Spring 2015

Categorizing Fetal Heart Rate Variability with and without Visual Aids

Amanda Jane Ashdown
Old Dominion University

CATEGORIZING FETAL HEART RATE VARIABILITY WITH AND WITHOUT VISUAL AIDS

by

Amanda Ashdown
B.S. May 2012, Old Dominion University

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
May 2015

Approved by:

Mark W. Scerbo (Director)

Debra A. Major (Member)

Christopher Brill (Member)

ABSTRACT

CATEGORIZING FETAL HEART RATE VARIABILITY WITH AND WITHOUT VISUAL AIDS

Amanda Ashdown
Old Dominion University, 2015
Director: Dr. Mark W. Scerbo

This present study examined the ability of clinicians and novices to correctly categorize fetal heart rate (FHR) variability with and without the use of exemplars. Clinicians and undergraduate students were asked to inspect FHR images and determine into which of four categories they belonged. Each participant took part in three conditions: one in which they were provided exemplars of prototypical FHR variability to use during their categorization task, another in which they were provided exemplars of nonprototypical FHR variability to use in their task, and a control condition in which no exemplars were available. The results showed that experts were more accurate and quicker in their category judgments than novices, but this difference was largely limited to the condition with no exemplars. The results also showed that participants correctly categorized more prototypical images than nonprototypical images and that the prototypical and nonprototypical cues were beneficial for experts and novices. The results suggest that providing clinicians with alignable, high similarity visual aids can improve judgments about FHR variability and potentially enhance safety in labor and delivery.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The ability of a person to perceive and identify shapes and patterns is extremely important. The capacity to mentally place relevant signals into correct categories is vital in many work situations. Military pilots need to recognize targets while flying, particularly with unmanned aerial vehicles (UAV) that require them to perceive objects on a computer screen. Also, clinicians rely on their ability to perceive abnormalities in x-ray and ultrasound images. A person working in any of these jobs needs to accurately perceive whether an item is a target of interest or something that can be ignored. Military personnel and clinicians need to be able to effectively categorize relevant and nonrelevant information, as stimuli can often be ambiguous in different situations.

The present study addressed issues faced by clinicians when interpreting fetal heart rate (FHR) tracings. Although there are guidelines established for assessing FHR, there is still potential for clinicians to misclassify FHR variability, which can result in inappropriate operative intervention (i.e., cesarean procedures).

A major issue for clinicians is the difficulty they face in interpreting and classifying FHR tracings as reassuring or ominous. Thus, research on categorization was examined to address the underlying cognitive process needed for this activity. Further, evidence shows that cues and visual aids can be beneficial when categorizing stimuli. Cues can direct attention to salient information and help individuals be more accurate and respond faster to stimuli. Accordingly, the present study also examined the effect of visual exemplars on the ability to categorize FHR variability for both novices and experts.

MATERNAL FETAL HEART RATE MONITORING

Historical Background

Over the years people have attempted to assess fetal well-being by monitoring fetal heart rate (FHR) activity. Presently, there are two methods commonly used for fetal assessment during labor. The first method was introduced after a Swiss surgeon, Francois Mayor, laid his ear on a pregnant woman's abdomen and heard fetal heart tones (Chez, Harvey, & Harvey, 2000). A couple years later, the use of a stethoscope to amplify fetal heartbeat and exclude other sounds (auscultation) was implemented in fetal assessment. Today, auscultation (listening to sounds from the heart with a stethoscope) is performed with low-risk patients every 30 minutes in the active phase of labor and every 15 minutes in the second stage of labor (Sweha, Hacker, & Nuovo, 1999).

The second method of fetal assessment is electronic fetal heart monitoring (EFM) introduced in 1958 (Sweha et al., 1999). EFM was used during the 1960s in an effort to improve fetal and neonatal outcomes by reducing neurological injury and death (Chez, Harvey, & Harvey, 2000; Miller, 2011; Sweha et al., 1999). This method uses an external transducer placed on the maternal abdomen and held in place by an elastic belt. The transducer uses Doppler ultrasound to detect fetal heart motion and is connected to a monitor that records the FHR, along with the mother's contractions, on a continuous strip of paper or a computer screen, as shown in Figure 1 (Sweha et al., 1999). The heart rate tracing is displayed over heavy vertical lines repeated at 60-sec intervals and lighter vertical lines repeated at 10-sec intervals (Chez et al., 2000). The top portion of the tracing shows the FHR in beats per minute (bpm) while the bottom portion displays the intensity of maternal contractions in millimeters of mercury (mmHg) as a function of time.

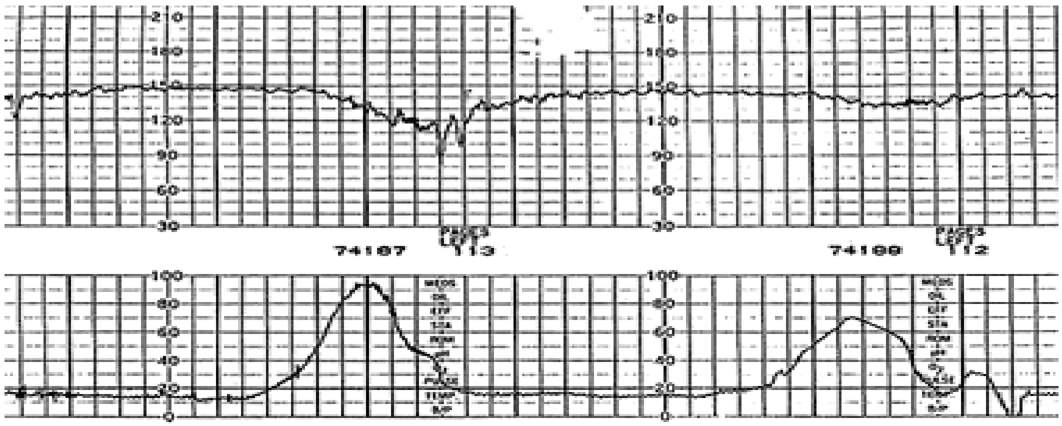

Figure 1. A paper scale of maternal fetal heart rate (MFHR). The FHR is displayed at the top and maternal contractions are displayed at the bottom (Sweha et al., 1999).

The introduction of EFM was thought to be beneficial because of the ability to continuously monitor the fetus and closely assess high-risk patients. Today, in North America the EFM process has become a standard for all patients designated high risk and has been widely applied to low-risk patients as well (Sweha et al., 1999; Miller, 2005; Bailey, 2009). The EFM procedure was reported in 2005 to have been used in 85.4 percent of births (Martin, Hamilton, Sutton, Ventura, Menacker, Kirmeyer, & Munson, 2007).

As interest in FHR monitoring grew, the importance of creating universal terminology became important. Therefore, in 2008, the National Institute of Child Health and Human Development (NICHD), the American College of Obstetricians and Gynecologists (ACOG), and the Society for Maternal-Fetal Medicine jointly sponsored a workshop focused on EFM to revisit interpretation and research recommendations for intrapartum EFM.

Fetal Heart Rate Characteristics

Interpretation of a tracing requires both qualitative and quantitative description of uterine contractions, baseline FHR, FHR variability, presence of accelerations, periodic or episodic decelerations, and changes or trends of FHR tracings over time (Macones, Hankins, Spong, Hauth, & Moore, 2008). Uterine contractions are quantified by the number present in a 10-minute window, averaged over 30 minutes. Normal uterine activity is defined as five or fewer contractions in 10 minutes (Macones et al., 2008).

Fetal heart rate patterns are defined by the characteristics of baseline, variability, accelerations, and decelerations (Macones et al., 2008). The baseline FHR is determined by the mean FHR rounded to 5 bpm during a 10-minute window. The normal FHR range is between 120 and 160 bpm. The baseline rate is considered changed if a shift persists for more than 15 minutes. When the baseline FHR is <110 bpm, it is called bradycardia and considered abnormal. Severe prolonged bradycardia of < 80 bpm that lasts for three minutes or longer is an ominous sign indicating severe hypoxia and is often a terminal event. A baseline greater than 160 bpm is defined as tachycardia and may be a sign of increased fetal stress when it persists for 10 minutes or longer (Sweha et al., 1999).

Normally, the FHR fluctuates from the baseline, which reflects a healthy nervous system and cardiac responsiveness. However, there can be fluctuations in FHR that are irregular in amplitude and frequency. There are four categories of variability: absent FHR variability with an undetectable amplitude range, minimal FHR variability with an amplitude range of 5 or fewer bpm, moderate FHR variability with an amplitude range of

6 to 25 bpm, and marked FHR variability with an amplitude range greater than 25 bpm (Macones et al., 2008). In addition to the variability, there are also several types of FHR patterns that clinicians need to detect.

One type of pattern is an acceleration, defined as an abrupt increase in FHR from its onset to its peak in less than 30 seconds (ACOG, 2010; Macones et al., 2008; Sweha, et al. 1999). At 32 weeks of gestation and beyond, an acceleration has a peak of 15 bpm or more above baseline, with a duration of 15 to 120 sec from onset to return (ACOG, 2010; Macones et al., 2008). Accelerations are usually associated with fetal movement, vaginal examinations, uterine contractions, umbilical vein compression, fetal scalp stimulation, or even external acoustic stimulation. The presence of accelerations is considered a reassuring sign of fetal well-being (Sweha et al., 1999).

Decelerations are classified as early, late, or variable based on specific characteristics. Early decelerations are caused by fetal head compression during uterine contraction, resulting in vagal stimulation and slowing of the heart rate (Sweha et al., 1999). Early decelerations are a gradual decrease in FHR lasting 30 seconds or more. This type of deceleration has a uniform shape and mirrors the contraction; with a slow onset that coincides with the start of the contraction and a slow return to the baseline that coincides with the end of the contraction (ACOG, 2010; Macones et al., 2008; Sweha et al., 1999). These decelerations are not associated with fetal distress and are considered reassuring (Sweha et al., 1999).

A late deceleration is a symmetric decrease in FHR lasting 30 seconds or more, beginning at or after the peak of the uterine contraction and returning to baseline only after the contraction has ended (ACOG, 2010; Macones et al., 2008; Sweha et al., 1999).

Late decelerations are associated with uteroplacental insufficiency and are provoked by uterine contractions. Any decrease in uterine blood flow or placental dysfunction, such as postdate gestation, preeclampsia, or chronic hypertension, can cause late decelerations (Sweha et al., 1999). All late decelerations are considered potentially ominous. A pattern of persistent late decelerations is nonreassuring requiring further evaluation of the fetus (Sweha et al., 1999).

Variable decelerations are abrupt decreases in FHR that vary in shape, depth, and timing in relation to uterine contractions. They typically have a decrease in FHR of at least 15 bpm and a duration of 15 to 120 seconds (Bailey, 2009; Sweha et al., 1999). Variable decelerations are caused by compression of the umbilical cord and are generally associated with a favorable outcome; however, a persistent variable deceleration pattern may lead to fetal distress if not corrected.

In 2008, the NICHD recommended a three-tier classification system to define FHR patterns (Macones et al., 2008). FHR tracings fall into Category I if the baseline is between 110–160 bpm, there is moderate variability, there are no late or variable decelerations, and early decelerations and accelerations are either present or absent (Fedorka, 2010; Macones et al., 2008). Category I FHR tracings are considered a reassuring sign that labor is progressing safely (Fedorka, 2010). Category II tracings include any of the following: a baseline rate indicative of bradycardia or tachycardia, minimal or absent baseline FHR variability not accompanied by recurrent decelerations, marked baseline variability, absence of induced acceleration after fetal stimulation, periodic or episodic decelerations, recurrent variable decelerations accompanied by minimal or moderate baseline variability, prolonged deceleration of more than two, recurrent late decelerations with moderate baseline, or variable decelerations with other characteristics, such as slow return to baseline (Fedorka, 2010; Macones et al., 2008). Category II FHR tracings may or not be problematic and should be closely monitored to assess whether they are normal or abnormal. Category III tracings appear to have absent baseline FHR variability and any of the following: recurrent late decelerations, recurrent variable decelerations, and bradycardia (Fedorka, 2010; Macones et al., 2008). Category III FHR tracings are abnormal and are regarded as ominous, often requiring clinical intervention. Because the FHR fluctuates over time, it is imperative to monitor a category change; FHR tracings may move back and forth between categories depending on the clinical situation and management strategies employed.

