

Quantifying Errors of Bias and Discriminability Emitted by Children during a
Matching-to-Sample Task

by

Courtney Hannula

A thesis submitted to the College of Psychology and Liberal Arts of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science

in

Applied Behavior Analysis and Organizational Behavior Management

Melbourne, Florida

July, 2018

We the undersigned committee hereby approve the attached thesis, "Quantifying Errors of Bias and Discriminability Emitted by Children During a Matching-to-Sample Task," by Courtney Hannula.

Christopher Podlesnik, Ph.D., BCBA-D
Associate Professor
School of Behavior Analysis

Corina Jimenez-Gomez, Ph.D., BCBA-D
Research Assistant Professor
School of Behavior Analysis

Adam T. Brewer, Ph.D., BCBA-D
Assistant Professor
School of Behavior Analysis

Darby Proctor, Ph.D.
Assistant Professor
School of Psychology

Lisa Steelman, Ph.D.
Senior Associate Dean
College of Psychology and Liberal Arts

Abstract

Title: Quantifying Errors of Bias and Discriminability Emitted by Children During a Matching-to-Sample Task

Author: Courtney Hannula

Advisor: Christopher A. Podlesnik, Ph.D., BCBA-D

Children diagnosed with Autism Spectrum Disorders (ASD) make errors during discrimination training regardless of antecedent or consequent procedures implemented to decrease errors. Further, these interventions are not guided by the source of errors. Two equations from Davison and Tustin's (1978) framework can quantify errors due to bias and discriminability, known as $\log b$ and $\log d$, respectively. This framework categorized errors emitted by children diagnosed with ASD during a matching-to-sample task. The task was displayed on a touchscreen device in which touching a sample stimulus at the beginning of each trial resulted in the appearance of two comparison stimuli. Researchers delivered reinforcement for touching the matching comparison stimulus. More similar sample stimuli were introduced during Phase 2 while keeping the comparison stimuli the same which affected sample discriminability only with little effect on biases for two of three participants. This framework accurately categorized errors emitted by children with ASD when levels of difficulty between the sample stimuli were manipulated. Future research might be able to use these equations to better categorize errors children with ASD exhibit during conditional discriminations. Future research

might also be able to improve teaching procedures by targeting interventions to mitigate or eliminate specific errors due to biases or reduced discriminability.

Keywords: discriminability, bias, conditional discriminations, matching-to-sample, children, autism spectrum disorder, errors, translational

Table of Contents

Table of Contents	v
List of Figures	vi
Acknowledgement	vii
Dedication	viii
Quantifying Errors of Bias and Discriminability Emitted by Children during a Matching-to-Sample Task	1
Method	16
Participants	16
Setting and Materials	17
Response Definition and Measurement	17
Procedures	20
Results	25
Discussion	30
Varied Responding as a Result of Difficult Discrimination	31
High Discriminability Yields Less Accurate Quantifications	31
Limitations	33
Implications for Clinical Practice and Research	34
References	36
Appendix	45

List of Figures

Figure 1.	48
Figure 2.	49
Figure 3.	50
Figure 4.	51
Figure 5.	52
Figure 6.	53

Acknowledgement

I would like to thank my supervisor, Dr. Christopher A. Podlesnik, for all his guidance and support during this process. Dr. Podlesnik, you showed me how writing is a process and it is okay and important to work through that process. You helped me to improve my writing skills and were always there to help. You taught me how to think critically about behavior analysis and science in general. Thank you so much for your quick edits and guidance throughout my time at FIT.

I would also like to thank Dr. Adam T. Brewer. Thank you so much for all of your help and encouragement while working on the paradigm programming for my thesis. I couldn't have done it without you. I am so grateful for the encouragement and praise you gave me throughout the whole process. Thanks for always checking in and keeping me going.

Thank you, Dr. Corina Jimenez-Gomez for always being so excited to see data in lab and for all the helpful discussions and comments throughout the entire process.

Thank you, Dr. Darby Proctor, for all of your helpful procedural comments for my methods and edits on my proposal and my defense. Your time and insight was valuable to my final product.

A special thanks to Dr. Blake A. Hutsell for his help with the data analysis and collaboration throughout the whole process as well.

Last, but not least, I would like to thank Weizhi Wu for her help running participants. Thank you for always being willing to squeeze my thesis into your busy days. Your help is greatly appreciated!

Dedication

I dedicate my thesis to my parents who taught me to have a love for learning and always pushed me to be the best I can be.

Quantifying Errors of Bias and Discriminability Emitted by Children during a Matching-to-Sample Task

Children diagnosed with Autism Spectrum Disorder (ASD) have unique and challenging behaviors that can interfere with learning new skills. According to the DSM-5, ASD is described as “having persistent deficits in social communication and social interaction across multiple contexts, and restricted, repetitive patterns of behavior, interests, or activities exhibited during early development which cause clinically significant impairment in social, occupational, or other important areas of current functioning” (American Psychiatric Association, 2013). In other words, individuals with ASD often have communication deficits or repetitive patterns of behavior that impede learning. Early Intensive Behavioral Intervention (EIBI), based on the principles of Applied Behavior Analysis (ABA), provides extra support for teaching skills to individuals with ASD (Lovaas, 1987; Howard, Sparkman, Cohen, Green, & Stanislaw, 2005). ABA is the systematic application of interventions based on principles of behavior that help improve behaviors of social significance (Baer, Wolf & Risley, 1968). It focuses on using scientifically validated interventions to improve skills and decrease repetitive and restrictive behaviors that interfere with an individual’s quality of life and ability to learn (Carr

& Durand, 1985; Hagopian, Fisher, Sullivan, Acquisto, & LeBlanc, 1998). In addition, interventions based on ABA can also help individuals to learn a variety of skills pertaining to expressive and receptive language, conditional discriminations, social communication, and daily living (Lovaas, 1987).

Discrimination and matching skills have been reported as important foundational skills to teach children with ASD (Fisher, Pawich, Dickes, Paden, & Toussaint, 2014). These skills are components of many other skills, including academics, socializing, communicating with others, and engaging in self-care routines (Green, 2001). A simple discrimination involves three terms (Davison & Nevin, 1999): antecedent, response, and consequence. Discriminations are trained by reinforcing one response in the presence of one antecedent stimulus and extinguishing the same response in the presence of other antecedent stimuli. For example, in the presence of a ball, a therapist delivers reinforcement for the individual's vocal-verbal response "ball." In contrast, in the presence of a block, the therapist does not reinforce the individual's vocal-verbal response of "ball." Discrimination is said to occur when the response occurs reliably in the presence of the correct stimulus and not in the presence of other stimuli (Green, 2001).

