than BIC. Figures 4.3 and 4.4 show the AIC weights for the *Differentiation* and *Unitization* category structures respectively. The participants trained on the *Differentiation* category structure first (Figure 4.3) show a progression toward more complex feature vocabularies across training. The proportion of participants whose judgments on the familiar stimuli is best accounted for by one of the ANALYTIC models decreased from 5/6 before training to 3/6 after *Differentiation* training, and to only 1/6 after *Unitization* training. A similar but less pronounced pattern was found for the participants trained on the *Unitization* category structure first. The proportion of participants whose judgments on the familiar stimuli is best accounted for by one of the ANALYTIC models decreased from 4/6 before training to 2/6 after *Unitization* training, but increased to 3/6 after *Differentiation* training. When collapsed across the order of category structure training, there is a marginally significant effect of category structure on the proportion of participants best fit by the models that use the ANALYTIC feature vocabulary ($\chi^2(2) = 6, p = 0.04$).[3]

The systematic shift from the ANALYTIC feature vocabulary to more complex feature vocabularies particularly after *Unitization* training does not appear for novel stimuli. The judgments of novel stimuli are better accounted for by the ANALYTIC feature vocabulary (26 of 36) than the more complex feature vocabulary models (8 of 36).

What can we conclude about which models best account for the data based on the disagreement between BIC and AIC model comparison? First, AIC and BIC agree that the best model of perceptual discriminations for these stimuli before any training is the ANALYTIC feature vocabulary. Once category training occurs, the AIC weights indicate

---

[3]This post-hoc test is only marginally significant, but a stronger result would be hard to observe because the χ^2 test is severely underpowered due to having only 12 observations in each cell.

that the fit of the more complex feature vocabularies increases relative to the ANALYTIC feature vocabulary but this improvement is not sufficient to outweigh the stronger penalty for model complexity imposed by the BIC (Burnham & Anderson, 2004).

4.6.3   Feature weights

Analytic features

Moving beyond which model best accounted for the data, do the best fitting parameters reveal anything about what is being learned? First we examine the feature weights from the ANALYTIC feature set. For the purposes of this analysis, the three features that map onto category relevant segments are grouped together into one average weight on one-segment category relevant features. A repeated-measures ANOVA was performed on the feature weights with three within-subject factors: 2 types of feature (category relevant and category irrelevant), 3 test phases (before any training, after the first, and after the second training phase), and 2 parameterization types (FIXED with one b and c value or FLEXIBLE with four b and c values each). The 2 possible first category structures (*Unitization* or *Differentiation*) were included as a between-subject factor.

There was a main effect of category relevance ($F(1, 10) = 7.34, p = 0.022$) and a marginal effect of training phase ($F(2, 20) = 3.55, p = 0.048$) and a marginally significant interaction between category relevance and training phase ($F(2, 20) = 3.55, p = 0.048$)[4]. These effects are shown in Figure 4.5. The effects of first category structure ($F(1, 10) < 1$) and parameterization ($F(1, 10) < 1$) were not significant as were all other interactions ($F < 1$).

---

[4]The main effect of training phase and the interaction between training phase and category relevance share the same F value because the feature weights for any individual are normalized and thus the category irrelevant feature weight is a function of the category relevant weight. Thus the interaction effect has the exact same amount of variation as the main effect in both the numerator and the denominator and thus the F ratio is identical.

A set of five post-hoc comparisons, using a Bonferroni corrected alpha value of 0.01, were conducted to understand the interaction between category relevance of the feature and phase. In the test phase before any training, there was no significant difference between category relevant (M = 0.25) and category irrelevant feature weights (M = 0.24; paired t-test $t(11) = 0.1, p = 0.87$). In the test phase after the first category structure was learned category relevant feature weights (M = 0.28) were significantly higher than category irrelevant weights (M = 0.15; paired t-test $t(11) = 3.3, p = 0.007$). The same pattern was found after the second training (category relevant M = 0.28; category irrelevant M = 0.16; paired t-test $t(11) = 3.7, p = 0.003$). There was not a significant interaction between the relevance of feature weights and testing after the first or second training phase $(t(11) = 0.3, p = 0.8)$.[5] Furthermore, there was no interaction between the category relevant and category irrelevant feature weights and if the most recently learned category structure was the *Unitization* or *Differentiation* structures (paired t-test $t(11) < 1$).

These results suggest that the ANALYTIC models consistently assigned higher feature weights to category relevant features than the category irrelevant feature. This difference was not found before any training but was found after both *Unitization* and *Differentiation* category training. Learning a second category structure did not continue to increase the weight assigned to the category relevant features.

A similar pattern of feature weights were found when fitting the ANALYTIC models to the whole-part judgments of novel stimuli shown in Figure 4.6. we performed the same five-factor repeated-measures ANOVA that was used for the familiar stimuli on the feature

---

[5]These post-hoc comparisons do not have the same issue with the category relevant and category irrelevant weights being related as the repeated-measures ANOVA did. These paired t-tests can each be re-conceptualized as a one-sample test comparing the weights to the overall average feature weight. Because of this, we feel the interaction between category relevance and training phase is the appropriate way to interpret the effect of category relevance.

weights from the novel stimuli. There was a main effect effect of training phase ($F(2, 20) = 4.6, p = 0.02$) and a significant interaction between category relevance and training phase ($F(2, 20) = 4.6, p = 0.02$). There was no main effect of category relevance ($F(1, 10) = 4.4, p = 0.06$) and all other main effects and interactions were not significant ($p > 0.2$). A set of five post-hoc comparisons, using a Bonferroni corrected alpha value of 0.01, were conducted to understand the interaction between category relevance and test phase. In the test phase before any training, there was no significant difference between category relevant (M = 0.23) and category irrelevant feature weights (M = 0.30; paired t-test $t(11) = 1.25, p = 0.23$). Category relevant feature weights (M = 0.27) were marginally higher than category irrelevant weights (M = 0.17; paired t-test $t(11) = 2.8, p = 0.017$) in the test phase after the first category structure was learned but the effect was not significant after the second training (category relevant M = 0.28; category irrelevant M = 0.16; paired t-test $t(11) = 2.7, p = 0.022$). There was not a significant interaction between the relevance of feature weights and testing after the first or second training phase ($t(11) < 1$). Furthermore, there was no interaction between the category relevant and category irrelevant feature weights and if the most recently learned category structure was the *Unitization* or *Differentiation* structures (paired t-test $t(11) = 1.4, p = 0.17$).

The feature weights for novel stimuli from the ANALYTIC model mirror the weights for the familiar stimuli but the effects do not seem to be as statistically reliable. The increase in weight to the category relevant features after category training, even for the novel stimuli, compliment the findings that improvements in the whole-part perceptual discriminations was not specific to familiar stimuli. The category training seems to have led to an increase in attention to segments in specific positions of novel stimuli structurally similar to the familiar stimuli.

Unitized category relevant features

The UNITIZED CATEGORY RELEVANT feature set consists of five features: the four features from the ANALYTIC feature set as well as a three-part feature that combines the three category relevant segments. Figure 4.7 shows three feature weights for each model: the one-segment category irrelevant feature, the three-segment unitized feature, and the average of the three one-segment category relevant features. A repeated-measures ANOVA was performed on the feature weights with four within-subject factors: 3 types of features (one-segment category relevant, one-segment category irrelevant, three-segment unit), 3 test phases (before any training, after the first, and after the second training phase), 2 feature combination rules (ADD the unitized feature to the one-segment features or RE-PLACE the one-segment features), and 2 parameterization types (FIXED with one b and c value or FLEXIBLE with four b and c values each). The 2 possible first category structures (*Unitization* or *Differentiation*) were included as a between-subject factor. There was a main effect of feature type ($F(2, 20) = 11.4, p < 0.005$) but no significant effect of training phase ($F(2, 20) = 1.6, p = 0.2$), the first type of category structure learned ($F(1, 10) = 1.7, p = 0.2$), the parameterization ($F(1, 10) = 3.6, p = 0.086$), or the feature combination rule ($F(1, 10) < 1$). The interaction between feature type and combination rule was significant ($F(2, 20) = 6.1, p = 0.009$), as was the interaction between feature type and the parameterization ($F(2, 20) = 6.6, p = 0.006$). The interaction between feature type and training phase was marginally significant ($F(4, 40) = 2.5, p = 0.062$) but all other interactions were not significant ($p > 0.3$).

Figure 4.8 shows the interaction between feature type and the number of parameters. The figure suggests the three-segment unit feature changes due to the number of parameters differently than the one-segment features. Figure 4.9 shows the interaction between feature type and the feature combination rule. The figure suggests the interaction is due to the

weight on the three-segment unit feature being higher under the replace rule than the additive rule and the one-segment category irrelevant feature being lower under the replace rule. Figure 4.10 shows the interaction between feature type and the training phase. The figure suggests this interaction is due to the change in weight on the one-segment category irrelevant segment. As in the ANALYTIC model, the weight on the category irrelevant feature decreases after training.

For the novel stimuli (Figure 4.11) we performed the same five-factor repeated-measures ANOVA that was used for the familiar stimuli. We find was a main effect of feature type ($F(2, 20) = 21.8, p < 0.005$) and an interaction between feature type and the number of parameters ($F(2, 20) = 4.4, p = 0.025$) as well as a three-way interaction between feature type, the number of parameters, and the rule for combining features ($F(2, 20) = 4.5, p = 0.025$). This interaction is shown in Figure 4.12 and appears to be driven by the low feature weight for the three-segment unit feature for the models where the unit feature is ADDED to the other features (instead of REPLACING them) and the b and c parameters are SHARED across all judgments.

Overall, the feature weights for the UNITIZED CATEGORY RELEVANT feature vocabulary are consistent. The relationship between the one-segment category relevant and irrelevant features mirrors the ANALYTIC feature weights, category relevant features receive higher weights than irrelevant once any category training occurs. This advantage also appears for novel stimuli, suggesting it is not specific to individual stimuli. The weight assigned to the three-segment feature depends on the flexibility of the decision rules. Models with the FIXED parameterization of b and the ADD combination rule produce the lowest weights for the three-segment feature because an additional matching or mismatching feature with a high weight will drive large differences in accuracy in a task where accuracy only varies approximately 20% across trial types. Models that are more more flexibly parameterized

can accommodate a larger weight on the three-segment feature.