Perceiving Fetal Heart Rate Tracings

It is important for clinicians to understand the different implications of FHR patterns to prevent misinterpretation and unnecessary clinical interventions. Clinicians who work in labor and delivery must visually inspect and interpret the FHR tracings, which can introduce a significant source of subjectivity (Menihan & Zottoli, 2001). Continuous EFM is intended to reveal potential problems in fetal well-being; however, misinterpretation of FHR tracing patterns may lead to other problems including risk for fetal injury, unnecessary intervention, or even death (Buscicchio et al., 2010; Sweha et al., 1999). The interpretation of FHR activity and guidelines for management are inconsistent and can result in an increase in inappropriate operative intervention, as well as increases in the cost of obstetrics and malpractice insurance (Barstow, Gauer, & Jamieson, 2010; Berkus et al.,1999; Menihan & Zottoli, 2001; Miller, 2011; Minkoff & Berkowitz, 2009; Sisco et al., 2009; Sweha et al., 1999; Weiss et al., 1997). Because there are no universally accepted definitions of fetal distress, EFM is associated with increased rates of surgical intervention (i.e., cesarean section) resulting in increased costs (Sweha et al., 1999). Since the introduction of EFM, the rate of cesarean delivery increased to 26.1 percent of all births, the highest rate ever reported in the United States (Martin et al., 2007).

Several studies show poor agreement in subjective interpretations of FHR patterns, even among experts (Ayres-de-Campo, Bernardes, Costa-Pereira, & Pereira-Leite, 1999; Bernardes et al., 1997; Gagnon, Campbell, & Hunse, 1993). Interpretation of tracings has been shown to vary among different practitioners interpreting the same tracing (inter-rater), and when the same practitioner examines the same tracing on consecutive occasions (intra-rater; Freeman, 2002). Figueras and colleagues (2005) measured the inter- and intra-observer agreement of visual analysis of fetal heart rate tracings and found poor reliability. Classifications of normal baseline and normal variability showed good agreement, but poor agreement was found for low FHR variability and number of decelerations present in the tracings (Figueras, Albela, Bonino, Palacio, Barrau, Hernandez, Casellas, Coll, & Cararach, 2005). In in a study of midwives, Devane and Lalor (2005) found that agreement was highest in the classification of decelerations and lowest in the assessment of baseline variability. Ayres-de-Campos and colleagues (1999) evaluated inconsistencies in classification of FHR tracings and clinical decisions among experts. The researchers examined inter-observer agreement in the interpretation of tracings with the International Federation of Gynecology and Obstetrics (FIGO) guidelines and found reasonable agreement in the classification of normal tracings compared to suspicious or pathologic tracings. Thus, clinicians had difficulty with the abnormal tracings, the very ones requiring more attention. Based on research showing poor agreement among clinicians for the classification of MFHR tracings and the difficulty clinicians experience when interpreting the tracings, it is important to examine further the ability to categorize the FHR tracings.

OBJECT PERCEPTION AND CATEGORIZATION

Historical Background

For years, scientists have investigated human ability to perceive objects and shapes in the environment. Our daily activities, and even survival, depend on the way we accurately recognize objects. The ability to perceive objects has been widely studied since the early 1900's with the introduction of Gestalt theory by Wertheimer and his colleagues (Hartmann, 1935; Wagemans, Feldman, Gepshtein, Kimchi, Pomerantz, van der Helm, & van Leeuwen, 2012; Wertheimer, 1938a). Wertheimer argued that structured wholes or Gestalten, rather than sensations, are the primary units of mental life (Wertheimer, 1938a). The Gestaltists' view is that humans perceive the simplest possible interpretation of elements in the environment; furthermore, we have the ability to immediately detect relationships such as symmetry and continuity, a phenomenon called perceptual organization (Lowe, 1985). The Gestalt principles of organization describe how figural properties are perceived as patterns. Since the emergence of Gestalt theory, a number of different principles of perceptual organization have been proposed in order to account for both static and dynamic aspects perceptual grouping (Wagemans et al., 2012). One principle that determines perceptual organization, according to Wertheimer (1938b), is similarity: The law of similarity states that similar features are grouped together.

Research on the principles of perceptual organization have helped us understand how we categorize and perceive patterns in our environment. Similarity is a very important principle of categorization, and our ability to make sense of our experiences and knowledge depends on our capability to categorize relevant information.

Categorization

Categories play a fundamental role in our daily lives and are the basis for decision making in most professions (e.g., air traffic control, healthcare, and engineering). Categories are sets of objects or events that have similar features and are grouped together because of their similarity (Rosch, 1975). Categorization refers to an individual's ability to assign objects or other stimulus patterns to categories (Paradis, Guo, Olden-Stahl, & Moulton, 2012). An important Gestalt principle of perceptual organization is that similar things will tend to be grouped together, and most theories of categorization share the assumption that similar examples tend to belong to the same category.

Rosch and Mervis (1975) conducted several studies to examine about how humans mentally organize the entities in their environment. They noted that features in the world are not distributed randomly across entities, but instead, tend to occur in clusters, and suggested that we group objects together that share such clusters of features to form categories. In their research, they found that the most prototypical members of categories are those which have the greatest family resemblance to other members in the category and have fewer attributes in common with members of other categories (Rosch & Mervis, 1975).

Many categorization models have received attention, two of which are the exemplar model and prototype model. In exemplar models, the learner stores mental representations of exemplars, grouped by category, and then classifies new objects on the basis of their similarity to the previously learned examples (Estes, 1986). The more similar the target instance is to concrete exemplars of a category, the more likely it will be placed in that category (Cohen & Basu, 1987). A new item is judged to be part of a category to the extent that it is sufficiently similar to an exemplar stored in memory (Storms, DeBoeck, & Ruts, 2000). In prototype models, the learner forms an abstract representation of each category represented in a series of learning experiences, then classifies new instances on the basis of their distances from the mental representation of their category prototypes (Estes, 1986). These models suggest that many categories do not have defining properties, but rather are organized around specific examples acquired during learning (i.e., exemplars) or around an average example (i.e., a prototype).

Early theories of categorization assumed that rules, prototypes, or exemplars were used exclusively to mentally represent categories of objects. More recently, hybrid theories of categorization have been proposed suggesting there are multiple ways categories can be represented. Johansen and Palmeri (2002) argue these can even represent shifts that occur during category learning. In three experiments, participants learned to categorize stimuli with feedback, and the researchers tracked how participants generalized their category knowledge by testing them on critical transfer items without feedback. The results revealed individual differences in the generalization patterns exhibited by subjects, and those generalizations changed systematically with experience. Early in learning, subjects generalized on the basis of single diagnostic dimensions, consistent with the use of simple categorization rules. Later in learning, subjects generalized in a manner consistent with the use of similarity-based exemplar retrieval, attending to multiple stimulus dimensions. Other models suggest a psychological transition from prototype-based to exemplar-based processing during category learning. In a series of experiments, Smith and Minda (1998) evaluated participants' categorization strategies and standard categorization models at successive stages of learning smaller, less differentiated categories and larger, more differentiated categories. Their results revealed that the prototype model had a strong early advantage that gave way slowly when learning larger, more differentiated categories and the exemplar model dominated in learning with small, less differentiated categories. Therefore, Smith and Minda's (1998) experiments provide evidence that early on, participants' performance is consistent with prototype-based processing with a gradual transition later on to strategies that feature exemplar processing given highly familiar training exemplars. Thus, what all these categorization models have in common is that people tend to judge the similarity of stimuli in order to group the stimuli together into categories.

Although similarity might be an explanation of how people categorize, observations suggest that similarity can be influenced by context. Medin, Goldstone, and Gentner (1993) conducted several experiments that demonstrate that properties recognized about a certain stimulus depend on its pairing with another stimulus, suggesting that similarity is dynamic and can change with experience. According to Medin and colleagues, similarity can be viewed as a guideline for categorization but can be overridden by other forms of knowledge. Not only has context been shown to influence categorization, but inductive inference can play a role in categorization as well.

Gelman and Markman (1986) examined whether inductions can be made without perceptual support, that is when an object does not look like other members of its category and when a property is unobservable. The researchers compared category membership and perceptual similarity in an induction reasoning task. Young children were shown pictures of two animals and were taught about different properties of each animal. Then they were asked which property was true of a new animal that was perceptually similar to one alternative but belonged to the category of a perceptually different alternative. For example, children had to decide whether a shark is more likely to breathe like a tropical fish because both are fish, or as a dolphin does because they look alike. Gelman and Markman (1986) found that by age 4, children can use categories to support inductive inferences even when category membership conflicts with appearances (similarity). Moreover, the children distinguished properties that support induction within a category (e.g., means of breathing) from those that are determined by perceptual appearances (e.g., weight).

The studies mentioned above suggest that perceptual similarity may serve as an initial classification strategy, but categorization can be refined and modified by knowledge. Evidence also suggests that categories can be organized around goals. Barsalou (1983, 1985) studied the organization of categories constructed in the service of goals and demonstrated that the determinants of a particular category's graded structure (i.e., members of a category varying in how good or typical they are of their category) can change with context. Whereas ideals may determine a category's graded structure in one context, central tendency may determine a different graded structure in another. Ideals are characteristics that exemplars should have if they are to best serve a goal associated with their category. Central tendency refers to any kind of information about a category's exemplars, which is another way to view an exemplar's family resemblance. Therefore, Barsalou (1983, 1985) argued the organization of categories can be determined by a person’s goal at a given time. Collectively, these studies illustrate that the categorization of objects and pattern stimuli is a dynamic process.

Categorization of Ambiguous Stimuli

The real world does not always have clear-cut stimuli that are easily categorized, but has instances where stimuli can be ambiguous and vague. When asked to indicate which items from a set of candidates belong to a particular natural language category, individuals disagree on which items should be considered category members (Verheyen & Storms, 2013). Alternatively, categorization differences are said to be due to *ambiguity* when individuals employ different criteria. Categorization differences are due to *vagueness* when individuals employ different cut-offs for separating members from non-members (Verheyen & Storms, 2013). There are important implications in examining how people perform when categorizing stimuli that cannot easily be distinguished.

One study had an objective to examine the RT-distance hypothesis which is motivated by decision-bound theories of categorization. Ashby, Boynton, and Lee (1994) examined response times (RT) of categories that fell near or far from the division point that separates the exemplars of the contrasting categories. Decision-bound theories of categorization assume that the perceptual effect of each presentation of a category exemplar can be represented as a point in a multidimensional perceptual space and that repeated presentations of the same exemplar do not always lead to the same perceptual effect. Furthermore, decision bound theory assumes that a practiced subject divides the perceptual space into regions and associates a category label with each region. On each trial, the subject categorizes an object by determining in which region the stimulus representation falls. The partition between two response regions is called the decision bound (Ashby, Boynton, & Lee, 1994; Maddox & Ashby, 1993). The RT-distance hypothesis states that RT decreases with the distance between the perceptual effect and the decision bound. Categorization response times were examined in three separate experiments and in each experiment exemplars varied on two physical dimensions. Three different types of stimuli were used: (1) horizontal and vertical line segments of varying length that were joined at an upper left corner, (2) rectangles of varying width and height, and (3) circles or semicircles of varying size with a radial arm of varying orientation. The results revealed that RT decreased with distance from the stimulus to the categorization decision bound. Thus, stimuli falling near a category bound have ambiguous category membership, hence categorization is slow, whereas stimuli far from the category boundary are easy to classify, and therefore, result in shorter response times.