Matching-to-sample is one procedure used to teach conditional discriminations. Each trial begins by presenting a sample stimulus. Depending on the specific protocol, an observing response may be required, such as touching the sample stimulus (Dube & McIlvane, 1999; Fisher, Kodak & Moore, 2007). To

receive reinforcer delivery, the participant would then need to select the comparison stimulus that matches or is related to the sample stimulus (Petursdottir & Aguilar, 2016). In basic research with pigeons on conditional-discrimination training, Jones and White (1992) used an apparatus with a center key and two side keys each with the ability to illuminate various colors. A trial began with the presentation of a red or green sample stimulus on the center key. A peck to the center key turned off the center key. The two side keys then illuminated to present the two comparison stimuli, one red and one green. A response to the key with the same color as the sample stimulus resulted in 3-s access to food, while a response to the key with a different color than the sample stimulus results in 3-s of chamber darkness and no food. The pigeons began responding most often on the matching (correct) comparison compared to the other (incorrect) comparison.

Gutowski and Stromer (2003) conducted an applied study arranging a matching-to-sample procedure to teach matching skills. Researchers used pictures of a dog, cat, and bee in a matching-to-sample task. For the purposes of simplicity, I will be discussing this procedure here and throughout the introduction in terms of only two stimuli, with pictures of a dog and cat. During each trial, one of the pictures (either dog or cat) was presented in the center as the sample stimulus. Touching the sample resulted in the appearance of one picture of a dog and one picture of a cat on either side. Selecting the picture that matched the sample (correct response) resulted in delivery of a pleasant auditory tone and a token to

exchange for coins for a vending machine while selecting the other picture (incorrect response) resulted in no reinforcement. Participants began responding more often on the matching (correct) comparison compared to the non-matching (incorrect) comparison.

Although there is extensive research on this matching-to-sample procedure, some individuals in both clinical and laboratory situations exhibit persistent errors during discrimination training. In basic laboratory situations with nonhuman primates, researchers have used antecedent and consequent manipulations to decrease errors occurring during conditional discrimination procedures. These included requiring multiple responses to the sample stimulus before the comparison stimuli are presented or implementing a few second time-out from reinforcement with a dark chamber after an incorrect response is emitted (Sidman, Cresson, & Willson-Morris, 1974; Sidman et al., 1982; Tomonaga, 1993; Truppa et al., 2011). Some also required the participant to re-do the same trial over again after an incorrect response until they emit a correct response (Da Silva Barros, De Faria Galvao, & McIlvane, 2002; Tomonaga, 1993; Truppa et al., 2011). All of these studies had some success increasing percent correct in order to move forward in their studies, but none of them reported they acquired objective information about types of errors occurring before implementing these procedures.

A large amount of research conducted with humans attempted to determine antecedent manipulations to mitigate errors. One antecedent manipulation is to

insert a differential observing response (DOR) to require the individual to attend more closely to the sample stimulus (e.g., Fisher et al., 2007). A second antecedent manipulation is to compare the effectiveness of starting with multiple stimuli versus progressively incorporating more stimuli into the array (Grow, Kodak, & Carr, 2014). This comparison manipulation attempted to evaluate if teaching all stimuli at one time or teaching one stimulus at a time resulted in fewer errors. A third antecedent manipulation is to compare the effectiveness of presenting the comparison stimuli before the sample stimuli or vice versa (Petursdottir & Aguilar, 2016). Overall, these manipulations can increase accuracy for some individuals but are generally idiosyncratic across participants (Doughty & Hopkins, 2011; Dube et al., 2010; Dube & McIlvane, 1999; Fisher et al., 2007; Grow et al., 2014; Petursdottir & Aguilar, 2016).

Researchers also evaluated a variety of strategies to attempt to decrease errors made during conditional discrimination training by manipulating consequences. Most research evaluating consequence manipulations to decrease errors compared different error-correction procedures for individuals with ASD (Carroll, Joachim, St. Peter, & Robinson, 2015; Kodak et al., 2016; Rodgers & Iwata 1991; Smith, Mruzek, Wheat, & Hughes, 2006; Townley-Cochran, Leaf, Leaf, Taubman, & McEachin, 2017). One example is comparing an error correction that results in the therapist presenting a model of the correct response contingent on an error response versus presenting a model of the correct response with a

requirement that the child repeats the correct response. A second, related consequence manipulation is to examine rapid assessments to identify which error correction to use before starting teaching (Kodak et al., 2016; McGhan & Lerman, 2013). These studies determined which standard error-correction procedure was more efficient for each individual based on the number of trials/sessions required to reach a mastery criterion and client preference. As with antecedent manipulations, consequence manipulations tend to decrease errors with idiosyncratic results across participants.

Although the purpose of most research on procedural manipulations is to determine the most effective and efficient procedure to reduce errors, these studies do not identify why these errors occur. Attempting to change behavior without understanding the cause of the behavior generally is inconsistent with the approach of applied behavior analytic research. For example, the treatment of problem behavior (e.g., aggression, self-injury) first necessitates an assessment of the consequences maintaining problem behavior (e.g., access to attention, escape from demanding tasks such as school work; Iwata et al., 1982/1994). Once the consequences maintaining problem behavior are identified using functional analysis (Beavers, Iwata, & Lerman, 2013; Hagopian, Rooker, Jessel, DeLeon, 2013; Hanley, Iwata, & McCord, 2003; Northrup et al., 1991), appropriate behavior can then be trained and maintained by the same consequence previously maintaining problem behavior (e.g., Hanley et al., 2003; Iwata, Dorsey, Slifer,

Bauman, & Richman, 1994). Thus, functional analysis provides an assessment tool to identify the variables maintaining problem behavior so appropriate behavior can effectively substitute for problem behavior. In the context of conditional discriminations, implementing antecedent and consequence manipulations to mitigate errors, without knowledge of the origin of the errors, ignores the source of errors and potentially limits the effectiveness of these training procedures.

Davison and Tustin (1978) proposed a quantitative framework to identify the source of errors exhibited during conditional discriminations (see also Davison & Nevin, 1999). During conditional discriminations, one source of errors can occur due to confusion (1) between the antecedent stimuli and responding to the comparison stimuli (i.e., stimulus-behavior relations) or (2) between responding to the comparison stimuli and reinforcer delivery (i.e., behavior-reinforcer relations). These are errors due to imperfect *discriminability*. Another source of errors during conditional discriminations can occur due to inherent preference for either one comparison stimulus (e.g., cat versus dog) or one comparison location (e.g., left versus right). These are errors due to *bias*. The potential benefit of using Davison and Tustin's quantitative framework is it could allow researchers to quantify the different sources of errors contributing to poor conditional-discrimination performance.