Powerset features

The POWERSET feature set that consists of 13 features: four one-segment features, four two-segment features, four three-segment features, and one feature that includes all four segments. The features are grouped for analysis by the number of segments they contain and if they contain the category irrelevant segment. The one-, two-, and three-segment feature groups each have at least one feature that contains the category irrelevant segment and one that does not contain it, but the four-segment feature always contains the category irrelevant segment. Figure 4.13 shows the best fitting feature weights for each group of features across all training conditions for the familiar stimuli. A repeated-measures ANOVA was conducted on the feature weights with three within-subject factors: 7 types of features (one-segment category relevant, one-segment category irrelevant, two-segment category relevant, two-segment category irrelevant, three-segment category relevant, three-segment category irrelevant, and four-segment), 3 test phases (before any training, after the first, and after the second training phase), and 2 parameterization types (FIXED with one b and c value or FLEXIBLE with four b and c values each). The 2 possible first category structures (*Unitization* or *Differentiation*) were included as a between-subject factor. There was a main effect of feature type ($F(6, 60) = 3.2, p = 0.008$) but no significant effect of training phase ($F(2, 20) = 2.4, p = 0.11$), the first type of category structure learned ($F(1, 10) < 1$), or the parameterization ($F(1, 10) = 2.5, p = 0.15$). The interaction between feature type and training phase was significant ($F(12, 120) = 2.3, p = 0.010$) but all other interactions were not significant ($p > 0.05$). Figure 4.14 shows this interaction but is difficult to interpret. One method of improving this analysis is to consider the number of segments in the feature and the category relevance of all segments as separate factors in the analysis. This requires

excluding the four-segment feature from the analysis to preserve the assumptions of the ANOVA then performing a repeated-measures ANOVA on the feature weights with three within-subject factors: a continuous variable indicating the number of segments (one, two, and three), 2 category relevance levels (relevant and irrelevant), and 3 test phases (before any training, after the first, and after the second training phase). The main effect of number of segments in the feature was significant (one-segment M = 0.087; two-segment M = 0.087; three-segment M = 0.063; $F(1, 11) = 5.7, p = 0.036$), as was the effect of category relevance (relevant M = 0.099; irrelevant M = 0.059; $F(1, 11) = 13.75, p = 0.0035$). There was no significant effect of test phase ($F(1, 22) < 1$) but it did have a significant interaction with category relevance ($F(2, 22) = 4.49, p = 0.023$) and a marginally significant interaction with number of segments ($F(2, 22) = 2.86, p = 0.079$). All other interactions were not significant ($p > 0.3$). These effects are shown in Figure 4.15 where the interaction between category relevance and test phase is due to category relevant features having a higher feature weight after any type of training, regardless of the number of segments in the feature.

The same analysis can be performed for the feature weights from the POWERSET model applied to the novel stimuli, shown in Figure 4.16. However, the larger number of free parameters in these models may be due to noise and not reflect true differences because the complexity of these models were not preferred by either the AIC or BIC model selection criteria. We performed a repeated-measures ANOVA on the feature weights with three within-subject factors: 7 types of features (one-segment category relevant, one-segment category irrelevant, two-segment category relevant, two-segment category irrelevant, three-segment category relevant, three-segment category irrelevant, four-segment), 3 test phases (before any training, after the first, and after the second training phase), and 2 parameterization types (FIXED with one b and c value or FLEXIBLE with four b and c values each). The 2 possible first category structures (*Unitization* or *Differentiation*) were included as a

118

between-subject factor. There was a main effect of feature type ($F(6, 60) = 5.26, p = 0.0002$) and all other effects and interactions were not significant ($p > 0.1$).

As was done for the familiar segments, the effect of feature type for the novel stimuli can be broken down into category relevance and the number of segments in each feature. This requires excluding the four-segment feature from the analysis to preserve the assumptions of the ANOVA then performing a repeated-measures ANOVA on the feature weights with two within-subject factors: a continuous variable indicating the number of segments (one, two, and three) and the 2 category relevance levels (relevant and irrelevant). The main effect of number of segments in the feature was significant (one-segment M = 0.099; two-segment M = 0.079; three-segment M = 0.050; $F(1, 11) = 5.9, p = 0.033$), but the effect of category relevance was not significant (relevant M = 0.088; irrelevant M = 0.064; $F(1, 11) = 4.3, p = 0.062$). There was a a significant interaction between category relevance and number of segments ($F(1, 11) = 7.8, p = 0.017$). The interaction between category relevance and the number of segments is shown in Figure 4.17 and suggests the interaction is due to category relevance having higher feature weights for the one-segment features only.

Perhaps the most interesting effect of feature weights for the POWERSET feature vocabulary is the shift across learning toward assigning more weight to category relevant features that span multiple segments (Figure 4.15). This effect is found for familiar stimuli but not for novel stimuli (Figure 4.16), suggesting a learning process that increasingly puts weight on larger features and is unique to trained stimuli.

4.7   Discussion

What do these models tell us about the changes in the perceptual representations people learn from training and use to make perceptual discriminations? First, behavior in this task is highly idiosyncratic. The perceptual representation that best accounts for an indi-

vidual's discriminations for familiar stimuli might not be the best set of features to explain judgments of novel stimuli or familiar stimuli after the next set of training. However, the weighted BIC and AIC scores of the models do provide some insight. Before any training, all model comparison techniques agree the best set of features to account for the perceptual discriminations is the ANALYTIC feature set in which each feature corresponds to a unique segment in each stimulus. Training seems to change performance such that the models with more perceptual features account for the data better according to the weighted AIC measure, but the BIC continues to favor the ANALYTIC feature vocabulary models due to model complexity.

The best fitting feature weights across all models also provide some insight into what is being learned. Across all datasets category training increases the weight on category relevant features relative to irrelevant features. This occurs regardless of the number of segments in a feature, or if the stimuli are novel or familiar. This suggests an attentional shift toward positions, not just toward specific segments, that are relevant for categorization (Posner, 1980).