Not only are stimuli responded to faster when they are further away from the category boundary, but stimuli that are similar to their prototypes, and therefore are more discriminable from other categories, are learned quicker. In Vandierendonck's (1984) study, participants learned to classify random patterns generated from two prototypes with either a short or long inter-prototype distance. The study revealed the tendency to call a pattern "new" increased with the distance between the pattern and its prototype. Learning was shown to be faster when the distance between the categories is larger (i.e., when the categories are more discriminable).

Although learning is faster when stimuli are similar to their prototypes, it is evident that learning with high exemplar diversity can aid in a person's ability to generalize to novel stimuli less typical of the category prototype. For example, Das-Smaal and De Swart (1984, 1986) argued that categorization models must be capable of representing variation among exemplars within a category. Their studies reveal that a central representation (prototype) is abstracted from the experienced exemplars of a category, and classification is based on distance from this prototype. They investigated forms of variation within categories, more specifically, the similarity of a dimensional value variant to a prototypical value (i.e., typicality of variants). Exemplars having the same dimensional values may differ with respect to how typically they exhibit these values. Their results revealed that typical exemplars were classified faster and with more certainty than less typical ones. However, following learning, broad range experience resulted in fewer classification errors than narrow range experience, due to better classification of both medium typical and (new) atypical stimuli in the broad range condition.

Similarly, Hahn, Bailey, and Elvin (2005) examined the effect of within-category diversity on a person's ability to learn perceptual categories, the inclination to generalize categories to novel items, and the ability to distinguish new items from old. The researchers manipulated exemplar diversity for one of two perceptual categories of schematic flower images. Category membership for these stimuli was determined by the flowers' head and stem areas. Participants learned to distinguish between pictures of both categories and were assessed with old and novel stimuli that were either similar or dissimilar to the prototype. In one training condition with low exemplar diversity, flowers presented for the reference category were very similar to the prototype of that category. In the other training condition, the flowers presented were more diverse and dissimilar from the prototype. Hahn et al. (2005) found that learning was impaired as exemplars differed from the prototype. Thus, training with high exemplar diversity made learning more difficult than training with low exemplar diversity. However, higher exemplar diversity during training increased generalization to novel stimuli outside the range of trained stimuli during test (Hahn et al., 2005).

Many variables influence the way in which humans classify certain exemplars into categories. People tend to categorize stimuli based on their experiences with learning specific categories. For example, people who only learned and experienced prototypical examples of a category may not be able to generalize to new, atypical stimuli that belong in the same category. Furthermore, people who have a broad range of experience with ambiguous stimuli are able to generalize to new stimuli and categorize them more reliably. Therefore, it is important to examine the way in which experts accurately classify different exemplars into their correct categories, and how they differ from novices in a categorization task.

Experts and Novices

Compared to novices, experts spend many years learning to classify objects on the basis of subtle perceptual cues, and categorization often becomes automatic because of the vast majority of examples they have experienced. Palmeri (1997) investigated the effects of exemplar similarity on the development of automaticity with a task in which participants judged the numerosity of random patterns of 6 to 11 dots. After several days of training, response times were the same at all levels of numerosity, indicating that automaticity had been achieved. Following training, participants were asked to judge the numerosity of old patterns and new patterns of varying similarity to the old patterns. Judgment response times were determined by the similarity of the transfer patterns to the old training patterns. Old patterns were judged just as quickly as those from the end of training, and new patterns were judged just as slowly as those from the start of training.

Also, new moderate-similarity and low-similarity distortions were judged with fast response times, in accord with their similarity to the old patterns. The researchers then investigated the influence of exemplar similarity on the development of automaticity. At each level of numerosity, people were trained with three types of patterns: moderate-similarity patterns with moderate-level distortions from the prototype, low-similarity patterns with significant distortions from the prototype, and unrelated patterns that were generated randomly. The results revealed that numerosity judgments became automatized more quickly for moderate-similarity patterns than for low-similarity or unrelated patterns. Further, throughout training, the moderate-similarity patterns were judged more quickly than the low-similarity or unrelated patterns (Palmeri, 1997). Therefore, these results suggest that experts are able to rapidly categorize similar novel dot patterns based on learned experiences.

The study by Palmeri (1997) showed that people can learn to categorize ambiguous stimuli such as dot patterns. However, in real world tasks using more meaningful stimuli, evidence shows that people in many domains process stimuli differently depending on their level of expertise. In an experiment that examined the role of radiological expertise in X-ray image perception, observers with four different levels of experience performed a recognition task on slides of faces and chest X-ray films (Myles-Worsley, Johnston, & Simons, 1988). Half of the X-ray films were normal and the other half revealed clinically significant abnormalities. Recognition for faces was uniformly high across all levels of radiological experience. The results revealed that memory for abnormal X-ray films increased with radiological experience and were equivalent to memory for faces. Moreover, expert radiologists appear to process X-ray images the way that we process faces, by quickly detecting and devoting processing resources to features that distinguish one stimulus from another. There is considerable ambiguity in the sensory input of X-ray images because of the low resolution of the radiographic image compared with a photographic image. However, Myles-Worsley et al. (1988) showed that when an experienced radiologist and an untrained observer view the same X-ray image, they presumably perceive it differently.

Furthermore, in a study aimed at identifying expertise in perceiving and interpreting complex, dynamic visual stimuli, Jarodzka, Scheiter, Gerjets, and Van Gog, (2010) had professors and novices examine four digital videos of swimming fish. After watching the video, performance was assessed by their ability to name locomotion pattern correctly, describe which body part had been moving, and describe how each body part had been moving. Compared to novices, experts were able to perform the task faster and more accurately, as indicated by their better description of locomotion patterns and their higher use of correct technical terms. Jarodzka et al., (2010) also found that experts attended to more relevant information than novices, who often attended to irrelevant information. Another study investigated diagnosis by clinicians of varying levels of expertise of authentic pediatric video cases of children with seizures and with disorders imitating seizures. The researchers found that the more experienced clinicians spent more of their time looking at relevant areas and were more accurate in visual diagnosis (Balslev, Jarodzka, Holmqvist, de Grave, Muijtjens, Eika, & Scherpbier, 2012).

Collectively, these studies reveal that experts and novices perform differently in classification tasks across many different domains. Evidence suggests that novices often have trouble discriminating relevant from irrelevant information in complex visual tasks, such as diagnosing radiological images or interpreting FHR tracings. Klein and Hoffman (1993) state that a major difference between novices and experts concerns what they perceive. Even when novices perceive all the relevant details, they fail to see the relevant relations among them. In real life complex visual tasks, a lot of information is present and not all information is relevant. Novices tend to focus on features that are obvious or salient rather than on those relevant to the task (Klein & Hoffman, 1993; Lowe, 1999).

Cues and External Memory Aids

Studies reveal that people have difficulty categorizing ambiguous stimuli. Novices tend to focus on obvious information and have difficulty in determining what is relevant to specific tasks, as well as difficulty in noticing when critical information is missing (Klein & Hoffman, 1993). Thus, researchers have studied the potential benefits of cues and external aids for improving a person's ability to correctly categorize relevant information. Research in human factors has shown that cues can be effective for improving signal detection in a wide variety of tasks. Cues are able to direct attention; therefore, when a cue is accurate, observers can attend to the cue, respond faster, and make fewer errors when detecting signals (Wickens & Hollands, 2000).In a series of experiments, Posner, Snyder, and Davidson (1980) demonstrated that detection efficiency was affected by cueing participants to where in space a stimulus would occur. Previous research has demonstrated that attending to a cued location in space leads to faster response times when a stimulus is presented in that location. Posner, Nissen, and Ogden (1978) showed that performance in detecting or discriminating a target significantly increased when the location of the target was previously cued. In their spatial cueing paradigm, a central arrow cue preceded the onset of a target. The cue correctly indicated the target's location on 80% of the trials (valid trials) and pointed to the opposite location in 20% of the trials (invalid trials). Reaction times were found to be faster to the valid trials. The results revealed that focusing attention on the cued location enhanced processing of the target stimulus, which resulted in faster responses and higher accuracy. Furthermore, in a recent study, Gunn, Warm, Nelson, Bolia, Schumsky, and Corcoran (2005) utilized a vigilance task in which threat detections (critical signals) required observers to perform a subsequent manual target acquisition task. The study revealed that visual, spatial-audio, and haptic forms of cues were effective in enhancing unmanned aerial vehicle (UAV) operators' performance in target acquisition. The speed with which observers detected threats increased for each of the cueing conditions compared to the no-cueing control.

External visual aids can not only improve performance on target detection tasks, but can also facilitate training. For example, Chaney and Teel (1967) examined the effectiveness of visual aids for training on an inspection task with experienced machine parts inspectors. Their results showed that inspection training by itself resulted in a 32% increase in detections, visual aids alone resulted in a 42% increase, and the use of both visual aids and training resulted in a 71% increase. The findings of this study suggest that the incorporation of visual aids into a training program can improve performance on visual detection tasks. Moreover, visual aids can be beneficial for improving a novice's performance by pointing out information relevant to the task at hand, since novices tend to focus on features that are obvious or salient (Klein & Hoffman, 1993).

The type of cue or exemplar is also important for training. Evidence shows that comparison learning is a promising approach for learning complex visual tasks (Gentner

& Gunn, 2001; Markman & Gentner, 1997). Comparison of contrasting images can help students to isolate relevant but less conspicuous information (Gentner & Gunn, 2001). According to structural alignment theory, when individuals compare stimuli, features and relations within one stimulus are systematically matched to features and relations in the other stimulus (i.e., aligned; Markman & Gentner, 1997). Differences between two stimuli become more salient as a result of this matching process; thus, information that is more salient is easier to notice, which helps discriminating relevant information (Gentner & Markman, 1997). For example, when medical students study radiological images of diseases, comparison of images with and without abnormalities can help them learn to discriminate relevant, disease-related information. Kok, de Bruin, Robben, and van Merriënboer (2013) found that on a visual diagnosis test, medical students who were allowed to study by comparing diseases on chest x-ray images with normal images were better able to diagnose focal diseases than students who could not make comparisons. The results show that comparisons with normal images made it easier to discriminate relevant information (Kok et al., 2013). Kurtz, Boukrina, and Gentner (2013) investigated the effect of presenting training items for comparison during supervised classification learning of novel relational categories. Their stimuli consisted of line-drawn images depicting rock arrangements made by fictional cultures. In a test phase measuring learning and transfer, the comparison group significantly outperformed a control group receiving an equivalent training session of single-item classification learning. Comparison-based learners also outperformed the control group learners on the ability to accurately classify items from a novel domain that was similar to the training materials.