Errors due to discriminability of stimulus-behavior relations result from confusion between samples. Accuracy may be lower if discriminability between the

sample stimuli is low. For example, Davison and McCarthy (1987) trained pigeons to peck the right key to access reinforcer delivery if the center key illuminated for 5 s and to peck the left key to access reinforcer delivery if the center key illuminated for twelve other durations ranging from 2.5 s to 57.5 s. Accuracy was lower when the difference between sample durations on the center key was small (e.g. 5 s vs. 7.5 s) than when greater (e.g., 5 s vs 57.5 s). Another example is with the matching-to-sample procedure with the dog and cat sample stimuli mentioned previously (Gutowski & Stromer, 2003). The dog and cat pictures might both have the same color fur or look similar in shape, which could lead to low discriminability between the two samples. Thus, more distinct sample stimuli produce greater accuracy due to greater stimulus-behavior discriminability (see also Davison & Nevin, 1999).

In contrast, errors due to discriminability of behavior-reinforcer relations result from confusion between the comparison stimuli and the following reinforcer deliveries. Accuracy could be lower if discriminability between the behavior-reinforcer relations is low. When there is lower discriminability due to reinforcement schedules being less distinct, accurate responses on the correct comparison stimuli could decrease and responding on the incorrect stimulus could increase. Peterson, Wheeler, and Trapold (1980) trained pigeons to match a green sample with vertical lines and a red sample with horizontal lines. A group of pigeons receiving food plus a tone for one of the correct responses and tone alone for the other correct responses performed more accurately than another group of

pigeons receiving food plus tone for both correct responses. Thus, more distinct consequences produced greater accuracy due to greater behavior-reinforcer discriminability (see also Davison & Nevin, 1999).

According to Davison and Tustin (1978), discriminability can be quantified by the equation, $\log d$, as shown below:

$$\log d = .5 \log \left[\left(\frac{C_1}{E_1} \right) \left(\frac{C_2}{E_2} \right) \right] \quad (1),$$

where C_1 and E_1 refer to correct and error responses, respectively, to the comparison stimulus (e.g. selecting dog picture or cat picture) following the first sample stimulus (S1; e.g. dog picture). Similarly, C_2 and E_2 refer to the correct and error responses to the comparison stimulus (e.g. selecting cat picture or dog picture) following the second sample stimulus (S2; e.g. cat picture) in a matching-to-sample task (Hutsell & Banks, 2015; Shahan & Podlesnik, 2006, 2007). $\log d$ shows any preference for the correct response that is independent of the reinforcer distribution or any biases the individual exhibits (Alsop & Rowley, 1996; see Appendix). Values of $\log d$ range from -1 to infinity and were used in the present investigation to quantify discriminability between the sample stimuli (i.e., stimulus-behavior discriminability) while reinforcer conditions were held constant (i.e., behavior-reinforcer discriminability). For an example with the dog and cat samples from Gutowski and Stromer (2003), the picture of the dog versus cat could be difficult to discriminate. This would result in a low value of $\log d$, (e.g. values near zero)

indicating less discriminability due to responding on the correct stimulus occurring less frequently. In contrast, the picture of the dog and the picture of the cat being more discriminable would result in a larger value of $\log d$, (e.g. values closer to one) indicating greater discriminability due to responding on the correct stimulus occurring more frequently.

According to Davison and Tustin (1978), a second source of errors is bias. Bias occurs from responding on one comparison stimulus or location more often than the other comparison or location, despite rate of reinforcement being equal for selecting either comparison after its corresponding sample stimulus (see also Davison & Nevin, 1999). Errors due to bias might occur even when there is high discriminability between the stimulus-behavior and behavior-reinforcer relations. Cumming and Berryman (1961) documented bias for both stimulus color and stimulus location in pigeons. The pigeons were trained to peck the comparison key with the same color as the sample key to obtain reinforcer delivery. All pigeons exhibited more frequent responding on one comparison location regardless of where the correct comparison stimulus appeared. As they experienced more sessions, a few pigeons also began to exhibit a stimulus bias. Despite equal reinforcement rates, several pigeons began responding more often on a particular color (e.g. blue, red, or green) comparison more often than the others.

According to Davison and Tustin (1978), bias can be quantified by the equation, $\log b$, which measures any preference for one of the comparison stimuli

independent from the reinforcer distributions. Bias for a particular comparison stimulus (e.g. dog vs. cat) or for a particular location (e.g. left vs. right) can occur.

Bias for a particular comparison is shown as,

$$\log b = .5 \log \left[\left(\frac{C_1}{C_2} \right) \left(\frac{E_1}{E_2} \right) \right] \quad (2),$$

where all terms appear as in Equation 1. In contrast, bias for a particular location is shown as,

$$\log b = .5 \log \left[\left(\frac{C_{right}}{C_{left}} \right) \left(\frac{E_{right}}{E_{left}} \right) \right] \quad (3),$$

where C_{right} and C_{left} refer to the correct responses to the comparison stimulus when it is on the right or left side, while E_{right} and E_{left} refer to the error responses to the comparison stimulus on the right or left side in a matching-to-sample task (Jones & White, 1992).

In the matching-to-sample example with the dog and cat pictures (Gutowski & Stromer, 2003), C_1 and E_1 refer to selecting the dog picture (correct) or selecting the cat picture (error) following the presentation of the dog picture. C_2 and E_2 refer to selecting the cat picture (correct) or selecting the dog picture (error) after the presentation of the cat picture. C_{right} and C_{left} refer to responding to the correct comparison (e.g. dog after display of dog or cat after display of cat) when it is on the right side or the left side, respectively. E_{right} and E_{left} refer to selecting the error comparison (e.g. cat after display of dog or dog after display of cat) when it is on the right or left side, respectively. These two equations measure the tendency for

the individual to emit more responses to one comparison or location in relation to the other comparison or location when the reinforcement ratios for responding on either are equal. A $\log b$ (stimulus) value of zero denotes no bias to a particular comparison stimulus. A $\log b$ (location) value of zero denotes no bias to a particular location. A value of less than or greater than zero when calculating $\log b$ (location) denotes bias for one of the locations (e.g. left or right; Baum, 1974). A value of less than or greater than zero when calculating $\log b$ (stimulus) denotes bias for a particular comparison (e.g. picture of dog or cat). In the proposed study, both types of biases were assessed.