The feature weights also shift toward more complex features with additional category training. In both the POWERSET and UNITIZED CATEGORY RELEVANT feature sets, category training led to an increasing proportion of weight being assigned to category relevant features that spanned more than one segment. Unlike the overall bias to shift attention to category relevant positions, the increase in weight to more complex features appears to be specific to judgments of familiar stimuli. There does not appear to be a strong difference between the changes in processing induced by learning *Unitization* and *Differentiation* category structures, the changes due to learning any structure were stronger than any difference between them.
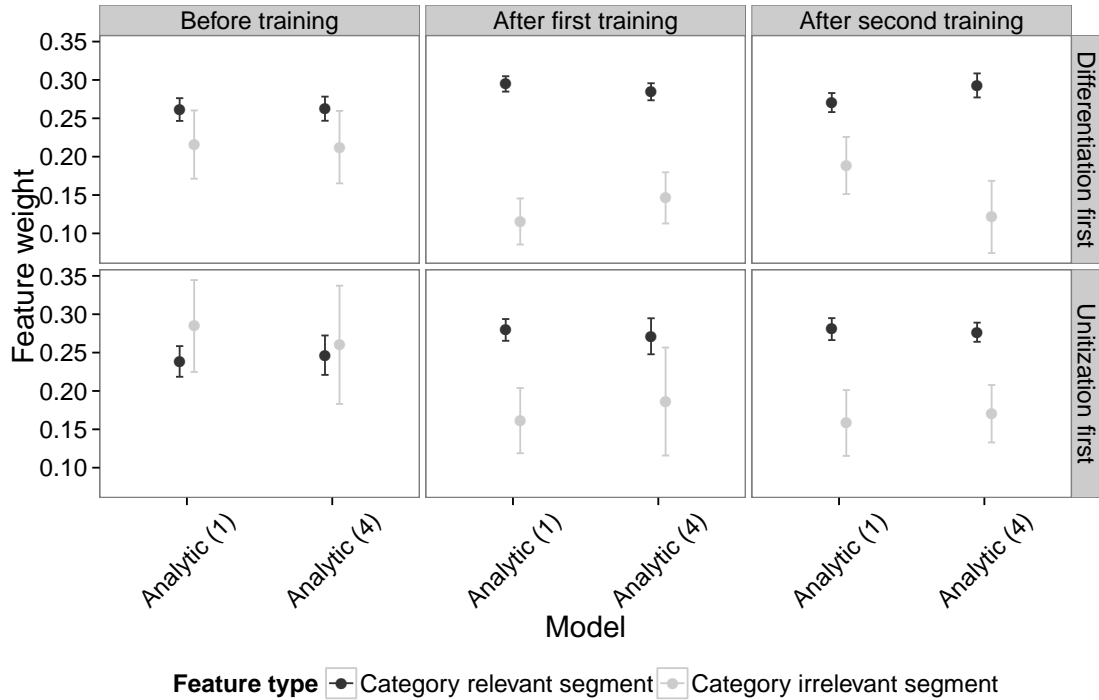
Figure 4.5: The best fitting feature weights for the ANALYTIC feature vocabulary to account for the whole-part discrimination judgments of familiar stimuli. The top row consists of participants who learned the *Differentiation* category structures first and the bottom row learned the *Unitization* category structure first. The "Analytic (1)" points are the weights from the FIXED b and c parameterization and the "Analytic (4)" points are the weights from the FLEXIBLE parameterization. The black points show the average of three category relevant features and the grey points show the weight on the single category irrelevant feature. All features in the ANALYTIC feature vocabulary are composed of exactly one segment. Error bars indicate standard error.
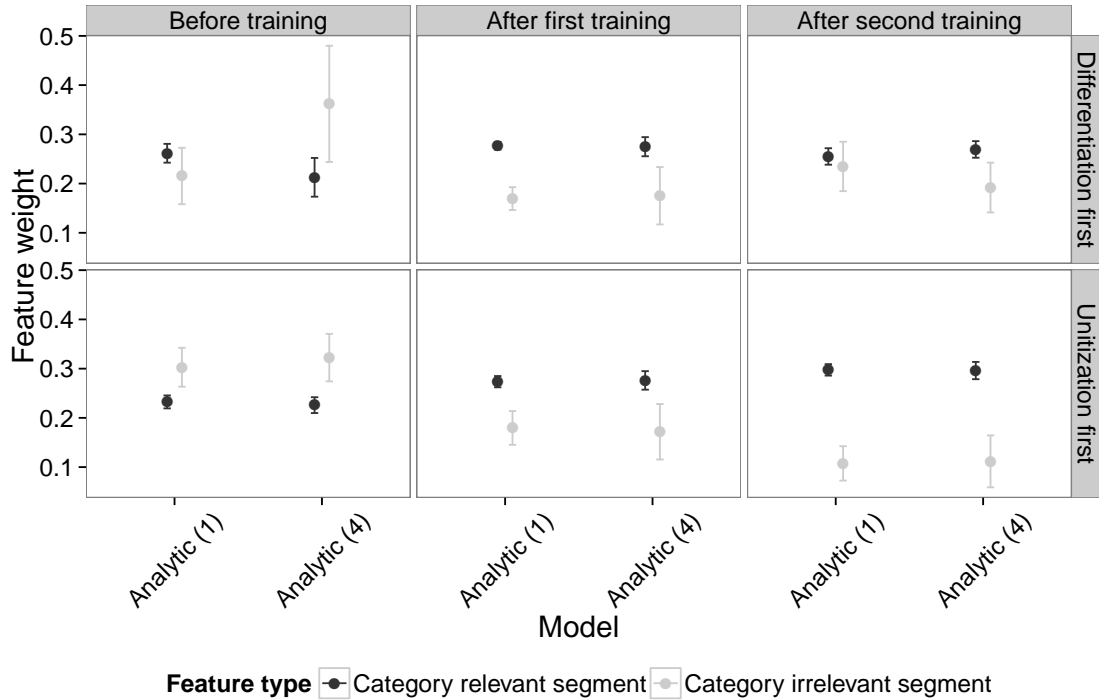
Figure 4.6: The best fitting feature weights for the ANALYTIC feature vocabulary to account for the whole-part discrimination judgments of novel stimuli. The top row consists of participants who learned the *Differentiation* category structures first and the bottom row learned the *Unitization* category structure first. The "Analytic (1)" points are the weights from the FIXED b and c parameterization and the "Analytic (4)" points are the weights from the FLEXIBLE parameterization. The black points show the average of three category relevant features and the grey points show the weight on the single category irrelevant feature. All features in the ANALYTIC feature vocabulary are composed of exactly one segment. Error bars indicate standard error.
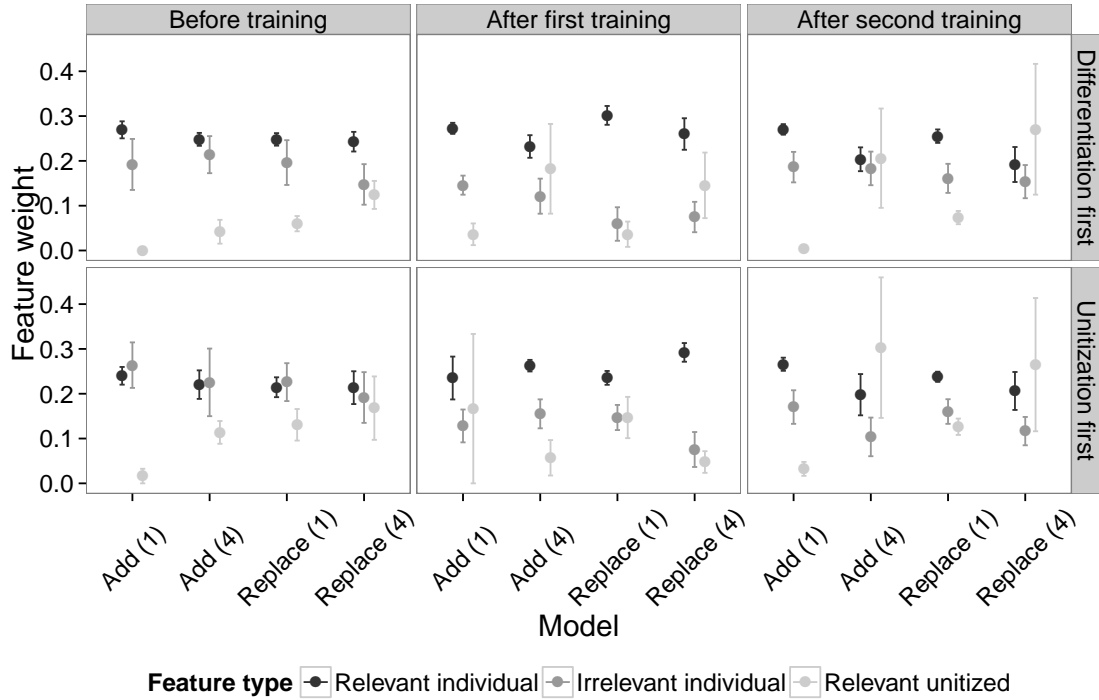
Figure 4.7: The best fitting feature weights for the UNITIZED CATEGORY RELEVANT feature vocabulary to account for the whole-part discrimination judgments of familiar stimuli. The top row consists of participants who learned the *Differentiation* category structures first and the bottom row learned the *Unitization* category structure first. The black points show the average of three one-segment category relevant features, the dark grey points show the weight on the single one-segment category irrelevant feature, and the light grey points indicate the weight assigned to the three-segment feature composed of all three category relevant features. The model name indicates both the combination rule for the three-segment feature (added to the component features or replacing them) as well as if the b and c parameters were FIXED (1) or FLEXIBLE (4). Error bars indicate standard error.
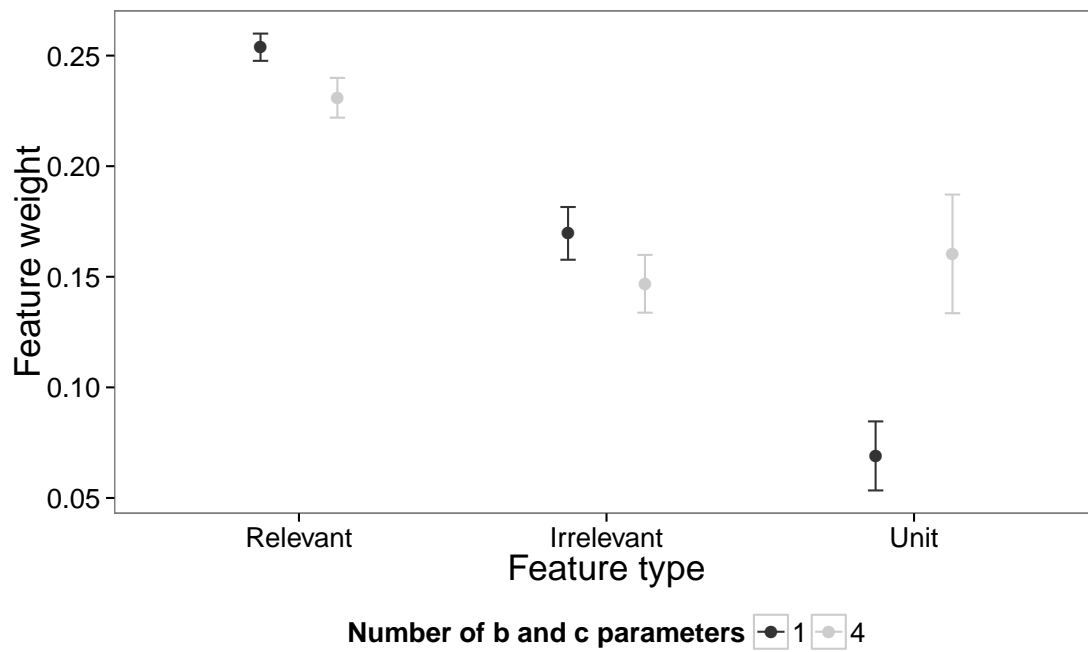
Figure 4.8: The interaction between feature type and parameterization for the UNITIZED CATEGORY RELEVANT feature vocabulary on judgements of familiar stimuli. Error bars indicate standard error.
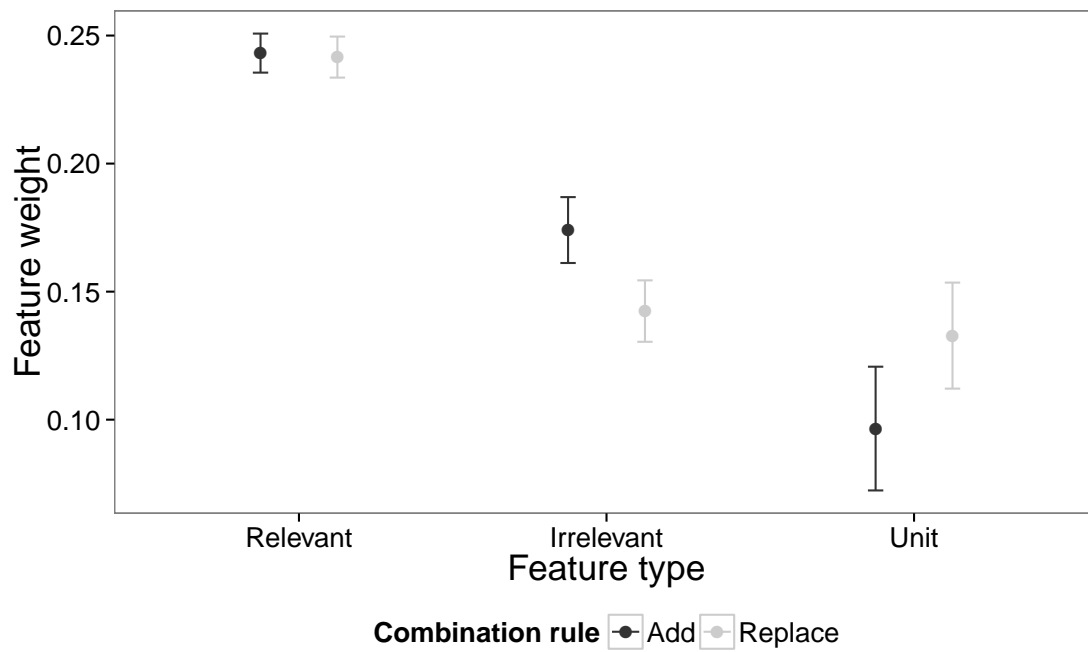
Figure 4.9: The interaction between feature type and feature combination rule for the UNI-TIZED CATEGORY RELEVANT feature vocabulary on judgements of familiar stimuli. Error bars indicate standard error.