Comparison training can facilitate detection performance but it turns out that placement of the comparison stimulus is important too. Kurtz and Gentner (2013) argue that an *alignable* (high-similarity) comparison standard can improve detection performance by pointing out relevant information. These researchers examined participants' ability to find an anomalous bone in drawings of animal skeletons. Target items including the anomaly were presented either alone or with a correct alignable standard. The correct standard was presented in either a regular (high alignable) or mirror-reversed (low alignable) manner. The high-alignable standard was identical to the target skeleton with the exception of the anomaly present. The low-alignable standard was always the mirror-reverse of the high-alignable standard. Thus, the two standards were equal in terms of the information present, but they differed in their perceptual alignability with the target. Their results showed increased accuracy when an alignable comparison standard was present during the detection task: participants showed better accuracy in detecting the anomalous feature when given a standard against which to compare the target item than when given only the target. Furthermore, results showed increased accuracy when the comparison standard was more easily alignable (high alignable). The alignable comparison standard helped participants detect relevant targets. Thus, the evidence suggests that comparison processes enhance people's ability to detect subtle anomalies in complex stimuli by highlighting key differences between the stimulus and the comparison standard. The researchers also investigated whether comparison as opposed to single-item training led to improved detection of anomalies. Their results showed that comparison training led to improved detection of anomalies in subsequent novel examples presented as isolated targets. Therefore, comparison-based learning is advantageous and detection of nonobvious anomalies can be improved by providing alignable standards next to targets. Collectively, the studies described above show that external visual aids that are aligned with relevant features in the training stimuli can be beneficial in pointing out relevant information and improving performance in target detection tasks.

The Present Study

The present study was designed to address how individuals discriminate between categories of FHR variability. There were four categories of FHR variability that participants needed to classify: absent, minimal, moderate and marked. Participants were instructed to identify the category in which each example belongs with and without the presence of alignable cues. The participants consisted of novices and experts in order to examine whether experience affects the way the FHR variability examples are categorized.

The first goal was to assess whether classification of FHR variability is more difficult as the examples deviate further from the prototype for each category. Research indicates that categorization of stimuli becomes more difficult as the distance between the stimulus and the prototype increases (Ashby et al., 1994; Das-Smaal & De Swart, 1984, 1986). The first hypothesis proposed that the classification of variability images was expected to be easier, and more accurate, if they resemble the prototype and was expected to become more difficult as the examples move away from the center and to the boundaries of a particular category, making it harder to distinguish from one category or another. Therefore, it was hypothesized that classification accuracy will be higher for the prototypical examples compared to the nonprototypical examples.

When detecting ambiguous and vague stimuli, research shows that participants benefit from comparison learning and using visual aids in a target detection task. Based on the evidence that external visual aids increase detection performance (Hall et al., 2012; Loft et al., 2013; Kok et al., 2012; Kurtz & Gentner, 2013), the second goal of this study was to examine categorization performance when given the opportunity to use exemplars. Participants were able to use an alignable standard cue to aid in the categorization task. Evidence shows that people, in general, benefit from using cues in a detection task (Chaney & Teel, 1967; Gunn et al., 2005; Posner et al., 1978; Posner et al., 1980). The cues in this study were prototypical and nonprototypical exemplars of each of the four FHR variability categories and were placed under each example as shown in Figure 2. Thus, participants should benefit from the cues since they can be used to aid in pattern matching. Therefore, the second hypothesis proposed that the presence of exemplars was expected to significantly improve participants' performance.

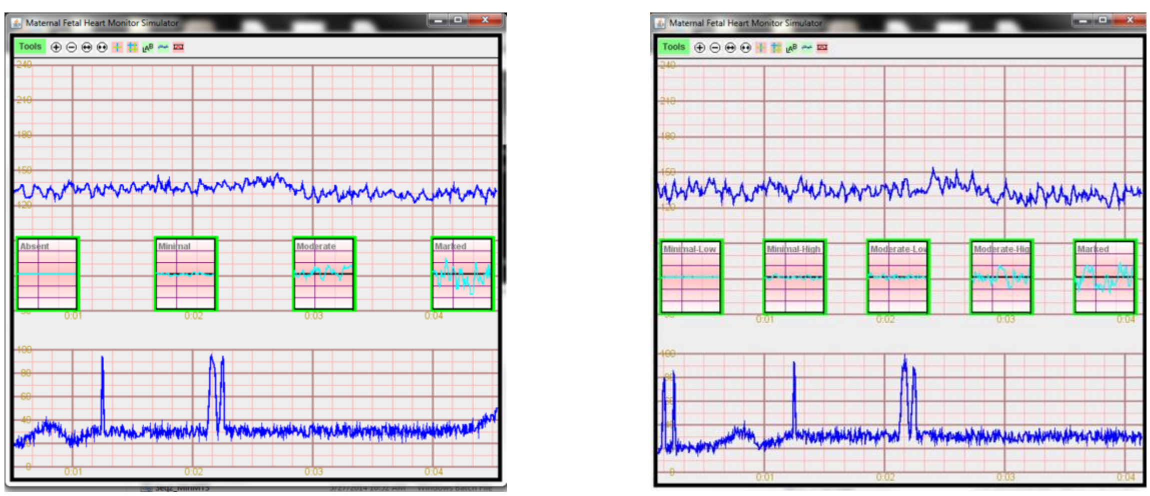

Figure 2: The left image shows moderate FHR variability with prototypical visual aids placed below the FHR and the right image shows moderate variability with nonprototypical visual aids placed below the FHR.

The ability to discriminate between categories is based not only on how similar the stimuli are to the category prototype, but evidence shows that people in many domains process stimuli differently depending on their level of expertise (Balslev et al., 2012; Jarodzka et al., 2010; Klein & Hoffman, 1993; Myles-Worsley et al., 1988). The third goal of this study was to examine how experts (i.e., clinicians) and novices differ when categorizing FHR variability with and without exemplars present.

Evidence suggests that experts are better than novices at noticing relevant information (Klein & Hoffman, 1993); therefore, clinicians may have an advantage over novices when categorizing the FHR variability. Because clinicians have extensive experience examining FHR tracings, they may have internalized prototypical representations of the primary categories of FHR variability. However, based on the categorization research, evidence shows that people will generally base their categorization decisions on how similar the given example is to the learned exemplar or prototype (Das-Smaal & De Swart, 1984, 1986; Cohen & Basu, 1987; Palmeri, 1997; Storms, DeBoeck, & Ruts, 2000; Johansen & Palmeri, 2002; Hahn et al., 2005). Thus, clinicians may be able to make their decisions using an internalized prototype of each category acquired over their years of experience. Evidence reveals that people are able to generalize new, moderately-similar stimuli and categorize them correctly when they are more familiar and have more experience with the stimuli (Palmeri, 1997; Hahn et al., 2005). Furthermore, experts are able to apply their experience and what they have learned in their domains in order to detect critical information by noticing relevant details that novices may miss (Klein & Hoffman, 1993). Therefore, the third hypothesis proposed that clinicians may be able to generalize when the FHR varies from the prototype and perform significantly better than novices when categorizing FHR variability, particularly when no cues are available.

Regarding the cues, however, the categorization of the FHR variability could be accomplished purely by perceptual pattern matching allowing both clinicians and novices to perceptually categorize the examples by matching them to the stimuli. Thus, it is possible when cues are present that clinicians would have no advantage over novices during the categorization task. However, FHR tracings may be vague, especially when the examples are not prototypical of the categories. Hall, Hannon, Leisk, Wolfberg, and House (2012) provided evidence that expert performance was improved when using an external aid. In their study, Hall et al. (2012) developed a prototype electronic ruler for the assessment of FHR variability on an electronic monitor. The electronic ruler consisted of horizontal bands that were sized and colored to embed the four FHR variability categories, and the FHR variability categories were represented with different colors to permit clinicians to rapidly assign a variability category to a segment of FHR data. The results of Hall and colleagues' (2012) study revealed that accuracy of expert variability assessment was significantly improved.

In the present study, the cues were changed to an alignable standard of prototypical examples of each variability category and may be superior since alignable cues have been shown to improve detection of subsequent novel examples by pointing out relevant information (Kok et al., 2013; Kurtz et al., 2013; Kurtz & Gentner, 2013). Thus, the alignable cues may aid clinicians' performance, but particularly when the examples are difficult and deviate from the prototypes. Thus, the fourth and final hypothesis was that there may be a benefit of exemplars for experts, but limited to FHR examples that are at the boundaries of each category and not the prototype of each category. Therefore, a 3-way interaction between prototypicality of the examples, the cueing conditions, and experience was expected: exemplars should be beneficial for everyone, but limited to the less prototypical examples for experts.

METHOD

Participants

Participants in this study consisted of a convenience sample of novices and experts. The novices were 41 Old Dominion University students (24 females, 17 males) recruited through the SONA research participation system, and the experts were 33 clinicians working in labor and delivery (21 nurses, 10 residents, 1 faculty physician, and 1 midwife) from Eastern Virginia Medical School (EVMS) and Sentara. All nurses, the midwife, and nine residents were female, and the faculty physician and one resident was male. To achieve a power of .80 with a medium effect size and an alpha of .05, G*Power 3.1 statistical software indicated a total of 34 participants were required for each group to observe a medium effect of .25 for these analyses (Cohen, 1988).

Design

The current study employed a 2 (expert vs. novice) × 3 (cue condition) × 2 (prototypicality) mixed design. Experience level was the between-subjects variable with two levels: experts, and novices. The within-subjects variables were the presence of cues and the prototypicality of examples. The dependent measures were response times and accuracy (i.e., the total number of examples each participant categorized correctly).

For this study, participants were required to categorize 270 images. Each participant viewed 180 example images presented with exemplars (prototypical or nonprototypical cues) and 90 images without exemplars. There were 10 images created for each category prototype (i.e., absent, minimal, moderate, and marked) and 10 images near the boundaries between the categories. Each image was presented with no time limit until the participants indicated which variability category the examples belong in by using the keyboard. Once participants selected a response, the next example immediately was presented.

FHR Display

The images used in this study were created using a MFHR simulator created by Belfore and colleagues (Belfore, Scerbo, & Anderson, 2007). The FHR signals were simulated by producing a stream of heart beats and then determining the average heart rate within a 4-sec interval. The interval between heart beats was a nominal interval representing the combination of the nominal heart rate, heart rate variation, and a random component whose standard deviation was equal to the (variability/100) * the nominal interval. This calculation ensured that the symmetry of the signals and variations scale well with heart rate and variation amplitude changes.

Static images (snaphsots) were produced from the dynamic tracing described above and were presented on the computer screen using the Superlab 5 software. Each static image had a white background and two red grids, one for the fetal heart rate and another below for the maternal contractions. The tracings were displayed in blue. An example image is displayed in Figure 3.

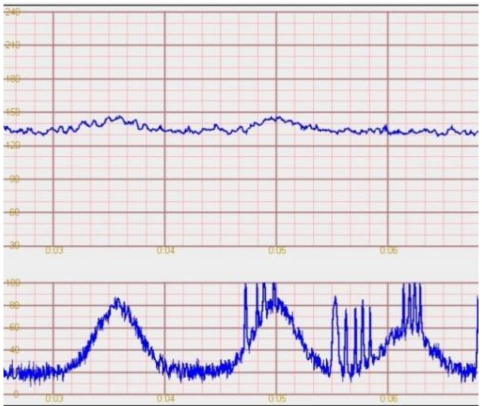

Figure 3: Image of moderate FHR variability. The FHR is shown in the top portion of the display while the maternal contractions are shown in the bottom portion of the window.

Fetal heart rate beat-to-beat variability differed according to the categories defined by the NICHD (Macones et al., 2008). The ranges defined by the NICHD and the values selected for the prototype and boundary images for this study are shown in Table 1.

Table 1

Standard variability ranges and the values selected for the prototypical and boundry variabilities.

Variability	**Prototypical**	**Boundary (Nonprototypical)**
Absent 0 bpm	0 bpm	
Minimal > 0 and ≤ 5 bpm	3 bpm	1 bpm & 4 bpm
Moderate 6-25 bpm	15 bpm	7 bpm & 20 bpm
Marked >25 bpm	35 bpm	30 bpm

The prototypical images were created by using the middle value within each category of FHR variability. The nonprototypical images were created by using values near the boundaries separating each FHR variability category. The visual exemplars used to aid participants in the categorization task were created by taking snapshots of each value used for the prototypical and nonprototypical images and labeling it with the correct variability category. They were designed to be placed side-by-side and under the FHR tracing (see Figure 2).