Although the quantitative framework of Davison and Tustin (1978) largely has been ignored in ABA research on conditional discriminations, Fisher et al. (2014) used a novel approach to mitigate errors based on assumptions from Davison and Tustin's (1978) framework (see also Davison & Nevin, 1999). Specifically, they focused on enhancing the discriminability of the behavior-reinforcer relations (e.g. reinforcement contingencies) in an attempt to decrease the number of errors performed by three children diagnosed with ASD during conditional discriminations. They arranged a token economy in which one edible reinforcer was put in a clear container in front of the participant for each correct response. Three consecutive correct responses resulted in delivery of all three edibles. To increase discriminability between the behavior-reinforcer relations, researchers removed all the accumulated edibles from the containers contingent on

an error response. This procedure increased the mean percent of correct responding during the treatment phase by 20-30% for all three participants. However, the participants still emitted errors during the intervention phase, as well as after the intervention was withdrawn.

A major limitation of Fisher et al.'s (2014) approach from the perspective of Davison and Tustin (1978) is this approach did not identify the patterns of responding producing the remaining errors. Although an assumption of the Davison and Tustin model provided the foundation for the procedure for increasing the discriminability of the behavior-reinforcer relations, Fisher et al. (2014) failed to fully use the insights the model has to offer. The model can identify and quantify sources of errors, which has not been a tactic examined in any applied research evaluating errors during discrimination training thus far. As a result, ongoing research has not attempted to or been able to mitigate the errors according to their origin. Therefore, Fisher et al.'s (2014) strategy joins most other studies of conditional-discrimination training that focus on a blanket approach to minimizing errors rather than identifying why the errors occur (e.g., Carroll et al., 2015; Fisher et al., 2014; Kangas & Branch, 2008; Lionello-DeNolf, Dube, & McIlvane, 2007; Lionello-DeNolf, Silva Barros, & McIlvane, 2008; Petursdottir & Aguilar, 2016; Rodgers & Iwata 1991; Smith et al., 2006; Townley-Cochran et al., 2017).

The framework created by Davison and Tustin (1978) to quantify errors based on discriminability and bias could allow clinicians and researchers to isolate

patterns of errors interfering with acquisition of conditional discriminations. This approach of identifying types of errors is consistent with how clinicians identify the variable(s) maintaining problem behavior before intervening to treat problem behavior (e.g. Iwata et al., 1982/1994), as described above. Taking this functional approach to identifying and mitigating errors could help clinicians and researchers to develop and implement procedures for more effectively teaching conditional discriminations to individuals with ASD.

Despite the promise for using Davison and Tusin's (1978) equations for quantifying discriminability and bias in early-intervention research and interventions, Equations 1, 2, and 3 have only been used to categorize errors in conditional-discrimination performance in basic laboratory research (see Davison and Nevin, 1999, for a review). Due to this lack of research on errors during discrimination tasks in EIBI research and treatment, the present study is the first investigation of the use of this quantitative framework to characterize errors in conditional-discrimination performance with a clinically relevant population. The present study arranged conditional discriminations on an automated touchscreen interface to implement a matching-to-sample procedure with children diagnosed with ASD. During training, this study arranged blue and yellow sample stimuli and matching blue and yellow comparison stimuli on the touchscreen interface. This training condition facilitated teaching participants how to respond correctly to matching-to-sample procedures before moving into experimental sessions. Once

conditional-discrimination performance was established reliably with high accuracy during the blue-yellow discrimination, experimental conditions began. In the first phase, darker red and lighter pink sample stimuli and red and pink comparison stimuli were presented in a symbolic matching-to-sample procedure. Once accuracy stabilized, more similar sample stimuli were introduced (more similar red and pink) while maintaining corresponding red and pink comparison stimuli. After accuracy stabilized, we implemented a reversal of conditions back to the same sample and comparison stimuli as in the first phase. Some researchers found that changing the difficulty of the sample stimuli affects $\log d$ without affecting $\log b$ (McCarthy & Davison, 1980). Therefore, calculations of $\log d$, $\log b$ (stimulus), and $\log b$ (location) occurred throughout each of the phases to characterize whether changing similarity of the sample stimuli impacted discriminability without impacting biases. Assessing the ability of $\log d$, $\log b$ (stimulus), and $\log b$ (location) to categorize errors during conditional discrimination performance in children will allow researchers to better understand the utility of these equations in guiding clinical decisions for teaching skills.

Method

Participants

Three children, Alfred, Harry, and Suzie, participated in this study. All participants were recruited from an Early Intensive Behavioral Intervention center. Participants all demonstrated the ability to follow simple instructions, sit or stand for five-minute sessions, and the ability to emit the gross motor response of pressing the touchscreen device. During consent meetings, all parents reported that none of the participants had a diagnosis of color blindness.

Alfred was 6 years old at the commencement of this study and had been receiving ABA services intermittently for 3 years with continuous service for the last 15 months. He was diagnosed with Autism Spectrum Disorder, Unspecified Disruptive Impulse-Control and Conduct Disorder, Stereotypic Movement Disorder with Self-Injury, and Phonological Disorder. He scored 160 (out of a possible 170) on the *Verbal Behavior Milestones Assessment and Placement Program* (VB-MAPP; Sundberg, 2008) with 15 out of 15 on the Visual Performance/Matching-to-Sample section. The second participant, Harry, was 4 years and 2 months at the beginning of the study. He was diagnosed with Autism Spectrum Disorder and had been receiving ABA services for 8 months. He scored 124 points (out of a possible 170) on the VB-MAPP with 11 out of 15 on the Visual Performance/Matching-to-Sample section. The last participant, Suzie, was 4 years and 2 months at the

beginning of the study. She also was diagnosed with Autism Spectrum Disorder and had been receiving ABA services for 11 months. She scored 162.5 points (out of a possible 170) on the VB-Mapp with 14.5 out of 15 on the Visual Performance/Matching-to-Sample section.

Setting and Materials

Sessions were run in a small room at the university EIBI center. Each room contained a table and chairs, edibles, and the touchscreen Windows-based laptop with Paradigm software being used for sessions and data collection, as well as a video camera.