Figure 4.10: The interaction between feature type and testing phase for the UNITIZED CATEGORY RELEVANT feature vocabulary on judgements of familiar stimuli. Error bars indicate standard error.
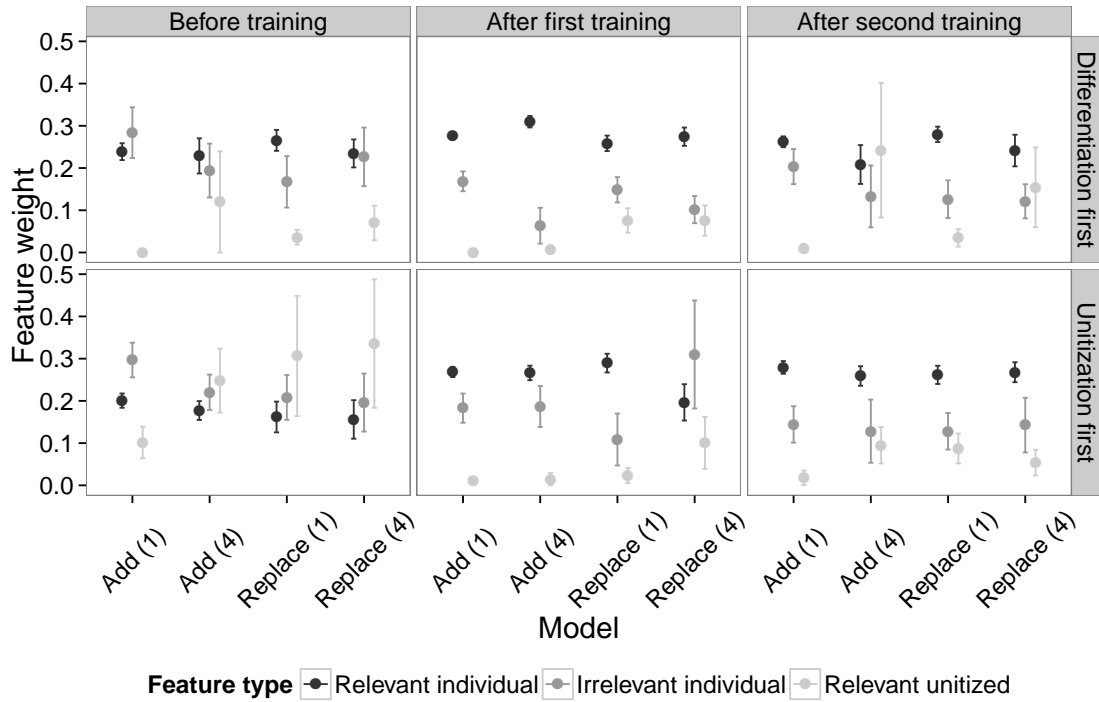
Figure 4.11: The best fitting feature weights for the UNITIZED CATEGORY RELEVANT feature vocabulary to account for the whole-part discrimination judgments of novel stimuli. The top row consists of participants who learned the *Differentiation* category structures first and the bottom row learned the *Unitization* category structure first. The black points show the average of three one-segment category relevant features, the dark grey points show the weight on the single one-segment category irrelevant feature, and the light grey points indicate the weight assigned to the three-segment feature composed of all three category relevant features. The model name indicates both the combination rule for the three-segment feature (added to the component features or replacing them) as well as if the b and c parameters were FIXED (1) or FLEXIBLE (4). Error bars indicate standard error.
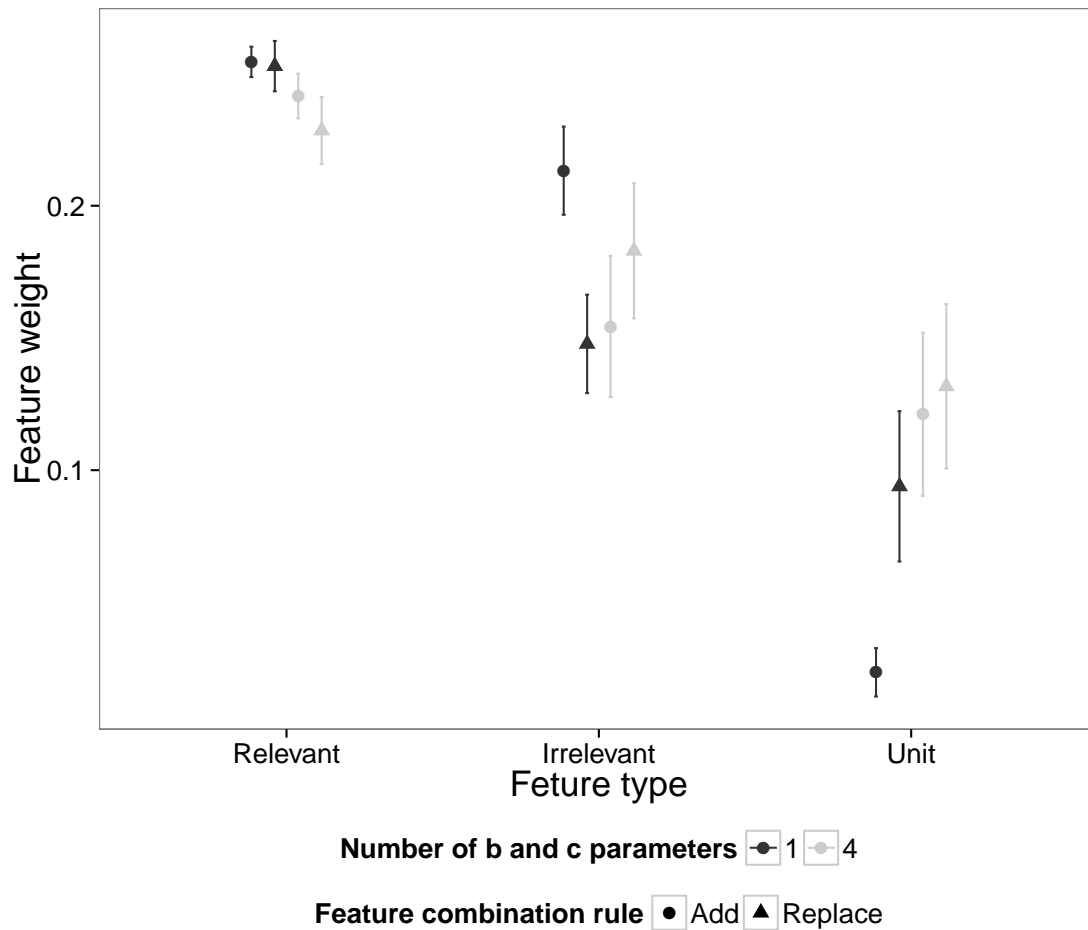
Figure 4.12: The three-way interaction between the feature type, feature combination rule, and decision parameterization for Unitized category relevant feature vocabulary on judgements of novel stimuli. Error bars indicate standard error.
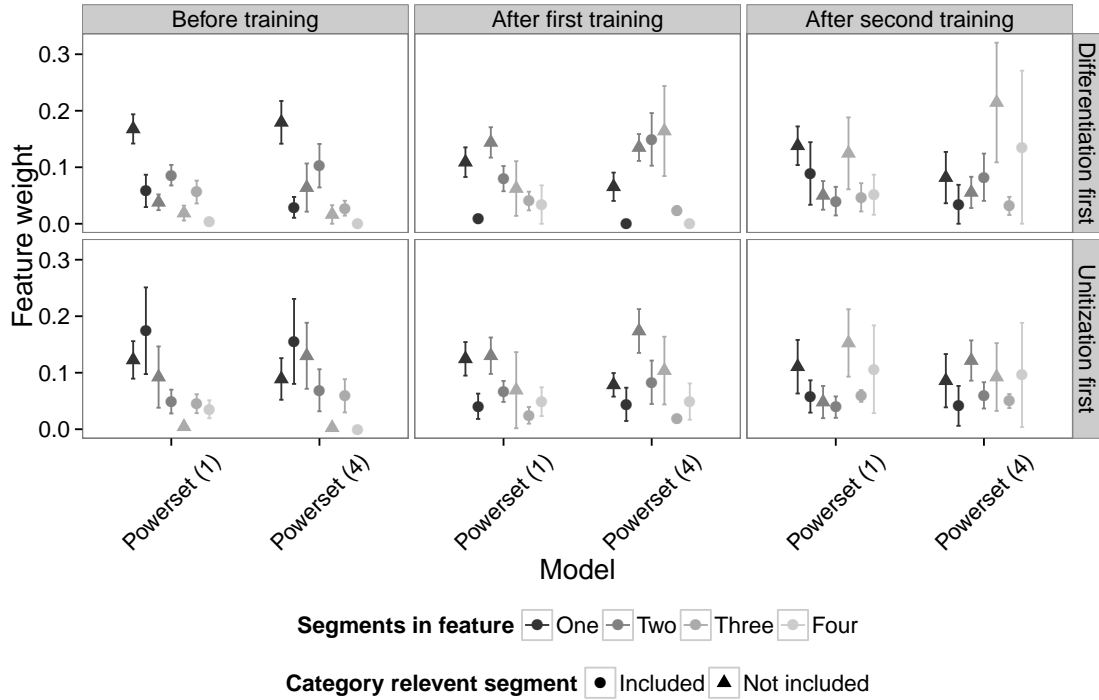
Figure 4.13: The best fitting feature weights for the POWERSET feature vocabulary to account for the whole-part discrimination judgments of familiar stimuli. The top row consists of participants who learned the *Differentiation* category structures first and the bottom row learned the *Unitization* category structure first. The color of the points indicate how many segments were included in the feature (one to four) and the shape of the point indicates if all segments were category relevant (circle) or contained the category irrelevant segment (triangle). The model name indicates if the b and c parameters were FIXED (1) or FLEXIBLE (4). Error bars indicate standard error.