Procedure

The experiment for the novices (undergraduate students) took place on campus at Old Dominion University. The undergraduate students recruited through SONA received course credit. After arriving at the laboratory, participants were asked to read and sign the Informed Consent Form (Appendix A). The participants were randomized across the initial three cue conditions using the Superlab 5 software. The researcher read the general instructions to each participant, repeating any instructions as necessary. The participants were then shown examples of each variability type and were given the opportunity to practice categorizing five examples of each cue condition with feedback. After the practice to get them familiar with the categories, the participants were seated at a computer to begin the first block (with or without the cues) of the experiment. They each had an opportunity to take a 5-minute break after the first block and then began the second block of the experiment.

The experts were asked to participate voluntarily by email and the experiment took place in a conference room at Sentara Hospital in Norfolk. The residents and nurses signed a consent form provided by the EVMS IRB before the start of the study. After arriving, the researcher read the instructions to each participant, repeating any instruction as necessary. The participants were then shown the same example images the novices were shown in order to become familiar with the FHR images produced by the simulator.

The participants were randomized across the initial three cue conditions. After the practice, the participants were seated at a computer to begin the first block (with or without the cues) of the experiment. They each had an opportunity to take a 5-minute break after the first block before beginning the second block of the experiment.

Each participant was required to judge and categorize each image by pressing a single key on the computer keyboard. Participants pressed the "A" key if they thought the image presented was of absent variability, "V" key if the image was of minimal variability, "M" key if the image was of moderate variability, and the "L" key if the image presented marked variability. The keys on the computer keyboard were labeled as absent, minimal, moderate, or marked on the above mentioned keys for each participant for their convenience. After each block was completed, the participants were taken back to the initial start screen informing them that the study had ended. Overall, the study lasted about 30 minutes.

RESULTS

All data were screened for outliers prior to running analyses and none were identified. The results were analyzed using a 2 (experience) × 3 (cue condition) × 2 (image type) analysis of variance (ANOVA) for a mixed design, with the cue conditions and the image type (prototypical or nonprototypical) as the within-subjects factors and the experience level as the between-subjects factor. The dependent measures analyzed were response times and accuracy (number of correctly categorized images). The alpha level was set at .05.

Accuracy

Accuracy was derived by obtaining the number of correctly categorized examples in each condition divided by the total number of example images each condition contained, or proportion of correct responses. The results of the ANOVA are shown in Table 2.

Table 2

Results of the Analysis of Variance for Proportion of Correct Responses

Source	SS	Df	MS	F	*p*	partial η^2
Between Subjects						
Experience (E)	0.15	1	.15	4.9	.03	.06
Error	2.19	72	.03			
Within Subjects						
Image (I)	3.17	1	3.17	186.6	.00	.72
I x E	0.01	1	.01	0.84	.36	.01
Error	1.22	72	.02			
Condition (C)	0.28	2	.14	25.24	.00	.26
C x E	0.04	2	.02	3.36	.04	.05
Error	0.80	144	.01			
I x C	0.08	1.82	.05	9.48	.00	.12
I x C x E	0.01	1.82	.01	1.37	.26	.02
Error	0.62	131.10	.01			

Levene's tests were used to check for equality of variance and the results concluded that there was adequate homogeneity of variance. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the interaction between image type and cue condition, $\chi 2 = 7.36$, $p < .05$. A Greenhouse-Geisser correction was used for this effect. For post hoc tests, the Bonferroni Sidak test was used. Descriptive statistics for accuracy are shown in Table 3.

Table 3

Means and Standard Deviations for Proportion of Correct Responses.

Experience	Image Type	Cue Condition	*M*	*SD*
Expert	Prototypical Image	Prototypical Cue	0.92	0.1
		Nonprototypical Cue	0.93	0.1
		No Cue	0.87	0.1
	Nonprototypical Image	Prototypical Cue	0.76	0.11
		Nonprototypical Cue	0.75	0.09
		No Cue	0.74	0.11
Novice	Prototypical Image	Prototypical Cue	0.88	0.1
		Nonprototypical Cue	0.94	0.07
		No Cue	0.82	0.13
	Nonprototypical Image	Prototypical Cue	0.73	0.12
		Nonprototypical Cue	0.71	0.1
		No Cue	0.67	0.12

Image Type. A significant effect for image type was observed. Both experts and novices correctly categorized more prototypical images (*M* = .90, *SD* = .10) than nonprototypical images (*M* = .73, *SD* = .11). A significant interaction was also observed for image type and cue condition. A plot of the interaction is shown in Figure 4. A test of simple effects showed that for the prototypical images, participants correctly categorized significantly more images in the nonprototypical cue condition (*M* = .94, *SD* = .08) compared to both the prototypical (*M*= .90, *SD* = .09) and no cue conditions (*M* = .85, *SD* = .12), and participants correctly categorized more images in the prototypical cue condition compared to the no cue condition. The test of simple effects also showed that for the nonprototypical images, participants correctly categorized more images in the prototypical (*M* = .74, *SD* = .12) and nonprototypical cue condition (*M* = .73, *SD* = .10) compared to the no cue condition (*M* = .70, *SD* = .12). Moreover, the test of simple effects also showed that the difference between prototypical and nonprototypical images were significant in all three cue conditions, with fewer examples categorized correctly for the nonprototypical images.

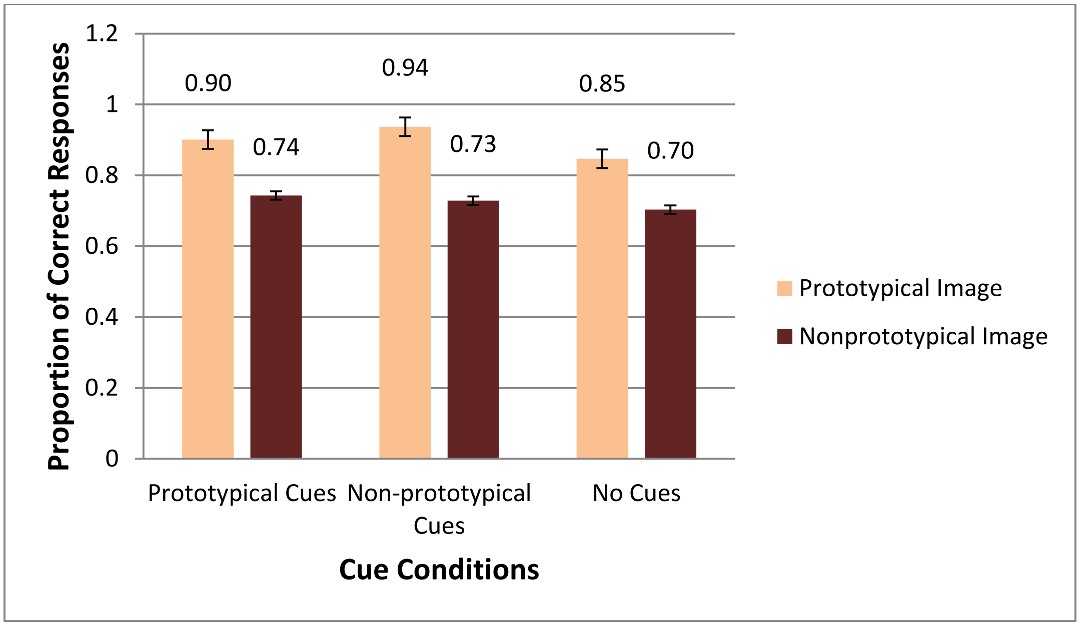

Figure 4. Mean proportion of correct responses for image type as a function of each cue condition.

Cue Conditions. A significant effect for cue conditions was observed. Participants correctly categorized more images in the prototypical (M = .82, SD = .11) and nonprototypical (M = .83, SD = .09) cue conditions compared to the no cue condition (M = .78, SD = .12).

Experience. A significant main effect for experience was observed. The experts correctly categorized more images (M = .83, SD = .10) compared to the novices (M = .79, *SD* = .11). A significant interaction for experience and cue condition was also observed. A plot of the interaction is shown in Figure 5. A test of simple effects showed that experts (*M* = .81, *SD* = .10) correctly categorized significantly more images in the no cue condition compared to novices (*M* = .74, *SD* = .12). The test of simple effects also showed that experts correctly categorized more images in the nonprototypical (*M* = .84, *SD* = .10) cue condition compared to the no cue condition (*M* = .81, *SD* = .10), and novices correctly categorized more images in the prototypical (*M* = .81, *SD* = .11) and nonprototypical (*M* = .83, *SD* = .08) cue conditions compared to the no cue condition (*M* = .74, *SD* = .12). There were no significant differences between experts and novices in the prototypical and nonprototypical cue conditions ($p > .05$).

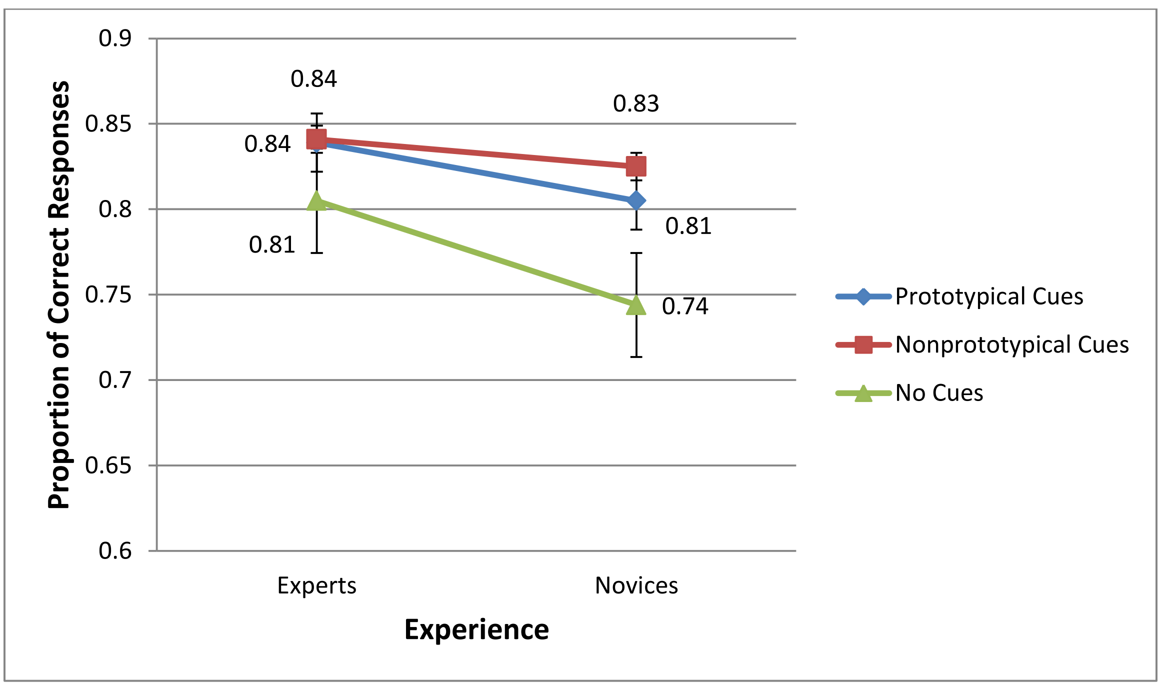

Figure 5. Mean proportion of correct responses for each cue condition as a function of experience.

Response Times

The results for response times are shown in Table 4. Levene's tests were used to check for equality of variance and the results concluded that there was adequate homogeneity of variance. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the interaction between image type and cue condition, $\chi 2$ = 6.59, $p < .05$. A Greenhouse-Geisser correction was used for this effect. For post hoc tests, the Bonferroni Sidak test was used.