Response Definition and Measurement

The primary dependent measures were the number of correct and incorrect responses during each session. Four response types were recorded for correct and incorrect responses based on comparison colors. These consisted of incorrect response of picking red comparison, incorrect response of picking pink comparison, correct response of picking red comparison, and correct response of picking pink comparison. Correct responses were defined as touching the comparison stimulus matching or corresponding with the sample stimulus that immediately preceded display of the comparison stimuli (e.g. pink comparison after lighter pink sample). Incorrect responses were defined as touching the comparison stimulus that did not match or correspond with the sample stimulus that

immediately preceded the comparison stimulus (e.g. red comparison after lighter pink sample). The location chosen (e.g. left or right) when participants responded correctly or incorrectly was also collected. The correct and incorrect responses were then used to calculate the following equations:

$$\log d = .5 \log \left[\left(\frac{C_1}{E_1} \right) \left(\frac{C_2}{E_2} \right) \right] \quad (1)$$

$$\log b = .5 \log \left[\left(\frac{C_1}{C_2} \right) \left(\frac{E_1}{E_2} \right) \right] \quad (2)$$

$$\log b = .5 \log \left[\left(\frac{C_{right}}{C_{left}} \right) \left(\frac{E_{right}}{E_{left}} \right) \right] \quad (3).$$

In Equations 1 and 2, C_1 and E_1 refer to correct and incorrect responses, respectively, to the comparison stimulus (e.g. touching red comparison or pink comparison) following the first sample stimulus (S1; e.g. darker red). Similarly, C_2 and E_2 refer to the correct and incorrect responses to the comparison stimulus (e.g. touching pink comparison or red comparison) following the second sample stimulus (S2; e.g. lighter pink). In Equation 3, C_{right} and C_{left} refer to the correct responses to the comparison stimulus (e.g. touching red comparison after darker red sample or touching pink comparison after lighter pink sample) when it is on the right or left side, while E_{right} and E_{left} refer to the incorrect responses (e.g. touching pink comparison after the darker red sample or touching the red comparison after the lighter pink sample) to the comparison stimulus on the right or left side, respectively.

Because the equations for calculating $\log d$, $\log b$ (stimulus), and $\log b$ (location) cannot be calculated if there is a value of zero in one of the four response categories, an error correction procedure was implemented for every session for all participants. This error correction procedure added 0.25 to each response category when calculating $\log d$, $\log b$ (stimulus), and $\log b$ (location).

Percent correct (or accuracy) was calculated by adding the total number of correct responses for the session, dividing it by the total number of trials responded to, and multiplying by 100. The number of reinforcer deliveries and any missed trials (e.g. participant did not respond before 30 s elapsed and software moved to next trial) during each session were also recorded.

Summations of each dependent measure (percent correct, $\log d$, $\log b$ (stimulus), and $\log b$ (location)) were calculated for the last seven sessions of each Phase for all participants. Seven sessions were used to calculate these measures because seven sessions were the least number of sessions needed to complete one phase across all participants. The sums were calculated with the last seven sessions of each Phase in order to ensure the least crossover effects on behavior from the previous Phase. In order to calculate these measures, sums of each of the four response categories were calculated for each Phase for each participant (e.g. Suzie exhibited 27 errors to the red comparison in total during Phase 2). These four response type sums were then used to calculate the same equations for percent

correct, $\log d$, $\log b$ (stimulus), and $\log b$ (location) as described above (see Equations 1, 2, and 3).

Procedures

Depending on availability, participants attended experimental sessions two to three times per week, with two to five 5-min sessions being conducted each day.

Preference assessment. Edibles (small pieces of preferred foods) served as the primary reinforcer for this study. Prior to the initial session, caregivers and clinicians were asked to report highly preferred edibles for each participant. Before each session, the experimenter conducted a brief multiple-stimulus-without-replacement (MSWO) preference assessment (Carr, Nicolson, & Higbee, 2000). The same choices were displayed in each MSWO for the remainder of the study. The first two edibles selected during the MSWO were randomly selected and given after each correct response during every session (Egel, 1981).

Training. During training, the primary researcher taught participants to respond to the matching-to-sample procedure on a touchscreen device under a fixed-ratio (FR) 1 schedule of edible reinforcement (delivered for every correct response) for selecting the correct match. Participants first learned to respond on a simple matching-to-sample with one sample stimulus and one comparison stimulus and then began training with one sample stimulus and two comparison stimuli. Depending on the skill level of the participant a prompting procedure was used to

help facilitate learning to respond appropriately during training only. At the beginning of training, the participant was instructed to “do this” with a model or physical prompt as needed. A most-to-least prompting strategy (e.g. full physical, partial physical, tap, gesture) was used to fade prompts (MacDuff, Krantz, McClannahan, 2001).

For the duration of Training phase for Harry, each session consisted of 25 trials. After initial data analyses and program configuration, the matching-to-sample was adjusted to consist of 24 trials during each session (e.g. 12 presentations of each sample) and remained at 24 trials for the remainder of the study for all participants. The order and location of the sample and comparison stimuli rotated according to a predetermined randomized list. This means the order of the sample stimuli were randomized each session and the location of the comparison stimuli were randomized on each trial, with a .5 probability of being on the left. Colors of the background, sample stimuli, and comparison stimuli were defined by RGB color values supported in all browsers. An RGB color value is specified with: R (red), G (green), and B (blue). Each parameter (R, G, B) defines the intensity of the color as an integer between 0 and 255. For example, R0G0B255 is labeled as blue, because the blue parameter is set to its highest value and the others are set to 0. Training began with the presentation of a R13G1B255 (blue) or R255G255B0 (yellow) sample stimulus in a central location on the R0G0B0 (black) background. Participants were taught to touch the sample stimulus

appearing on the touchscreen to display one correct (matching) comparison stimulus on either the right or left side of the screen. Touching both the sample stimulus and then the one comparison stimulus resulted in the appearance of a star on the screen for five seconds and delivery of social praise (e.g. “you got it!”) and one of the top two edible reinforcers identified by the MSWO preference assessment by the researcher according to the FR1 schedule of reinforcement. Once accurate responding occurred reliably and independently for 95% of trials or higher across two consecutive sessions, the next part of training was introduced.

The second part of training began with the presentation of a blue or yellow sample stimulus in a central location on the black background. Touching the sample stimulus resulted in the appearance of two comparison stimuli each on one side of the screen. One of the comparisons (correct response) was identical to the sample stimulus. Touching the correct response resulted in appearance of a star on the screen for five seconds and delivery of an edible and social praise by the researcher. Touching the non-identical comparison (incorrect) resulted in a black screen for the same duration as the star consequence and no edible delivery. A brief 3-second black screen followed both consequence displays (star and black screen) before the next trial began. Participants began experimental sessions after responding occurred independently at 90% accuracy or above for two consecutive sessions or stable independent responding occurred for three consecutive sessions below 90% accurate.

Experimental Sessions. Experimental sessions were identical to training sessions with the exception of providing no prompting, changing the color of the stimuli, and consisting of symbolic matching-to-sample trials rather than identity matching-to-sample trials. Highly different samples during Phase 1 allowed participants to learn to respond accurately during the symbolic matching-to-sample procedure with stimuli different from training while categorizing any errors they made due to discriminability or bias. More similar sample stimuli were presented during Phase 2 with the same corresponding comparison stimuli as in Phase 1. In Phase 3, a reversal to the same highly different samples as in Phase 1 occurred.