Figure 4.14: The best fitting feature weights for the POWERSET feature vocabulary for judgments of familiar stimuli. Error bars indicate standard error.

Figure 4.15: The best fitting feature weights for the POWERSET feature vocabulary for judgments of familiar stimuli across testing phase and number of segments in the feature. The color indicates if all segments in a feature are category relevant (black) or one is the category irrelevant segment (grey). Error bars indicate standard error.
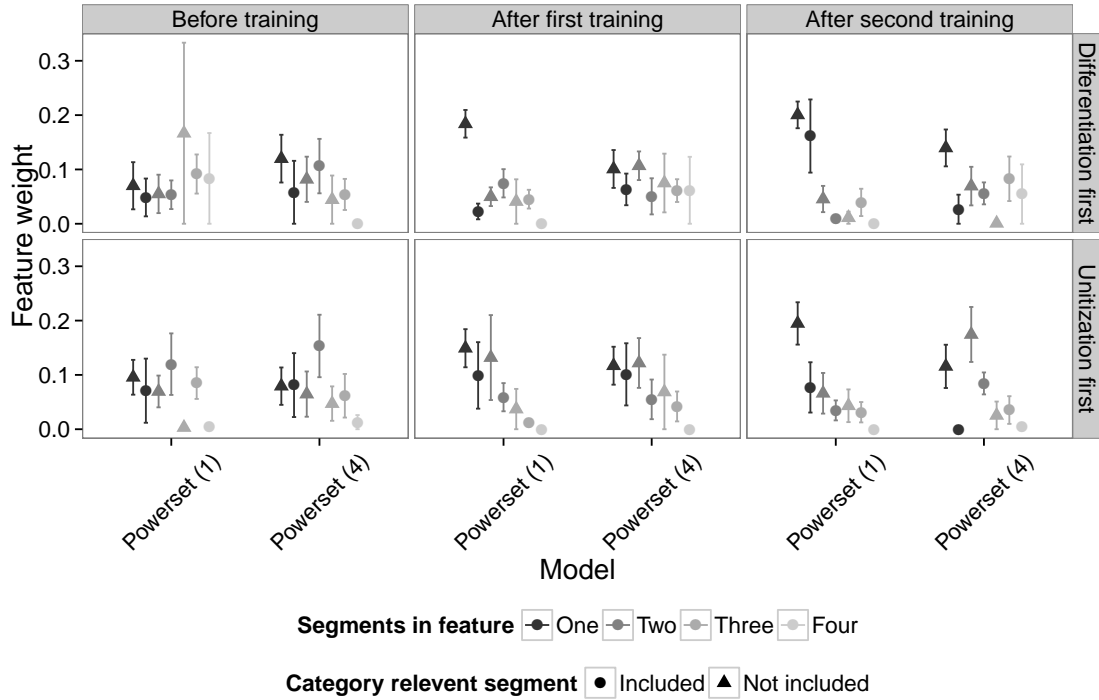
Figure 4.16: The best fitting feature weights for the POWERSET feature vocabulary to account for the whole-part discrimination judgments of novel stimuli. The top row consists of participants who learned the *Differentiation* category structures first and the bottom row learned the *Unitization* category structure first. The color of the points indicate how many segments were included in the feature (one to four) and the shape of the point indicates if all segments were category relevant (circle) or contained the category irrelevant segment (triangle). The model name indicates if the *b* and *c* parameters were FIXED (1) or FLEXIBLE (4). Error bars indicate standard error.
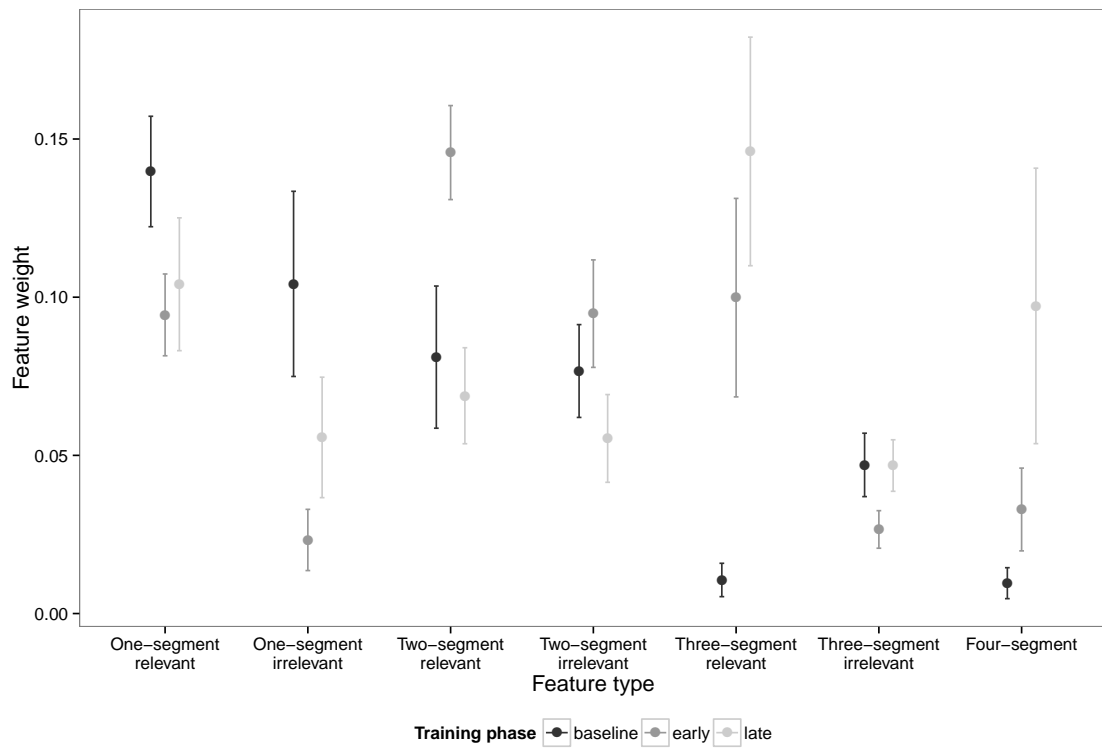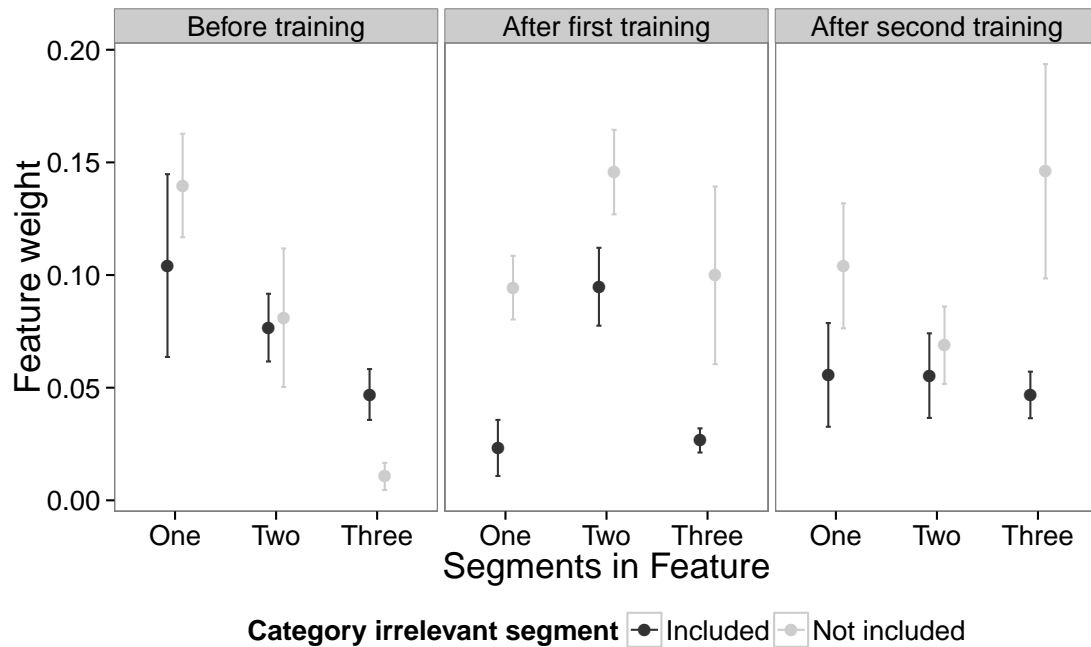
Figure 4.17: The best fitting feature weights for the POWERSET feature vocabulary for judgments of novel stimuli across the number of segments in the feature. The color indicates if all segments in a feature are category relevant (black) or one is the category irrelevant segment (grey). Error bars indicate standard error.
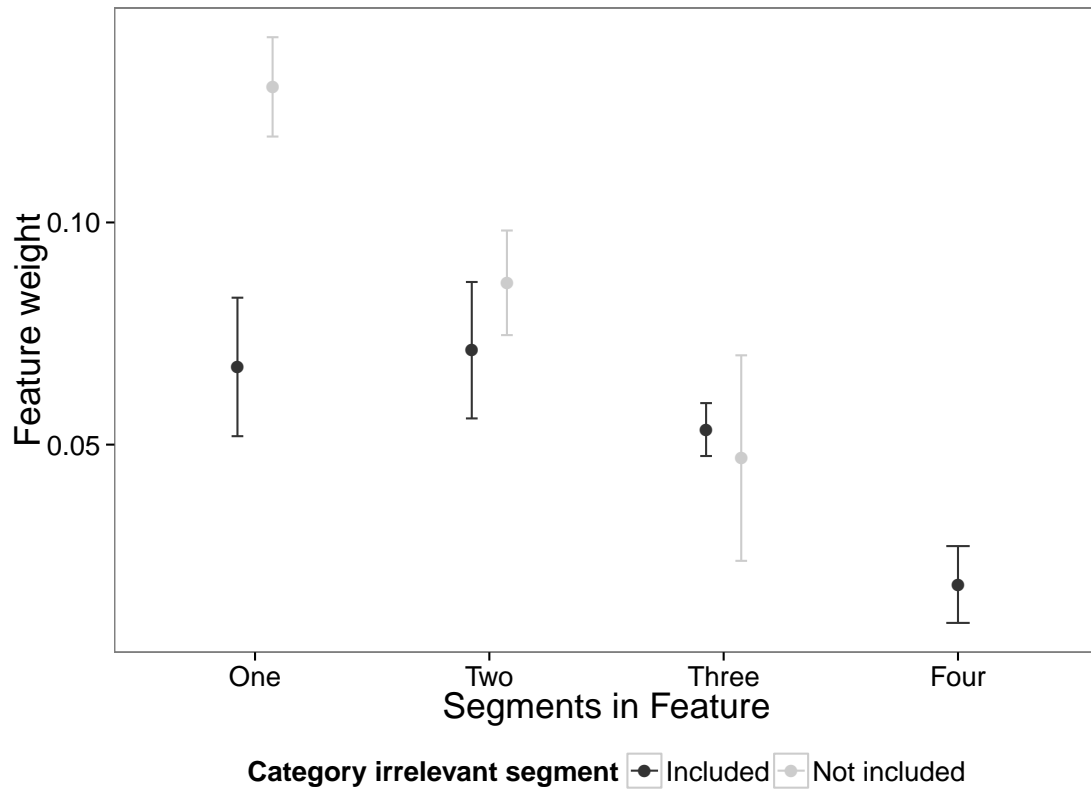
CHAPTER 5