Table 4

Results of the Analysis of Variance for Response Times

Source	SS	df	MS	F	*p*	partial η^2
Between Subjects						
Experience (E)	40.62	1	40.62	5.78	.01	.10
Error	506.41	72	7.03			
Within Subjects						
Image (I)	26.52	1	26.52	31.46	.00	.30
I x E	1.95	1	1.95	2.32	.13	.03
Error	60.7	72	.84			
Condition (C)	7.85	2	3.93	2.42	.09	.03
C x E	0.03	2	.01	0.01	.99	.00
Error	233.86	144	1.62			
I x C	2.67	1.84	1.45	4.24	.02	.06
I x C x E	0.84	1.84	.46	1.33	.27	.02
Error	45.37	132.27	.34			

A significant main effect for experience was observed which shows that novices (*M* = 2.89, *SD* = 1.62) took significantly longer to respond to images compared to experts (*M* = 2.28, *SD* = .99). A significant main effect for image type was also observed. Both experts and novices took significantly longer to respond to the nonprototypical images (*M* = 2.83, *SD* = 1.58) compared to the prototypical images (*M* = 2.34, *SD* = 1.23). A significant interaction for image type and cue condition was observed. A plot of the interaction is shown in Figure 6. Regarding prototypical images, a test of simple effects showed that participants took longer to respond with the nonprototypical cues (*M* = 2.47, *SD* = 1.24) than with the prototypical cues (*M* = 2.16, *SD* = 1.00). As for the nonprototypical images, participants took longer to respond with the nonprototypical cues (*M* = 3.07, *SD* = 1.74) than they did without cues (*M* = 2.66, *SD* = 1.55). A test of simple effects also showed that the difference between prototypical and nonprototypical images was significant in all three cue conditions, with slower responses to the nonprototypical images. No other effects were significant.

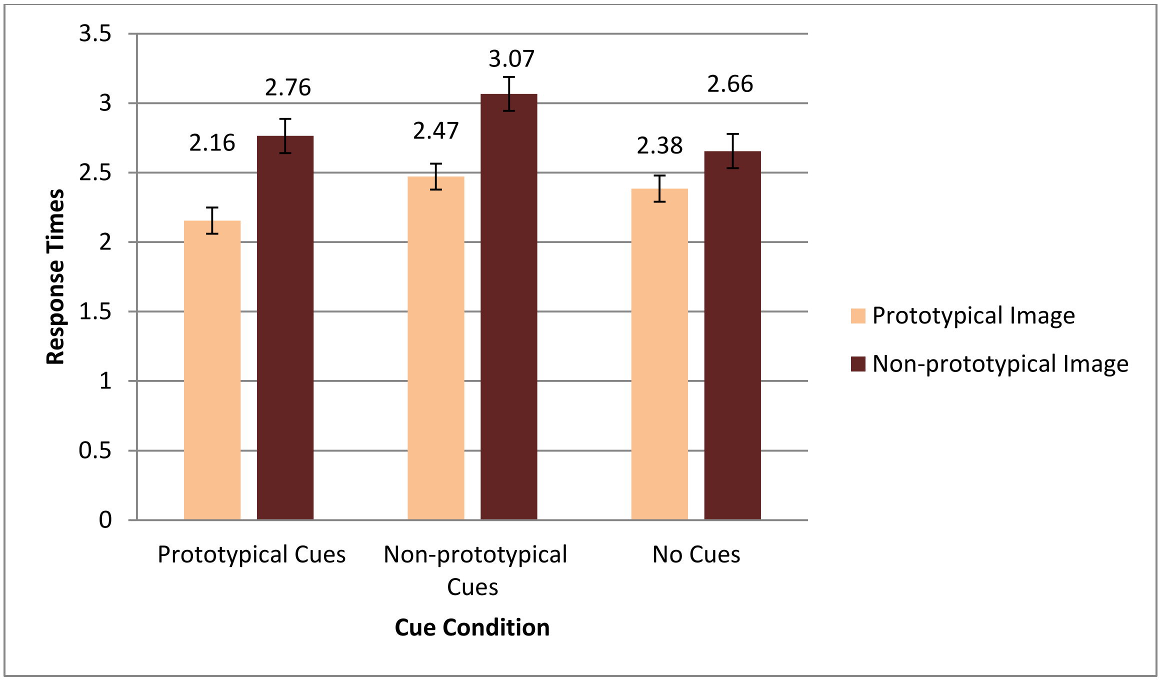

Figure 6. Mean response times (sec) for image type as a function of cue condition.

DISCUSSION

The purpose of the present study was to examine how well individuals discriminate between categories of FHR variability. There were four categories of FHR variability that participants needed to classify: absent, minimal, moderate and marked. Participants were instructed to identify the category in which each example image of a FHR tracing belonged with and without the presence of alignable cues. Participants took part in three conditions: no cue, prototypical cue, and nonprototypical cue. The prototypical cues were four images that represented the prototype of each of the four FHR variability categories. The nonprototypical cues were five images that represented the boundaries of each of the four FHR variability categories. The participants in the present study consisted of novices and experts in order to examine whether experience affects the way the FHR variability examples are categorized.

Image Type

The first goal was to assess whether classification of FHR variability is more difficult as the examples deviate further from the prototype for each category. Previous research suggests that stimuli falling near the category boundaries have ambiguous category membership, making categorization of the stimuli slower compared to stimuli far from the category boundaries, where categorization is an easier and faster process (Ashby et al., 1994; Das-Smaal & De Swart, 1984, 1986). Therefore, it was anticipated that classification accuracy would be higher and response times shorter for the prototypical examples compared to the nonprototypical examples. The results from the present study supported the first hypothesis. The results show that both experts and novices correctly categorized significantly more prototypical images compared to the nonprototypical images, which supports other findings and suggests that examples of stimuli that are similar to their prototypes are easier to categorize and are learned quicker (Das-Smaal & De Swart, 1984, 1986; Vandierendonck, 1984). The results also show that participants took longer to respond to the nonprototypical images as compared to the prototypical FHR images. These results support numerous studies that have found that examples of stimuli that resemble their prototypes are responded to faster (Ashby et al., 1994; Das-Smaal & De Swart, 1984, 1986). Theories of categorization suggest that individuals tend to judge the similarity of stimuli to group them into categories (Cohen & Basu, 1987; Estes, 1986; Johansen & Palmeri, 2002; Smith & Minda, 1998). Previous research has found that individuals' categorization performance is consistent with prototype-based possessing early on in learning (Smith & Minda, 1998). In prototype models of categorization, individuals learn to categorize examples by comparing new examples to an average example of each category (i.e., a prototype; Estes, 1986). The novices in the present study were given a prototypical example image of each of the four FHR variability categories, followed by practice images to categorize before they started the experiment. The novices were required to categorize each practice image before they could start the experiment to ensure familiarity with the categories. Thus, because the novices were able to correctly categorize more prototypical FHR images compared to the nonprototypical images, the results of the present study suggest that novices may have made their judgments based on memory by comparing each FHR example image to the prototype of each category.

The results of the present study also reveal that the experts correctly categorized more FHR images than the novices in all conditions; however, the experts correctly categorized significantly fewer nonprototypical images compared to prototypical images in all three conditions. Based on previous research, which suggests that learning with high exemplar diversity can aid the ability to generalize to novel stimuli less typical of the category prototype (Das-Smaal and De Swart, 1984; 1986; Hahn et al., 2005), it was expected that the experienced clinicians would be able to focus on the relevant details of the images and correctly categorize more nonprototypical images in the no cue and prototypical cue condition (Balslev et al., 2012; Klein et al., 1993; Lowe, 1999). The experts were already familiar with the stimuli and had the advantage of years of learning from multiple examples; thus, they should have been able to organize the FHR variability categories around examples acquired through years of experience and outperform the novices at categorizing nonprototypical images. Therefore, the results suggest that categorization of the FHR images may require more perceptual processing, guided by how similar each example is to the learned prototype of each category, making it difficult to categorize nonprototypical images when needing to rely on memory.

Cue vs. No Cues

The second goal of this study was to examine categorization performance when given the opportunity to use exemplars. Previous research suggests that cues are able to direct attention to relevant information and help a person respond faster and make fewer errors when detecting stimuli in visual tasks (Chaney & Teel, 1967; Hall et al., 2012; Loft et al., 2013; Kok et al., 2012; Kurtz & Gentner, 2013; Gunn, et al., 2005; Posner, et al., 1978; Posner, et al., 1980; Wickens & Hollands, 1999). It was expected that the presence of the exemplars would lead to faster response times and better accuracy. The results of the current study supported the second hypothesis and revealed that novices and experts correctly categorized significantly more FHR variability images when given the opportunity to use the exemplars. The results also revealed that experts and novices correctly categorized more prototypical images in the nonprototypical cue condition and more nonprototypical images in the prototypical cue condition. According to structural alignment theory, the participants were able to use the alignable (high-similarity) cues which helped them to compare the FHR images to the exemplars, or cues, and detect subtle differences (Markman & Gentner, 1997). By comparing the features of the FHR examples and the cues, the participants were able to isolate relevant information in the ambiguous stimuli, which made the differences become more salient (Gentner & Gunn, 2001; Kok et al., 2013; Kurtz & Gentner, 2013). Thus, by making differences between the image and cues more salient as a result of this matching process, discriminating relevant information became easier allowing the participants to correctly identify the categories to which the images belonged.

The results also showed that participants took longer to respond to the prototypical images in the nonprototypical cue condition compared to the prototypical cue condition, and participants took longer to respond to the nonprototypical images in the nonprototypical cue condition compared to the no aid condition. One possible explanation for the longer response times in the nonprototypical cue conditions could be that the exemplars of the category boundaries were harder to discriminate from one another because of how similar they looked, requiring more attention, and time, to match the image to the appropriate exemplar (Ashby et al., 1994; Maddox & Ashby, 1993).

Prototypical vs. Nonprototypical Cues

The results showed that all participants correctly categorized significantly more images in both the prototypical and nonprototypical cue conditions compared to the no cue condition. There was also a significant interaction for image type and cue condition. More specifically, for the prototypical images, participants categorized more images in the nonprototypical cue condition compared to both the prototypical and no cue conditions, and participants correctly categorized more images in the prototypical cue condition compared to the no cue condition. For the nonprototypical images, the results showed that participants correctly categorized more images in the prototypical and nonprototypical cue condition compared to the no cue condition. Thus, the experts and novices were able to benefit from the cues by using them as an aid for pattern matching. The alignable cues allowed participants to compare the example to the relevant information in the cues and detect the less conspicuous information in order to categorize the images correctly (Gentner & Gunn, 2001; Markman & Gentner, 1997). The results also showed that experts correctly categorized significantly more images in the nonprototypical cue condition compared to the no cue condition, and novices correctly categorized significantly more images in the prototypical and nonprototypical cue conditions compared to the no cue condition. These results support previous research in which cues improved performance when categorizing MFHR images (Kennedy, Anderson-Montoya, Scerbo, Prytz, Belfore, Abuhamad, Davis, & Chauhan, 2012; Hall et al., 2012), and suggest that the categorization of the FHR variability could be accomplished by pattern recognition (Lowe, 1985; Wertheimer, 1938b). Both clinicians and novices were able to categorize the examples by matching them to the cues;

therefore, when the cues were present, the clinicians had no appreciable advantage over the novices. Moreover, the results showed that when given the prototypical cues, the novices were able to perform just as well as the experts that categorized FHR variability without the cues. Thus, the current study demonstrated that not only do cues aid a participant's performance, but the alignable cues, which were highly similar examples of the simulated FHR images, used in this study are beneficial to both novices and experts when categorizing MFHR variability.