Phases 1 and 3: Different In the first and third phases, the sample stimuli were R255G51B51 (darker red) and R255G100B100 (lighter pink) and comparison stimuli were R188G0B0 (red comparison) and R255G155B155 (pink comparison). Touching the red comparison after the darker red sample and touching the pink comparison after the lighter pink sample (correct responses) resulted in the appearance of the star for five seconds, as well as delivery of social praise and edible from the researcher according to the FR1 schedule of reinforcement. Touching the red comparison after the lighter pink sample and touching the pink comparison after the darker red sample (incorrect responses) resulted in the black screen for five seconds and no edible delivery. A black screen followed either consequence (star or black screen) for 3 seconds before the next trial began. Following Phase 1, Phase 2 began once accurate responding reached stability with

no increasing or decreasing trends using visual inspection (Sidman, 1960).

Following Phase 2, Phase 3 began once responding reached stability once again.

Phase 2: Similar. Phase 2 introduced more similar sample stimuli, R255G70B70 (lighter red) and R255G95B95 (darker pink) compared to Phase 1 while maintaining identical comparison stimuli as Phase 1. Touching the red comparison after the lighter red sample or the pink comparison after the darker pink sample (correct responses) resulted in a star appearing on the screen for five seconds and the delivery of edible and social praise by the researcher according to the FR1 schedule of reinforcement. Touching the red comparison after the darker pink sample or the pink comparison after the lighter red sample (incorrect responses) resulted in the black screen for five seconds and no reinforcer delivery. Both consequence displays (the star or the black screen) were followed by a black screen for 3 seconds before the next trial begins. Phase 3 began as described above once accurate responding stabilized according to visual inspection.

Results

Figures 1, 2, and 3 depict the results for Alfred, Harry, and Suzie, respectively. All aspects of the figures are identical with each x-axis depicting sessions and each individual y-axis depicting percent correct, values of $\log d$, $\log b$ (stimulus), or $\log b$ (location). Training depicts sessions with the blue and yellow sample and comparison stimuli. Phases 1 and 3 depict sessions with the highly different darker red and lighter pink sample stimuli with corresponding red and pink comparison stimuli. Phase 2 depicts sessions with the more similar lighter red and darker pink sample stimuli and corresponding red and pink comparison stimuli.

Suzie, Harry, and Alfred required four, nine, and thirteen sessions of Training, respectively, to meet the mastery criteria to move into experimental sessions. Only Suzie meet our predetermined criteria of two consecutive sessions of 90% or higher correct responding. Both Harry's and Alfred's responding were not quite that high but were accurate enough to conclude they were under control by the stimuli and had learned to respond appropriately to the task. During Training, trends of $\log d$ correspond with trends of percent correct for each individual, indicating how $\log d$ impacts percent correct considerably with relatively little change in stimulus or location bias for two of the three participants.

During Training, both Alfred's (see Figure 1) and Suzie's (see Figure 2) responding displayed $\log b$ (stimulus) and $\log b$ (location) values that remained

near zero with little variability, indicating little to no bias to either stimulus or location. Harry's responding (see Figure 3) displayed a little more variability for both $\log b$ (stimulus) and $\log b$ (location), indicating a small bias to one stimulus for a few sessions of Training and a small bias towards one location throughout Training.

In Phase 1, Suzie and Alfred both displayed high and stable levels of accuracy as shown by percent correct. Harry had a small decreasing trend in percent correct which stabilized near 80% in the last three sessions of Phase 1. $\log d$ values followed the general trend of percent correct for all participants with high levels near 0.75, 1, and 1.5 for Harry, Alfred, and Suzie, respectively. For both Alfred and Harry, $\log b$ (stimulus) values display a little variability and $\log b$ (location) displayed more variability ranging from about -.5 to .5, indicating both of these participants had a small bias towards the red comparison stimulus during a few sessions and a bias towards a particular location throughout Phase 1. Alfred's location bias was primarily the left side and Harry's location bias was primarily to the right side until the last three sessions of Phase 1 where it switched to the left side. Suzie, however, did not display any biases throughout Phase 1, as both $\log b$ (stimulus) and $\log b$ (location) remained near zero.

At the start of Phase 2, accurate responding, displayed as percent correct, and $\log d$ decreased immediately and continued at low levels throughout Phase 2, for all participants. This shows how the presentation of more similar sample stimuli

significantly impacts percent correct and $\log d$. Harry and Suzie both display some variability, but like Alfred, the general trends for percent correct hover around chance levels and are much lower compared to the level of percent correct in Phase 1 for each participant. Both Alfred and Suzie displayed $\log b$ (stimulus) values that remained near zero throughout Phase 2, indicating little to no bias to a particular stimulus. Suzie's values of $\log b$ (stimulus) started to decrease in the last three sessions of Phase 2, indicating a small bias towards the pink comparison stimulus developed. Harry displayed more variability during Phase 2, with $\log b$ (stimulus) values remaining between -0.5 and -1.4, indicating a large bias towards the pink comparison stimulus. All participants displayed a small amount of variability in values of $\log b$ (location), but for Alfred and Harry these values were not drastically different than variability of values of $\log b$ (location) in Phase 1. Suzie displayed only two sessions with $\log b$ (location) values different than in Phase 1. This indicates that there was not a large difference in bias to a particular location across phases for any of the participants. There was not a large difference in bias to a particular stimulus for both Alfred and Suzie, but Harry had reliably different stimulus bias during Phase 2.

During Phase 3, Harry's accurate responding, as displayed by percent correct, and $\log d$ increased immediately, returning to near the same levels as in Phase 1. While, Alfred and Suzie did not have a large increase in their first session of Phase 3, accurate responding quickly increased and remained high. Similar to

levels of variability seen in Phase 1, Suzie displayed some variability in values of percent correct in Phase 3. For Alfred, accurate responding stabilized at 95% correct (e.g. one error per session) for the last six sessions of Phase 3. This indicates the presentation of the highly different sample stimuli greatly impacts accuracy and $\log d$. Harry displayed little variability in values of $\log b$ (stimulus), indicating little to no bias for a particular stimulus throughout Phase 3 similar to values seen during Phase 1. Suzie displayed some variability, indicating a small bias towards one stimulus during two sessions. However, these values are similar to values of $\log b$ (stimulus) in all Phases for Suzie. Harry also displayed a small amount of variability in values of $\log b$ (location), indicating a small bias towards the right side, as seen throughout all phases for Harry. While Alfred initially showed little variability in values of $\log b$ (stimulus) and $\log b$ (location) in Phase 3, during the last six sessions of Phase 3, these values displayed some variability. Because Alfred only exhibited one error during each of these six sessions, this displayed a bias towards the pink comparison in three sessions, the red comparison in three sessions, the right side in two sessions and the left side in four sessions.