General Discussion

Our perceptual system adjusts based on experience and alters our perceptual representations to increase their utility. This learning can take many forms, and in this dissertation we focus on two processes that drive slow-changing adjustments of the perceptual system: unitization and differentiation. The empirical evidence for both processes comes from a diverse set of stimuli and tasks but evidence for both unitization and differentiation has not been shown within the same stimuli and task design. Despite the effects of unitization and differentiation processes emerging in different learning environments, a number of prominent computational models of perceptual learning treat differentiation and unitization as processes that emerge from the same perceptual learning mechanism. These models claim that the context, specifically the category structure being learned, can determine if perceptual representations are joined together to form unitized features or separated apart to differentiate features from the same set of stimuli.

In the second chapter, we tested the role of category structure for both processes in a set of experiments that manipulate the category structure and the order of category training to induce differentiation and unitization learning from the same set of stimuli. These experiments use stimuli and category structures similar to those in previous studies by Goldstone (2000) that have found evidence for unitization but are complex enough for differentiation processes to emerge. The experimental results do not show strong evidence that learning

a category structure designed to promote unitization has a dramatically different impact than learning a category structure designed to promote differentiation. The most straightforward prediction of the computational models of perceptual learning, that accuracy would be higher for whole-part judgments in which the part stimuli corresponded to category relevant features, did not show any difference between the two category structures in either the one-session or multi-session experiments.

This is not to say that people did not learn the correct category structure or that learning did not influence their perceptual discrimination judgments. Participants learned the correct categorization for both category structures and this learning led to an increase in attention to positions that were relevant for categorization, an improvement in judgments involving the focus category after *Unitization* category structure training, and improved perceptual discrimination of familiar stimuli relative to novel stimuli. Of these three learning effects, the improvement in processing familiar stimuli is the effect most consistent with perceptual learning. The advantage for processing familiar stimuli relative to novel stimuli is consistent with other evidence for the differentiation perceptual learning process that is stimulus specific (Fine & Jacobs, 2002; Goldstone, 1994). Interestingly, this differentiation process occurs for both category structures, suggesting it may not be driven in this task by a particular category structure. The remaining two learning effects are consistent with categorization models that shift attention among features and dimensions (Kruschke, 1992; Nosofsky, 1986) without changing the underlying perceptual representation. Though Goldstone (1998) argues attention shifting among features is a form of perceptual learning by improving the utility of perceptual representations, these effects are not evidence for a change in the set of perceptual features.

The third chapter was devoted to generating a set of candidate feature representations from a computational model of perceptual learning. This model included a mechanism

to infer perceptual features that combined unitization and differentiation into one process that was shaped by category structure. Though this model struggled to infer a set of stable features for the *Unitization* category structure, the model did infer qualitatively different sets of perceptual features based on the category structure. The features the model inferred were consistent with the existing literature that predicted the *Unitization* category structure would include a category relevant feature that was composed of multiple components and the *Differentiation* category structure would produce a set of component features. These sets of perceptual features were used as the basis for the feature representations in computational models fit to the behavioral data.

In the fourth chapter, we developed a computational modeling framework to explicitly predict the behavior using different sets of perceptual features. These models were fit to the perceptual discrimination judgment data from Experiment 4 before and after each type of category training and compared across perceptual features. The best-fitting attention weights across all perceptual features found training increased the weight to features composed of category relevant segments. This increase was found for both familiar and novel stimuli, suggesting this attentional shift, as in the empirical results, was not specific to the familiar stimuli but extended to novel segments in the same positions. The feature representations that contained composite features saw an increase in weight to more complex features after category training and specifically after the *Unitization* category structure training. These changes in weights within a representation are consistent with shifting toward adding composite features.

The comparisons between computational models containing different sets of perceptual features were more ambiguous than the feature weights. The weighted BIC model comparison, which strongly favors less complex models, preferred the most basic perceptual representation for all testing conditions and both novel and familiar stimuli. This model

mapped features to segments in the stimuli and was consistent with static perceptual representations. The weighted AIC model comparison preferred the perceptual representations with more features for judgments of familiar stimuli after learning a category structure. The *Unitization* category structure increased the proportion of people who were best accounted for by a perceptual representation with composite feature regardless of when it occurred in training, but the *Differentiation* category structure only did so when it was the first category structure being learned. The same increase in composite perceptual features of the best fitting model did not occur for novel stimuli. It is difficult to interpret the disconnect between the AIC and BIC model comparisons beyond concluding that perceptual representations might be shifting but the evidence does not overwhelmingly favor the models that require additional complexity of more features. There appears to be strong individual differences in what perceptual features best account for whole-part discrimination performance after each type category training for these stimuli.

5.1 Future directions

The work presented here does not clearly identify if people were learning to differentiate and unitize new features in these tasks. However, this work does provide a framework with many advantages for future work to better understand the mechanisms of perceptual learning and the influence of category information on them. First, perceptual discrimination judgments address the recent controversy surrounding task demands accounting for many top-down perceptual effects (Firestone & Scholl, 2015) by being a simple speeded judgment task that has no clear demands due to training. Whole-part judgments are not dependent on the category structure because participants are explicitly instructed that all of the unoccluded part stimulus must match. Regardless of the category people most recently learned, on each whole-part trial they are judging if the part stimulus matches the whole without re-

gard to category membership. Second, perceptual discrimination judgments can potentially separate the learning effects of perceptual processes from learning labels and language in category tasks (Roberson & Davidoff, 2000). By manipulating what segments are present in the part stimulus, it is possible to have trials in which the part stimulus can be categorized and assigned a label and others where as many segments are present but the category is ambiguous. This allows for manipulations where the number of matching segments can be constant across trials for perceptual judgments but the category label information changes between trials. This manipulation was not done in the current work but is possible in this framework. Also, a critical set of controls that constrained the interpretation of the empirical data were the perceptual judgments done before any training and the novel stimuli included in all testing phases. Without a pre-training assessment, the general advantage in discrimination of the participants who received the *Unitization* category training first would have been indistinguishable from a benefit of learning that structure first. This straightforward methodological control is often used in psychometric studies of perceptual learning (Fine & Jacobs, 2002) but less often for category-induced perceptual learning (Pevtzow & Goldstone, 1994). Finally, the relevance of the empirical results were assessed via model comparison with models that include a variety of sets of perceptual features, some of which were predicted by computational models of perceptual learning. Quantitative model comparison is common for comparing representations in category learning paradigms (Nosofsky & Johansen, 2000), but has not previously been applied to perceptual learning results. Thus far, computational modeling in perceptual learning has focused on applying a single model and demonstrating it can account for a given phenomena (Austerweil & Griffiths, 2013; Goldstone, 2003; Dosher et al., 2013; Orbán et al., 2008).

Whole-part discrimination tasks are particularly appropriate for computational modeling in future work on understanding perceptual learning because they are a form of simple

decision tasks that have been extensively modeled with computational models that predict not only choice, but reaction time (Ratcliff, 1978; Nosofsky & Palmeri, 1997; Brown & Heathcote, 2008). One limitation with the current study was the lack of a clear signal in accuracy that disambiguated between the models of perceptual features. Measuring and predicting reaction time distributions can add constraints that may disambiguate between models (Goldstone, 2000). Models that predict both choices and reaction times would add additional power to disambiguation.