Experience

The third goal of this study was to examine how clinicians and novices differ when categorizing FHR variability with and without exemplars. Previous research has found that experts are better than novices at noticing relevant information in discrimination tasks (Klein & Hoffman, 1993; Myles-Worsley et al., 1988). Moreover, people who have more experience with specific stimuli are better able to generalize new, moderately-similar stimuli and categorize them correctly (Palmeri, 1997; Hahn et al., 2005). Therefore, it was proposed that experienced clinicians would be more adept at generalizing to nonprototypical examples of the FHR variability categories and perform significantly better than novices when making judgments, particularly when no cues were available. However, differences between experts and novices could be modified by the presence of cues. When cues are available, the categorization of FHR variability could be accomplished purely by perceptual pattern matching allowing both clinicians and novices to perceptually categorize the examples by matching them to the stimuli (Estes, 1986; Rosch & Mervis, 1975). The alternative hypothesis was that clinicians would have no advantage over novices when cues were available. The results partially supported the third hypothesis and revealed a significant interaction for experience and cue condition, more specifically, experts correctly categorized more images and in less time in the no cue condition compared to novices. The experts could draw upon their extensive experience with the ambiguous and vague FHR tracings, and therefore, were more adept at making their decisions. However, the presence of cues did indeed impact performance. The advantage of experts over novices in the no cue condition was eliminated when cues were available. Thus, the availability of cues enabled novices to make up for their lack of experience and make categorization decisions comparable to those of clinicians.

Although the experts correctly categorized more images compared to the novices in all conditions, the experts did not perform significantly better than the novices in either of the cue conditions. Thus, the results suggest that for the experts and novices alike, categorization of the FHR images can be facilitated by a perceptual matching process that is guided by how similar each example is to the learned prototype of each category and then influenced by how similar each example is to the cues.

The final goal of the present study was to examine whether the alignable cues would be beneficial to the experts. A 3-way interaction among prototypicality of the examples, the cueing conditions, and experience was expected. It was hypothesized that there might be a benefit of alignable cues for experts, but limited to FHR examples that are at the boundaries of each category and not the prototype of each category. However, the fourth hypothesis was partially supported because no interaction was detected, suggesting that even though the experts have more experience with the ambiguous stimuli, categorizing nonprototypical examples of FHR variability is still a difficult task. Although the results did not show the expected interaction, the results did show that experts correctly categorized significantly more images in the nonprototypical cue condition compared to the no cue condition. Therefore, the experts benefited from the cues on both prototypical and nonprototypical images.

The results of the present study suggest that the experts were able to make their decisions using an internalized prototype of each FHR variability category acquired over their years of experience when no cues were available; however, categorizing nonprototypical examples was still a difficult task. The experts were less able to generalize to nonprototypical images, suggesting that when the images are farther away from the prototype of each category and closer to the boundaries, the task becomes more perceptual. In the cue conditions, the experts may have benefited from their experience, but the additional information provided by the cues further aided their performance by allowing them to pattern match. Furthermore, novices, who have no experience with FHR tracings, benefitted from both prototypical and nonprototypical cues and were able to perform at the level of experts with no cues.

Limitations and Future Research

The goal of the present study was to examine theories of categorization to predict how individuals would classify different levels of FHR variability. Previous research has indicated that the placement of a cue is important when comparing stimuli against a comparison standard (Kurtz & Gentner, 2013). A limitation was that the cues used in the present study were fixed and appeared under the FHR images. If the participants were able to move the cues and place them nearer or on top of the FHR images, performance may have been improved because structurally aligning the two figures would render the differences more salient (Gentner, & Gunn, 2001). Therefore, subsequent studies should be designed to allow more control over cue placement. Also, the participants in the present study were presented with sets of static images of FHR tracings. However, this paradigm differs from standard clinical practice. When maternity patients are monitored by EFM, the clinicians inspect dynamic tracings of the FHR. Therefore, the task of categorizing static images of FHR variability did not resemble the actual dynamic task clinicians perform on a regular basis in labor and delivery units because the participants had a limited view (5 minutes of a tracing), whereas in hospitals, clinicians have an opportunity to inspect hours' worth of data. By narrowing and restricting the view, participants were not given the opportunity to use the additional information available in a longer FHR tracing. For instance, it may be possible to benefit from looking at a longer tracing because there is more context available for detecting changes in the variability. Another potential limitation was that the experts differed in their level of experience. The experts consisted of residents, nurses, and midwifes. A closer look at performance within the expert group showed that the residents ($M = .80$, $SD = .10$) correctly categorized significantly more nonprototypical images compared to the nurses ($M = .72$, $SD = .10$). Future studies should look at how the training differs between residents and nurses, because nurses are often the first to inspect the FHR tracings before they are given to doctors for further examination. Furthermore, it would also be beneficial to look at nurses solely because they may benefit more from the cues.

Another goal of the present study was to examine the benefits of providing cues to aid in categorization. Participants benefited more from the nonprototypical cues when categorizing prototypical example images and benefited more from the prototypical cues when categorizing the nonprototypical example images. This suggests that individuals may have been comparing and contrasting the differences between the cues and examples in order to detect relevant information. More research is needed to examine whether participants would benefit from a combination of both prototypical and nonprototypical cues. A combination of both types of cues could help to further distinguish the relevant information by providing contrasting sets of exemplars (Gentner & Gunn, 2001). Furthermore, because comparison-based learning/training has been shown to lead to improved detection performance of non-obvious anomalies in subsequent novel examples (Kok et al., 2013; Kurtz & Gentner, 2013), it would be beneficial to give clinicians the cues and examine whether using the cues while inspecting FHR tracings affects the outcome of how they treat their patients.

Theoretical and Clinical Implications

One of the main goals of the current study was to assess how well individuals categorize FHR variability with the aid of a visual cue. The results suggest that categorizing FHR variability is a difficult task because FHR tracings are vague and ambiguous stimuli; clinicians may employ different criteria for distinguishing categories of FHR variability. Based on theories of category learning (Das Smaal & De Swart, 1984; 1986; Hahn et al., 2005; Palmeri, 1997; Vandierendonck, 1984), research suggests that people who have a broad range of experience with categorizing ambiguous stimuli, such as clinicians in labor and delivery units, should be able to generalize to new stimuli and categorize them more reliably. Furthermore, theories of expertise suggest that experts who have a broad range of experience with specific stimuli should be able to detect the less-conspicuous and relevant information in order to distinguish one stimulus from another (Balslev et al., 2010; Jarodzka et al., 2010; Klein & Hoffman, 1993; Myles-

Worsley et al., 1988). However, the results of the present study reveal that the experts categorized significantly fewer nonprototypical compared to prototypical FHR images and only correctly categorized about 81% of the FHR images when no cues were available. This suggests that the clinicians may have an internalized prototype of each variability category, however, it doesn't help them when the examples are at the border of each category, where categorization is more difficult. Moreover, theories of expertise suggest that clinicians should be able to outperform novices who have not had as much training and experience with the stimuli. However, the results of the present study reveal that the novices, when presented with cues, were able to perform just as well as the experts with no cues. The results suggest that some FHR images may just be too ambiguous for clinicians to rely solely on memory for making category decisions. Thus, even experts could use additional cues in order to reliably categorize FHR variability since performance improved when an alignable cue were present.

In a clinical setting, the failure to make the distinction between the different types of variability may lead to improper care of the patient. More specifically, clinicians must be able to judge the degree of variability in the tracing in order to correctly determine to which category of the three-tier FHR interpretation system the tracing belongs. If the variability is categorized incorrectly then there is a risk of misdiagnosis and unnecessary intervention, such as performing a cesarean section. Therefore, the presence of a visual cue can improve clinicians' abilities to properly categorize FHR variability when the tracings are prototypical examples of the categories, and may provide a benefit when the FHR variability is not prototypical.

CONCLUSION

The main goal of this study was to examine the effect of cues on the ability of individuals to categorize prototypical and nonprototypical examples of FHR tracings and also to examine whether there were differences due to experience. Although experts performed better on categorizing the images and responded faster, the cues increased performance for both the experts and novices. Both image type and the presence of cues had a significant impact on performance. The current study suggests that the clinicians could benefit from training with alignable cues on categorizing FHR variability, which could improve later clinical decision making when monitoring FHR tracings to enhance the safety in labor and delivery.

REFERENCES

American College of Obstetricians and Gynecologists. (2010). Practice bulletin number 116: Management of intrapartum fetal heart rate tracings. *Obstetrics & Gynecology*, *116*, 1232–1240.

Ayres-de-Campos, D., Bernardes, J., Costa-Pereira, A., & Pereira-Leite, L. (1999). Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. *British Journal of Obstetrics and Gynaecology, 106,* 1307–1310.

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, *55*, 11-27.

Bailey, R. E. (2009). Intrapartum Fetal Monitoring. *American Family Physician, 80,* 1388-1396.

Barstow, C., Gauer, R., & Jamieson, B. (2010). How does electronic fetal heart rate monitoring affect labor and delivery outcomes? *Journal of Family Practice, 59,* 653a-653b.

Belfore II, L. A., Scerbo, M. W., & Anderson, B. L. (2007, September). A fetal heart rate monitor simulator. ModSim World 2007, Virginia Beach, VA.

Berkus, M. D., Langer, O., Samueloff, A., Xenakis, E. M., & Field, N. T. (1999). Electronic fetal monitoring: what's reassuring? *Acta Obstetricia et Gynecologica Scandinavica*, *78*, 15–21.

Bernardes, J., Costa-Pereira, A., Ayres-de-Campos, D., van Geijn, H. P., & Pereira-Leite, L. (1997). Evaluation of interobserver agreement of cardiotocograms. *International Journal of Gynecology & Obstetrics*, *57*, 33–37.

Balslev, T., Jarodzka, H., Holmqvist, K., de Grave, W., Muijtjens, A. M., Eika, B., &

Scherpbier, J. (2012). Visual expertise in paediatric neurology. *European Journal of Paediatric Neurology, 16*, 161-166.

Chaney, F. B., & Teel, K. S. (1967). Improving inspector performance through training and visual aids. *Journal of Applied Psychology, 51*, 311-315.

Chez, B. F., Harvey, M. G., & Harvey, C. J. (2000). Intrapartum fetal monitoring: past, present, and future. *The Journal of Perinatal & Neonatal Nursing, 14*, 1-18.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cohen, J. B., & Basu, K. (1987). Alternative models of categorization: toward a contingent processing framework. *Journal of Consumer Research*, 455-472.

Devane, D. & Lalor, J. (2005). Midwives' visual interpretation of intrapartum cardiotocographs: intra- and inter-observer agreement. *Journal of Advanced Nursing, 52*, 133–141.

Das-Smaal, E. A., & De Swart, J. H. (1984). Variation within categories. *Acta psychologica*, *57*, 165-192.

Das-Smaal, E. A., & De Swart, J. H. (1986). Effects of contrasting category, conjoint frequency and typicality on categorization. *Acta psychologica*, *62*, 15-40.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18*, 500-549.

Fedorka, P. (2010). Electronic fetal monitoring: an update. *Journal of Legal Nurse Consulting*, *21*, 15-18.

Figueras, F., Albela, S., Bonino, M. P., Palacio, M., Barrau, E., Hernandez, S., Casellas, C., Coll, O., Cararach, V. (2005). Visual analysis of antrepartum fetal heart rate tracings: Inter- and intra-observer agreement and impact of knowledge of neonatal outcome. *Journal of Perinatal Medicine*, *33*, 241-245.