Figures 4, 5, and 6 depict summaries of percent correct, $\log d$, $\log b$ (stimulus), and $\log b$ (location) for the last seven sessions of each Phase, including Training, for each participant. The gray bars depict Training. The black bars depict Phases 1 and 3 with highly different sample stimuli and the white bars depict Phase 2 with more similar sample stimuli. All three participants display a large decrease in

percent correct and values of $\log d$ during Phase 2 when more similar stimuli are presented. All participants display some variability in $\log b$ (stimulus) across phases indicating it is possible that changes in bias may have played a role in changes of percent correct in conjunction with the impact of changes in $\log d$. Especially for Harry who had a large difference in values of $\log b$ (stimulus) during Phase 2. Suzie displays little variability in $\log b$ (location) across phases indicating no location bias present during all phases. Both Harry and Alfred display relatively large location biases during one phase (e.g. Phase 1 for Alfred and Phase 3 for Harry).

Overall, for Alfred and Suzie there were not any large differences in level or trend of either $\log b$ (stimulus) or $\log b$ (location) across phases. There were some individual sessions that displayed a larger bias but generally values were stable within and across phases. For Harry, there was a large difference in his values of $\log b$ (stimulus) during Phase 2, indicating that a bias towards a particular stimulus developed. When less similar stimuli were presented in Phase 3, values of $\log b$ (stimulus) quickly increased to near zero levels, suggesting that the levels of variability of $\log b$ (stimulus) occurred mostly when the stimuli were much more similar. Overall, percent correct and $\log d$ during Phase 2 was greatly impacted by the presentation of more similar sample stimuli. This shows how changes in percent correct are mostly due to changes in $\log d$ and typically do not impact biases.

Discussion

This study used Equations 1, 2, and 3 from Davison and Tustin (1978) to categorize errors due to bias and discriminability during conditional discriminations, in particular during simple matching-to-sample tasks with either highly different sample stimuli (Phases 1 and 3) or more similar sample stimuli (Phase 2). The results show the ability of $\log d$ and $\log b$ to categorize the errors made during these tasks. As we hypothesized, the results show how percent correct (accuracy) is impacted mostly by discriminability and doesn't generally impact any biases.

These findings are similar to the basic literature evaluating $\log d$ and $\log b$ (Davison & Nevin, 1999; Hutsell & Banks, 2015). For example, Hutsell and Banks manipulated the discriminability of their nonmatching-to-sample procedure (e.g. match the sample to the non-identical comparison). They introduced longer delays between the offset of the sample stimuli and the onset of the comparison stimuli. With the longest delays, $\log d$ values approached zero. The results of the present study show how changing the sample stimuli from highly different darker red and lighter pink to more similar lighter red and darker pink impacted accuracy (percent correct) through decreases in $\log d$ but did not reliably impact stimulus or location biases for two of the three participants (McCarthy & Davison, 1980).

Varied Responding as a Result of Difficult Discrimination

Although the changes in $\log d$ during Phase 2 did not reliably impact stimulus or location biases for two of the three participants, we did observe a large change in stimulus bias for Harry and a few small changes in stimulus or location biases for Alfred and Suzie. Harry demonstrated a sharp decrease in values of $\log b$ (stimulus) throughout Phase 2, indicating an almost exclusive bias to the pink comparison stimulus. Suzie also displayed slightly larger bias towards the left, then right side in a few sessions and the pink comparison stimulus in three sessions of Phase 2. It may be possible that this varied responding is a result of each participant's past history with schedules of reinforcement. These historical variables may influence behavior once the more similar stimuli are introduced even with the FR1 schedule of reinforcement for correct responses in place (Freeman & Lattal, 1992). When the stimuli presented were too difficult to distinguish, past history of gaining reinforcement for particular patterns of behavior might have influenced correct responding.

High Discriminability Yields Less Accurate Quantifications

In line with research on discrimination tasks, the present study tracked four different types of responses to calculate $\log d$, $\log b$ (stimulus), and $\log b$ (location) (Alsop & Rowley, 1996; Davison & Nevin, 1999; Davison & Tustin, 1978; Hutsell & Banks, 2015). These four responses include correct and incorrect selections for

both the red and pink comparison stimuli following their corresponding (matching) sample stimulus. These equations use either the proportion of correct responses or the ratio of correct to incorrect responses to calculate biases and discriminability (Brown & White, 2005). These equations cannot be calculated if there is a value of zero in one of the four response categories. Each of the participants exhibited zero errors to one of the response categories for at least eight sessions. For Alfred and Suzie this often occurred during instances of very high discriminability, creating one or less errors in many sessions (e.g. last six sessions of Phase 3 for Alfred). For Harry, this more often was exhibited in sessions with higher bias for one of the stimuli (e.g. first few sessions of Phase 2). For this reason, the present study implemented an error correction procedure of adding a constant value of .25 to all response categories for all sessions for each participant (Alsop, 2004; Brown & White, 2005). Unfortunately, research findings show using a correction procedure can underestimate the values of $\log d$ and $\log b$. Fewer trials can also contribute to larger underestimations of these values. Most basic research with nonhuman animals can run hundreds of trials per sessions, however, the present study only ran 24 trials per session. Therefore, it is important to interpret the results with caution, as a larger number of trials would better estimate true values of $\log d$ and $\log b$. In order to mitigate this problem, future studies could attempt to run more trials per session depending on the population participating (see Brown & White, 2005, for a discussion). Another possible solution is to introduce stimuli in Phases 1 and 3 that

are slightly more similar, therefore participants would still respond with more accuracy than in Phase 2, but not with such accuracy that they would make no errors.

Limitations

The current study did not counterbalance the stimuli that were associated with each condition across participants, nor did it counterbalance the order of the conditions. In studies with nonhuman primates, it is common procedure to counterbalance the stimuli and the order of the conditions across participants when using a within-subjects design (D'amato, Salmon, Loukas, & Tomie, 1985; Proctor, Williamson, Latzman, de Waal, & Brosnan, 2014). This procedure often varies which stimuli are samples and which are comparisons across participants in order to control for any potential influence the particular stimuli have on responding. For example, one study presented two stimuli (e.g. triangle and dot) as the sample stimuli for two participants and then as the comparison stimuli for two other participants with the same matching relations occurring across participants (D'amato et al., 1985). This procedure can create additional experimental control to analyze across participants. Although this study could not counterbalance in the same way, future studies could use multiple sets of stimuli for all of the participants to ensure there are no variables directly related to the stimuli impacting responding. It is possible that the order of the conditions in the present study impacted responding. In order to control for any possible impact the order of conditions

could have, future studies could start some participants in Phase 2 (more similar stimuli) and some participants in Phase 1 to see if any differences in responding occur.