Focusing exclusively on whole-part discrimination trials was not the only limitation of the experimental design. Most experiments that show evidence of perceptual unitization have a single unitized feature per feature. This is true for visual search tasks (Shiffrin & Lightfoot, 1997), category learning tasks (Goldstone, 2000; Pevtzow & Goldstone, 1994), and even in fast-learning experiments where the imprinting process leads to learning composite features (Schyns & Rodet, 1997; Schyns & Murphy, 1994). The exception to this are experiments without categories where clusters of stimuli can be formed with each having their own feature (Fiser & Aslin, 2001, 2002; Austerweil & Griffiths, 2011). Yet in the experimental design in the second chapter, the unitization category structure had two equally frequent composite features in the focus category: the ABC and WXY features. Categories that are defined by more than one feature are more difficult to learn in a regular categorization task (Shepard et al., 1961) and having a many-to-one mapping between features and category labels may have harmed perceptual learning by confusing the processes. Furthermore, the set of whole-part discrimination trials tested were not optimized for comparing the computational models of perceptual discrimination, many irrelevant judgments were included. This was done to control for the frequency of all segments across whole-part judgments, but it left many discrimination trial types did not disambiguate between different sets of perceptual features. These extra trials increased the number of observations

in the penalty for model complexity in the weighted BIC comparisons and impacted model comparisons but did not differentiate between models.

5.2 Conclusion

The literature on perceptual learning identified two learning processes, unitization and differentiation, that characterize seemingly opposite changes in perceptual representation. Yet most prominent computational models of perceptual learning hypothesize both processes are the result of a unified perceptual learning mechanism. This predicts that the task and context determine if perceptual learning will result in forming unitized or differentiated perceptual features, not inherent properties of the stimuli themselves. We tested this prediction in a series of experiments that manipulated the category structure across a set of stimuli and the order categories were learned. Using these stimuli and category structures, we inferred a set of perceptual features for each category structure from an existing computational model of perceptual learning. Finally, we compared the fit of a model assuming each of those sets of perceptual features using a novel model comparison framework. The results of the empirical and modeling work do not show strong evidence that different sets of perceptual features were learned in the two category structures, though some evidence of perceptual learning was found due to both category structures. The most reliable difference due to learning category structures was a shift in attention both to stimulus components and also to whole stimuli. We conclude by discussing the advantages and limitations of this framework for studying perceptual learning and how it could be adapted in the future to better understand perceptual learning in the context of categorization.

References

Adini, Y., Sagi, D., & Tsodyks, M. (2002). Context-enabled learning in the human visual system. *Nature*, *415*(6873), 790–793.

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision research*, *36*(21), 3487–3500.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723.

Austerweil, J. L., & Griffiths, T. L. (2011). A rational model of the effects of distributional information on feature learning. *Cognitive psychology*, *63*(4), 173–209.

Austerweil, J. L., & Griffiths, T. L. (2013). A nonparametric bayesian framework for constructing flexible feature representations. *Psychological review*, *120*(4), 817.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, *218*(4573), 697–698.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision research*, *27*(6), 953–965.

Barton, J. J., Keenan, J. P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. *British journal of psychology*, *92*(3), 527–549.

Beard, B. L., Levi, D. M., & Reich, L. N. (1995). Perceptual learning in parafoveal vision.

Vision research, *35*(12), 1679–1690.

Bennefl, R. G., & Westheimer, G. (1991). The effect of training on visual alignment discrimination and grating resolution. *Perception & Psychophysics*, *49*(6), 541–546.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, *8*(6), 551–565.

Blanes-Vidal, V., Hansen, M. N., Adamsen, A. P. S., Feilberg, A., Petersen, S. O., & Jensen, B. B. (2009). Characterization of odor released during handling of swine slurry: Part i. relationship between odorants and perceived odor concentrations. *Atmospheric Environment*, *43*(18), 2997–3005.

Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological research*, *46*(3), 207–222.

Bradshaw, J. L., & Wallacei, G. (1971). Models for the processing and identification of faces. *Perception & Psychophysics*, *9*(5), 443–448.

Braunstein, M. L., Hoffman, D. D., & Saidpour, A. (1989). Parts of visual objects: An experimental test of the minima rule. *Perception*, *18*(6), 817–826.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153–178.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in cognitive sciences*, *10*(4), 159–166.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding aic and bic in model selection. *Sociological methods & research*, *33*(2), 261–304.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision research*, *45*(4), 431–448.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology*, *16*(4), 257.

Cohen, A. L., & Nosofsky, R. M. (2000). An exemplar-retrieval model of speeded same–different judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1549.

Curtis, D. W., Stevens, D. A., & Lawless, H. T. (1984). Perceived intensity of the taste of sugar mixtures and acid mixtures. *Chemical Senses*, *9*(2), 107–120.

Czerwinski, M., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *The American journal of psychology*, 271–315.

De Valois, K. K. (1977). Spatial frequency adaptation can enhance contrast sensitivity. *Vision research*, *17*(9), 1057–1065.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107.

Dosher, B. A., Jeter, P., Liu, J., & Lu, Z.-L. (2013). An integrated reweighting theory of perceptual learning. *Proceedings of the National Academy of Sciences*, *110*(33), 13678–13683.

Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences*, *95*(23), 13988–13993.

Doshi-Velez, F., & Ghahramani, Z. (2009). Accelerated sampling for the indian buffet process. In *Proceedings of the 26th annual international conference on machine learning* (pp. 273–280).

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Fahle, M. (1994). Human pattern recognition: parallel processing and perceptual learning.

PERCEPTION-LONDON-, *23*, 411–411.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision research*, *33*(3), 397–412.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is" special" about face perception? *Psychological review*, *105*(3), 482.

Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision research*, *23*(2), 145–150.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of vision*, *2*(2), 5.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency.

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision research*, *21*(7), 1149–1158.

Firestone, C., & Scholl, B. J. (2014). "top-down" effects where none should be found the el greco fallacy in perception research. *Psychological science*, *25*(1), 38–46.

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for 'top-down'effects. *The Behavioral and brain sciences*, 1–77.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological science*, *12*(6), 499–504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *PERCEPTION-LONDON-*, *29*(2), 159–170.

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: Object

specificity and size invariance. *Vision research*, *40*(5), 473–484.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature neuroscience*, *6*(4), 428–432.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience*, *3*(2), 191–197.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision research*, *37*(12), 1673–1682.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 431.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. *Nature neuroscience*, *2*(6), 568–573.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble'experts: a framework for studying expert object recognition processes. *Vision research*, *38*(15), 2401–2428.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 721–741.

Gerganov, A., Grinberg, M., Quinn, P. C., & Goldstone, R. L. (2007). Simulating conceptually-guided perceptual learning. In *Proceedings of the twenty-ninth annual conference of the cognitive science society* (pp. 287–292).

Ghahramani, Z., & Griffiths, T. L. (2005). Infinite latent feature models and the indian

buffet process. In *Advances in neural information processing systems* (pp. 475–482).

Ghose, G. M., Yang, T., & Maunsell, J. H. (2002). Physiological correlates of perceptual learning in monkey v1 and v2. *Journal of neurophysiology*, *87*(4), 1867–1888.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* Appleton-Century-Crofts.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological review*, *62*(1), 32.

Girard, M. H., Collet, L., Bouchet, P., & Pernier, J. (1994). Auditory selective attention in the human cochlea. *Brain research*, *633*(1), 353–356.

Goldstone, R. L. (1994, Sep). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 1–23.

Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, *49*(1), 585–612.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 86.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. *Perceptual organization in vision: Behavioral and neural perspectives*, 233–278.

Goldstone, R. L., Gerganov, A., Landy, D., & Roberts, M. E. (2009). Learning to see and conceive. *The new cognitive sciences*, 163–188.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). Novel approach to nonlinear/non-

gaussian bayesian state estimation. In *Iee proceedings f (radar and signal processing)* (Vol. 140, pp. 107–113).

Green, B. G., Dalton, P., Cowart, B., Shaffer, G., Rankin, K., & Higgins, J. (1996). Evaluating the 'labeled magnitude scale' for measuring sensations of taste and smell. *Chemical senses*, *21*(3), 323–334.

Griffiths, T. L., & Ghahramani, Z. (2011). The indian buffet process: An introduction and review. *The Journal of Machine Learning Research*, *12*, 1185–1224.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*(1), 1–32.

Hanley, J. R., & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin & Review*, *18*(2), 355–363.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. *Categorical perception: The groundwork of cognition*, 1–52.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach.* John Wiley & Sons.

Helmholtz, H. v. (1910/1924). *Treatise on physiological optics, third edition.* Optical society of America.

Kami, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*(6443), 250–252.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*(11), 4966–4970.

Knapp, B. R., Nosofsky, R. M., & Busey, T. A. (2006). Recognizing distinctive faces: A hybrid-similarity exemplar model account. *Memory & cognition*, *34*(4), 877–889.

Kruschke, J. K. (1992). Alcove: an exemplar-based connectionist model of category learning. *Psychological review*, *99*(1), 22.

Kumar, T., & Glaser, D. (1993). Initial performance, learning and observer variability for hyperacuity tasks. *Vision Research*, *33*(16), 2287–2300.

Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants.* John Wiley & Sons.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). Vot discrimination by four to six and a half month old infants from spanish environments. *Journal of Experimental Child Psychology*, *20*(2), 215–225.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, *54*(5), 358.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of experimental psychology*, *61*(5), 379.

Lin, E. L., & Murphy, G. L. (1997). Effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 1153.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain: a network model of category learning. *Psychological review*, *111*(2), 309.

Luce, R. (1963). *Detection and recognition* (E. G. RD Luce RR Bush, Ed.). New York:

Wiley.

Mayer, M. J. (1983). Practice improves adults' sensitivity to diagonals. *Vision research*, *23*(5), 547–550.

McClelland, J. L., Rumelhart, D. E., & Group, P. R. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1 & 2). MIT Press.

McKee, S. P., & Westhe, G. (1978). Improvement in vernier acuity with practice. *Perception & psychophysics*, *24*(3), 258–262.

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

Metherate, R., & Weinberger, N. M. (1990). Cholinergic modulation of responses to single tones produces tone-specific receptive field alterations in cat auditory cortex. *Synapse*, *6*(2), 133–145.