Freeman, R. K. (2002). Problems with intrapartum fetal heart rate monitoring interpretation and patient management. *Obstetrics & Gynecology*, *100*, 813–826.

Gagnon, R., Campbell, M. K., & Hunse, C. (1993). A comparison between visual and computer analysis of antepartum fetal heart rate tracings. *American Journal of Obstetrics & Gynecology*, *168*, 842–847.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183-209.

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition*, *29*, 565-577.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45-56

Gunn, D. V., Warm, J. S., Nelson, W. T., Bolia, R. S., Schumsky, D. A., & Corcoran, K. J. (2005). Target acquisition with UAVs: vigilance displays and advanced cuing interfaces. *Human Factors*, *47*, 488-97.

Hahn, U., Bailey, T. M., & Elvin, L. B. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*, 289-302.

Hall, L. M., Hannon, D. J., Leisk, G. G., Wolfberg, A. J., & House, M. D. (2012). Measurement of Fetal Heart Rate Variability on an Electronic Monitor Using a Prototype Electronic Ruler. *American Journal of Perinatology*, *29*, 409-414.

Hartmann, G. W. (1935). *Gestalt psychology: A survey of facts and principles*. New York: Ronald Press.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*, 146-154.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning?. *Cognitive Psychology*, *45*, 482-553.

Kennedy, R. A., Anderson-Montoya, B. L., Scerbo, M. W., Prytz, E., Belfore, L. A., Abuhamad. A. Z., Davis, S.S., & Chauhan, S.P. (2012). The influence of visual aids on detecting early and late decelerations in maternal-fetal heart rate patterns. *Proceedings of the Human Factors & Ergonomics Society 56th Annual Meeting*, 1317-1321.

Klein, G. A., & Hoffman, R. R. (1993). Seeing the invisible: Perceptual-cognitive aspects of expertise. In M. Rabinowitz (Ed.), *Cognitive Science Foundations of Instruction*. (pp. 203-226). Hillsdale, NJ: Erlbaum.

Kok, E. M., de Bruin, A. B., Robben, S. G., & van Merriënboer, J. J. (2013) Learning radiological appearances of diseases: does comparison help? *Learning and Instruction*, 90-97.

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison Promotes Learning and Transfer of Relational Categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1303-1310.

Kurtz, K. J., & Gentner, D. (2013). Detecting anomalous features in complex stimuli: The role of structured comparison. *Journal of Experimental Psychology: Applied*, *19*, 219.

Loft, S., Smith, R. E., & Remington, R. W. (2013). Minimizing the disruptive effects of prospective memory in simulated air traffic control. *Journal of Experimental Psychology: Applied*, *19*, 254.

Lowe, D. G. (1985). *Perceptual organization and visual recognition*. Boston, MA: Kluwer.

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education*, *14*, 225-244.

Macones, G. A., Hankins, G. D. V., Spong, C. Y., Hauth, J., Moore, T. (2008). The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines. *Obstetrics & Gynecology, 112*, 661–666.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*, 49-70.

Markman, A. B., & Gentner, D. (2000). Structure mapping in the comparison process. *American Journal of Psychology*, *113*, 501-538.

Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Menacker, F., Kirmeyer, S., & Munson, M. L. (2007). Births: final data for 2005. *National vital statistics reports, 56* (6), 1-103.

Miller, L. A. (2005). System errors in intrapartum electronic fetal monitoring: A case review. *Journal of Midwifery & Women's Health, 50*, 507-516.

Miller, D. A. (2011). Intrapartum fetal heart rate definitions and interpretation: evolving consensus. *Clinical obstetrics and gynecology*, *54,* 16-21.

Miller, D. A., & Miller, L. A. (2012). Electronic fetal heart rate monitoring: applying principles of patient safety. *American Journal of Obstetrics and Gynecology, 206,* 278-283.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254.

Menihan, C. A. & Zottoli, E. K. (2001). *Electronic Fetal Monitoring: Concepts and Applications*. Philadelphia: Lippincott.

Minkoff, H., & Berkowitz, R. (2009). Fetal monitoring bundle. *Obstetrics & Gynecology, 114*, 1332–1335.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 553.

Nelson, K.B., Dambrosia, J.M., Ting, T.Y. & Grether, J. K. (1996) Uncertain value of electronic fetal monitoring in predicting cerebral palsy. *New England Journal of Medicine, 334*, 613–618.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23,* 324.

Palmer, S. E. (1992). Common region: A new principle of perceptual organization. *Cognitive Psychology*, *24*, 436–447.

Palmer, S. E., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*, 29–55.

Paradis, R. D., Guo, J. K., Olden-Stahl, J., & Moulton, J. (2012). Cognitive Category Learning. *Procedia Computer Science*, *12*, 188-193.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. Pick & E. Saltzmann (Eds.), *Modes of perceiving and processing information* (pp. 137–157). Hillsdale, New Jersey: Lawrence Erlbaum.

Posner, M., I., Snyder, C., R., Davidson, B., J.(1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109,* 160-174.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Sisco, K. M., Cahill, A. G., Stamilio, D. M., & Macones, G. A. (2009). Is continuous monitoring the answer to incidentally observed fetal heart rate decelerations? *The Journal of Maternal-Fetal and Neonatal Medicine*, *22*, 405–409.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411.

Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language*, *42*, 51-73.

Sweha, A., Hacker,T. W., & Nuovo, J. (1999). Interpretation of the electronic fetal heart rate during labor. *American Family Physician, 59*, 2487-2500.

Vandierendonck, A. (1984). Concept learning, memory and transfer of dot pattern categories. *Acta Psychologica*, *55*, 71-83.

Verheyen, S., & Storms, G. (2013). A Mixture Approach to Vagueness and Ambiguity. *PloS one*, *8*, 1-13.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, *138*, 1218–1252.

Wertheimer , M. (1938a). Gestalt theory. In W. D. Ellis (Ed.), *A Source Book of Gestalt Psychology* (pp. 1-11). London, England: Kegan Paul, Trench, Trubner & Company.

Wertheimer, M. (1938b). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A Source Book of Gestalt Psychology* (pp. 71-88). Kegan Paul, Trench, Trubner & Company.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

APPENDIX

OLD DOMINION UNIVERSITY INFORMED CONSENT FORM

IRB Identifier:

Informed Consent Document
Old Dominion University

PROJECT TITLE: Inspecting Maternal-Fetal Tracings

RESEARCHERS:
Mark W. Scerbo, Ph.D., Responsible Project Investigator, Professor, College of Sciences, Psychology Department

Co-investigators:

Amanda Ashdown, Graduate Student, College of Science, Psychology Department

DESCRIPTION OF RESEARCH STUDY:
A maternal fetal heart rate (MFHR) monitor depicts a fetus's heartbeat in conjunction with maternal contractions. The purpose of this experiment is to study the ability of individuals to categorize MFHR variability.

You will be asked to monitor a computer screen displaying simulated MFHR signals for four types of variability. You will also be asked to fill out a background information form and perform a computer task. The entire session will last approximately 30 minutes. Approximately 49 students will participate in this study.

EXCLUSIONARY CRITERIA:
To participate in this study you must be an undergraduate student at ODU. You must be 18 years of age or older. You must also have normal or corrected-to-normal vision. If you wear contacts or glasses you must have these with you when you participate. You must not have participated in any previous MFHR studies.

RISKS:
Risks associated with this study are similar to those of normal computer usage, such as eye strain. However, there is still potential for risks that have not yet been identified.

BENEFITS:
There are no direct benefits for participation. However, you may potentially benefit from learning how psychological research is conducted, and learning about the types of MFHR signals that are monitored during labor and delivery.

COSTS AND PAYMENTS:
If you decide to participate in this study, you will receive 1 Psychology department research credit, which may be applied to course requirements or extra credit in certain Psychology

courses. Equivalent credits may be obtained in other ways. You do not have to participate in this study, or any Psychology Department study, in order to obtain this credit.

CONFIDENTIALITY:
Your results from this study will be kept confidential by the researcher. All identifiers that can link you to your responses will be removed. The results of this study may be used in reports, presentations, and publications; however, your identity will remain anonymous.

WITHDRAWAL PRIVILEGE:
If you wish to stop participating it is OK. At any point during the study you may state that you wish to withdraw and you will not be penalized.

COMPENSATION FOR ILLNESS AND INJURY:
If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of any harm, injury, or illnesses arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in any research project, you may contact Dr. Mark W. Scerbo (757) 683-4217 or Dr. George Maihafer, the current IRB chair at (757) 683-4520 at Old Dominion University, who will be glad to review the matter with you. The Office of Research may be contacted at any time (757) 683-3460.

VOLUNTARY CONSENT:
By signing this form, you are saying several things. You are saying that you have read this form or have had it read to you, that you are satisfied that you understand this form, the research study, and its risks and benefits. The researchers should have answered any questions you may have had about this research. If you have any questions later on, then the researchers should be able to answer them: Dr. Mark W. Scerbo (757) 683-4217.

If at any time you feel pressured to participate, or if you have any questions about your rights or this form, then you should call Dr. George Maihafer, the current IRB chair, at (757) 683-4520, or the Old Dominion University Office of Research, at 757-683-3460.

And importantly, by signing below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

------------------------------------	-----------------------------------	------------
Participant's Name	Participant's Signature	Date

<u>INVESTIGATOR'S STATEMENT</u>

I certify that I have explained to this subject the nature and purpose of this research, including benefits, risks, costs, and any experimental procedures. I have described the rights and protections afforded to human subjects and have done nothing to pressure, coerce, or falsely entice this subject into participating. I am aware of my obligations under state and federal laws, and promise compliance. I have answered the subject's questions and have encouraged him/her to ask additional questions at any time during the course of this study. I have witnessed the above signature(s) on this consent form.

------------------------------------ ----------------------------------- ---------
---
Investigator’s Name Investigator’s Signature Date

VITA

Amanda Ashdown

Phone: (757) 8316177
Email: aashd001@odu.edu, aashdown@odu.edu

EDUCATION

2008 - 2012	B.S. - Psychology Major- Psychology Minor- Human Services Old Dominion University Psychology Department Norfolk, VA

PROFESSIONAL EXPERIENCE

2010 – Present	Graduate Research Assistant Supervisor: Mark W. Scerbo, Ph.D. Old Dominion University Psychology Department Norfolk, VA 23529
2013-Present	Graduate Teaching Assistant Old Dominion University Psychology Department Norfolk, VA 23529

TECHNICAL REPORTS, PROCEEDINGS ARTICALS, HONOR'S THESIS

Prytz, E., Anderson-Montoya, B., Kennedy, B., Montano, M., Ashdown, A., Warvel, L., & Scerbo, M. (2012). *Virtual I.V. Self-Directed Learning System*. In Parodi, A. et al. (Eds.) LIVES Lab: Process Development for a Conceptual Framework Driven, Process and Product Analysis Laboratory for Medical & Healthcare Simulators and Simulations, *Technical report, Virginia Modeling, Simulation, Analysis, and Simulation Center.*

Ashdown, A., Young, S. (2014). Airplane System Awareness. Unpublished technical report, NASA Langley, Hampton, VA.

Ashdown, A., Scerbo, M. W., Anderson, B. L., Belfore II, L. A., & Abuhamad, A. Z (2014). User Placement of a Visual Aid for Detecting Critical Signals in Fetal-Heart Rate Tracings: A Yoke-Control Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

Ashdown, A. J. (2012). Detecting Critical Patterns in Maternal-Fetal Heart Rate Signals with Visual Aids. Unpublished master's thesis, Old Dominion University, Norfolk, VA.

PROFESSIONAL AFFILIATIONS

Human Factors and Ergonomics Society (HFES) Student Affiliate,
ODU HFES Student Chapter, Vice President 2014-2015, President 2015-2016