Minsky, M., & Seymour, P. (1969). Perceptrons.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of japanese and english. *Perception & Psychophysics*, *18*(5), 331–340.

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial vision*, *10*(1), 51–58.

Nelson, J. I. (1977). The plasticity of correspondence: after-effects, illusions and horopter shifts in depth perception. *Journal of theoretical biology*, *66*(2), 203–266.

Newell, F. N., & Bülthoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, *85*(2), 113–143.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, memory, and cognition*, *10*(1), 104.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization rela-

tionship. *Journal of Experimental Psychology: General*, *115*(1), 39.

Nosofsky, R. M. (1988a). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: learning, memory, and cognition*, *14*(4), 700.

Nosofsky, R. M. (1988b). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of experimental psychology: human perception and performance*, *17*(1), 3.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological review*, *104*(2), 266.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2006). Bayesian model learning in human visual perception. *Advances in neural information processing systems*, *18*, 1043.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745–2750.

Overbosch, P. (1986). A theoretical model for perceived intensity in human taste and smell as a function of time. *Chemical senses*, *11*(3), 315–329.

Özgen, E. (2004). Language, learning, and color perception. *Current Directions in Psychological Science*, *13*(3), 95–98.

Özgen, E., & Davies, I. R. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental*

Psychology: General, *131*(4), 477.

Petrov, A. A., Dosher, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological review*, *112*(4), 715.

Petrov, A. A., Dosher, B. A., & Lu, Z.-L. (2006). Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision research*, *46*(19), 3177–3197.

Pevtzow, R., & Goldstone, R. L. (1994, Jan). Categorization and the parsing of objects. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 717–722.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3–25.

Puel, J.-L., Bonfils, P., & Pujol, R. (1988). Selective attention modifies the active micromechanical properties of the cochlea. *Brain research*, *447*(2), 380–383.

Quinn, P. C., & Bhatt, R. S. (2005). Good continuation affects discrimination of visual pattern information in young infants. *Perception & psychophysics*, *67*(7), 1171–1176.

Quinn, P. C., Burke, S., & Rush, A. (1993). Part—whole perception in early infancy: Evidence for perceptual grouping produced by lightness similarity. *Infant Behavior and Development*, *16*(1), 19–42.

Quinn, P. C., & Schyns, P. G. (2003). What goes up may come down: Perceptual process and knowledge access in the organization of complex visual patterns by young infants. *Cognitive Science*, *27*(6), 923–935.

Quinn, P. C., Schyns, P. G., & Goldstone, R. L. (2006). The interplay between perceptual organization and categorization in the representation of complex visual patterns by

young infants. *Journal of Experimental Child Psychology*, *95*(2), 117–127.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Sam: A theory of probabilistic search of associative memory. *The psychology of learning and motivation: Advances in research and theory*, *14*, 207–262.

Ramachandran, V., & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception*.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Rips, L. J. (1989). Similarity, typicality, and categorization. *Similarity and analogical reasoning*, 21–59.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*(6), 977–986.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, *13*(3), 250–257.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, *5*, 3.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive science*, *9*(1), 75–112.

Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: what is learned? *Vision research*, *35*(4), 519–527.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological review*, *117*(4), 1144.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in v1 neurons. *Nature*, *412*(6846), 549–553.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Schyns, P. G., & Murphy, G. (1994). The ontogeny of part representation in object concepts. *Psychology of learning and motivation*, *31*, 305-349.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 681.

Sekuler, R., & Machamer, J. (1983). Detection and identification of moving targets. *Vision research*, *23*(3), 229–238.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of mathematical psychology*, *1*(1), 54–87.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1.

Shiffrin, R. M., & Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological review*, *84*(2), 127.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: Rem—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*(3), 278.

Shiu, L.-P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally

local but dependent on cognitive set. *Perception & psychophysics*, *52*(5), 582–588.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological science*, *12*(1), 43–47.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*(2), 225–245.

Teich, A. F., & Qian, N. (2003). Learning and adaptation in a recurrent model of v1 orientation selectivity. *Journal of Neurophysiology*, *89*(4), 2086–2100.

Tversky, A. (1977). Features of similarity. *Psychological review*, *84*(4), 327.

Vogels, R., & Orban, G. A. (1985). The effect of practice on the oblique effect in line orientation judgments. *Vision research*, *25*(11), 1679–1687.

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*(6858), 844–848.

Weinberger, N. M., Javid, R., & Lepan, B. (1993). Long-term retention of learning-induced receptive-field plasticity in the auditory cortex. *Proceedings of the National Academy of Sciences*, *90*(6), 2394–2398.

Weinstein, S. (1968). Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality. In D. R. Kenshalo (Ed.), *The skin senses* (pp. 195 – 222). Springfield, Il: Thomas.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental psychology*, *24*(5), 672.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49–63.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007).

Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.

Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of vision*, *13*(7), 1.

Witzel, C., & Gegenfurtner, K. R. (2015). Categorical facilitation with equally discriminable colors. *Journal of vision*, *15*(8), 22–22.

Wood, F., Griffiths, T. L., & Ghahramani, Z. (2006). A non-parametric Bayesian method for inferring hidden causes. In *Proceedings of the 22nd conference on uncertainty in artificial intelligence* (pp. 536–543).

Wright, W. D. (1947). Researches on normal and defective colour vision.

Xiao, L.-Q., Zhang, J.-Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology*, *18*(24), 1922–1926.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*(1), 141.

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 1022.

Zhang, J.-Y., Zhang, G.-L., Xiao, L.-Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *The Journal of Neuroscience*, *30*(37), 12323–12328.

Zhang, T., Xiao, L.-Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Decoupling location specificity from perceptual learning of orientation discrimination. *Vision research*, *50*(4), 368–374.

Zhaoping, L., Herzog, M. H., & Dayan, P. (2003). Nonlinear ideal observation and recur-

rent preprocessing in perceptual learning. *Network: Computation in Neural Systems*, *14*(2), 233–247.

Andrew Hendrickson

Contact Information

School of Psychology

University of Adelaide

North Terrace Campus, Hughes Building, Level 4

South Australia, 5005

Australia

drew.hendrickson@adelaide.edu.au

Education

Indiana University, Bloomington, Indiana, USA

Ph.D., November 2015

Psychological and Brain Sciences and Cognitive Science Program,

Advisor: Dr. Robert Goldstone

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

B.S., May 2005

Computer Science and Cognitive Science

Publications

1. Hendrickson, A.T., Navarro, D.J. and Perfors, A. (in press). Sensitivity to hypothesis size during information search Decision

2. Hendrickson, A.T., Navarro, D.J. and Donkin, C. (2015). Quantifying the time course of similarity. In DC Noelle, R Dale, AS Warlaumont, J Yoshimi, T Matlock, CD Jennings and PP Maglio (Ed.) Proceedings of the 37th Annual Conference of the Cognitive Science Society (pp. 908-913)

3. Hendrickson, A. T., Perfors, A., Navarro, D. J. (2014). Adaptive information source selection during hypothesis testing. In P Bellow, M Guarani, M McShane and B Scassellati (Ed.) Proceedings of the 36th Annual Conference of the Cognitive Science Society

4. Langsford, S., Hendrickson, A. T., Perfors, A., Navarro, D. J. (2014). People are sensitive to hypothesis sparsity during category discrimination. In P Bellow, M Guarani, M McShane and B Scassellati (Ed.) Proceedings of the 36th Annual Conference of the Cognitive Science Society

5. Todd, P. M., Hills, T. T., Hendrickson, A. T. (2013). Modeling reproductive decisions with simple heuristics. For Demographic Research, Special Collection on "Theoretical Foundations of the Analysis of Fertility."

6. Vong, W.K., Hendrickson, A. T., Perfors, A., Navarro, D. J. (2013). The role of sampling assumptions in generalization with multiple categories. In M Knauff, M Pauen, N Sebanz and I Wachsmuth (Ed.) Proceedings of the 35th Annual Conference of the Cognitive Science Society (pp. 3699-3704)

7. Hendrickson, A. T., Carvalo, P. F, Goldstone, R. L. (2012). Going to extremes: The influence of unsupervised categories on the mental characterization of faces and asymmetries in perceptual discrimination. Proceedings of the 34th Annual Conference of the Cognitive Science Society.

8. Hendrickson, Kachergis, G., Faucey, C. M., Goldstone, R. L. (2012) Re-learning labeled categories reveals structured representations. Proceedings of the 34th Annual Conference of the Cognitive Science Society.

9. Hendrickson, A. T., Todd, P. M., Fortenberry, J. D. (2010) Modeling decision making in mate choice within and across relationships. Proceedings of the 22nd annual meeting of the Human Behavior and Evolution Society.

10. Hendrickson, A. T., Kachergis, G., Gureckis, T. M., Goldstone, R. L. (2010) Is categorical perception really verbally mediated perception? Proceedings of the 32nd Annual Conference of the Cognitive Science Society.

11. Goldstone, R. L., Hendrickson, A. T. (2009) Categorical Perception. Wiley Interdisciplinary Reviews: Cognitive Science, 1, 69 – 78.

12. Hendrickson, A. T., Goldstone, R. L. (2009) Perceptual Unitization in Part-Whole Judgments. Proceedings of the 31st Annual Conference of the Cognitive Science Society.

13. Hendrickson, A. T., Goldstone, R. L. (2007) Feature Learning Within Categorization. Proceedings of the 29th Annual Conference of the Cognitive Science Society.

Awards

"IGERT" Fellowship

National Science Foundation, Indiana University, 2010 – 2015

Pre-doctoral Training Award

Indiana Clinical and Translational Research Institute, Indiana University, 2010-2011

Student Research Grant Award

The Kinsey Institute, Indiana University, 2010

Summer Fellowship

Cognitive Science, Indiana University, 2009, 2010

Ad Hoc Reviewing

Frontiers in Psychology, Journal of Memory and Cognition, Journal of Adaptive Behavior,

Proceedings of the Cognitive Science Society, Psychological Bulletin