Implications for Clinical Practice and Research

When researchers or practitioners determine that a child is making errors during conditional discriminations, antecedent and consequent manipulations can be put in place to attempt to decrease errors (Carroll et al., 2015; Fisher et al., 2014; Kangas & Branch, 2008; Lionello-DeNolf et al., 2007; Lionello-DeNolf et al., 2008; McIlvane & Duve, 2003; Petursdottir & Aguilar, 2016; Rodgers & Iwata 1991; Smith et al., 2006; Townley-Cochran et al., 2017). In the literature so far, researchers have not used a systematic approach for tracking errors to guide their decisions about how to mitigate errors, but the approach the present study used both identifies and quantifies such errors. The framework used to quantify errors in the present study, yields data-based information that practitioners could use to guide their clinical decisions when clients are not making progress. For example, identifying whether a participant is making persistent errors because of biases or reduced discriminability can inform the practitioner or researcher to make either antecedent or consequent manipulations based on the kinds of errors emitted by the individual. If these equations depict that errors are due to low discriminability, practitioners could increase the saliency of the two sample stimuli to increase accuracy (see Lionello-DeNolf et al., 2008). If these equations depict that an

individual is erroring due to a location bias, blocking a response to the location bias and prompting the correct response could create a longer history of receiving reinforcement for the correct response, as observed during repetitive error correction procedures in past literature (Kangas & Branch, 2008; Rodgers & Iwata, 1991). This information could be very helpful in particular for practitioners working with clients with ASD, who are often directly taught conditional discriminations (Green, 2001). This framework can give practitioners and researchers a way to systematically evaluate errors clients are exhibiting. Future studies should evaluate the utility of this framework in more natural settings, potentially with clinically relevant stimuli and programming.

References

- Alsop, B. (2004). Signal-detection analyses of conditional discriminations and delayed matching-to-sample performance. *Journal of the Experimental Analysis of Behavior, 82* (1), 57-69.
- Alsop, B., & Rowley, R. (1996). Types of responding in a signal-detection task. *Journal of the Experimental Analysis of Behavior, 65*, 561-574.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.
- Baer, D., Wolf, M., & Risley, T. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91-97.
- Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22*, 231-242.
- Beavers, G., Iwata, B., & Lerman, D. (2013). Thirty years of research on the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 46*(1), 1-21.
- Brown, G. S., & White, K. G. (2005) The optimal correction for estimating extreme discriminability. *Behavior Research Methods, 37* (3), 436-449.

- Carr, E. & Durand, M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*, 111-126.
- Carr, J., Nicolson, A., & Higbee, T. (2000). Evaluation of a brief multiple-stimulus preference assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33*, 353-357.
- Carroll, R., Joachim, B., St. Peter, C., & Robinson, N. (2015). A comparison of error-correction procedures on skill acquisition during discrete-trial instruction. *Journal of Applied Behavior Analysis, 48(2)*, 257-273.
- Cumming, W., & Berryman, R. (1961). Some data on matching behavior in the pigeon. *Journal of the Experimental Analysis of Behavior, 4*, 281-284.
- Da Silva Barros, R., De Faria Galvao, O., & McIlvane, W. (2002) Generalized identity matching-to-sample in cebus apella. *The Psychological Record, 52*, 441-460.
- D'amato, M. R., Salmon, D. P., Loukas, E. & Tomie, A. (1985) Symmetry and transitivity of conditional relations in monkeys (cebus apella) and pigeons (columba livia). *Journal of the Experimental Analysis of Behavior, 44(1)*, 35-47.
- Davison, M., & McCarthy, D. (1987). The interaction of stimulus and reinforcer control in complex temporal discriminations. *Journal of the Experimental Analysis of Behavior, 48*, 97-116.

- Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior*, 71, 439-482.
- Davison, M., & Tustin, R. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of Behavior*, 29, 331-336.
- Doughty, A., & Hopkins, M. (2011). Reducing stimulus overselectivity through an increased observing-response requirement. *Journal of Applied Behavior Analysis*, 44, 653-657.
- Dube, W., Dickson, C., Balsamo, L., Lombard O'Donnell, K., Tomanari, G., Farren, K., ... McIlvane, W. (2010). Observing behavior and atypically restricted stimulus control. *Journal of Experimental Analysis of Behavior*, 94, 297-313.
- Dube, W., & McIlvane, W. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis*, 32, 25-33.
- Egel, A. (1981). Reinforcer variation: Implications for motivating developmentally disabled children. *Journal of Applied Behavior Analysis*, 14, 345-350.
- Fisher, W., Kodak, T., & Moore, J. (2007). Embedding an identity matching task within a prompting hierarchy to facilitate acquisition of conditional discriminations in children with autism. *Journal of Applied Behavior Analysis*, 40 (3), 489-499.

- Fisher, W., Pawich, T., Dickes, N., Paden, A., & Toussaint, K. (2014). Increasing the saliency of behavior-consequence relations for children with autism who exhibit persistent errors. *Journal of Applied Behavior Analysis, 47*, 1-11.
- Freeman, T., & Lattal, K. (1992). Stimulus control of behavioral history. *Journal of the Experimental Analysis of Behavior, 57* (1), 5-15.
- Green, G. (2001). Behavior analytic instruction for learners with autism: Advances in stimulus control technology. *Focus on Autism and Other Developmental Disabilities, 16* (2), 72-85.
- Grow, L., Kodak, L., & Carr, J. (2014). A comparison of methods for teaching receptive labeling to children with autism spectrum disorders: A systematic replication. *Journal of Applied Behavior Analysis, 47* (3), 600-605.
- Gutowski, S. & Stomer, R. (2003). Delayed matching to two-picture samples by individuals with and without disabilities: An analysis of the role of naming. *Journal of Applied Behavior Analysis, 36*(4), 487-505.
- Hagopian, L., Fisher, W., Sullivan, M. T., Acquisto, J., & LeBlanc, L. (1998). Effectiveness of functional communication with and without extinction and punishment: A summary of 21 inpatient cases. *Journal of Applied Behavior Analysis, 31*, 211-235.
- Hagopian, L., Rooker, G., Jessel, J., & DeLeon, I. (2013). Initial functional analysis outcomes and modifications in pursuit of differentiation: A summary of 176 inpatient cases. *Journal of Applied Behavior Analysis, 46*(1), 81-100.

- Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36*, 147-185.
- Howard, J., Sparkman, C., Cohen, H., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*, 359-383.
- Hutsell, B. & Banks, M. (2015). Effects of environmental and pharmacological manipulations on a novel delayed nonmatching-to-sample 'working memory' procedure in unrestrained rhesus monkeys. *Journal of Neuroscience Methods, 251*, 62-71.
- Iwata, B., Dorsey, M., Slifer, K., & Richman, G. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197-209.
- Jones, B. M. & White, K.G. (1992). Sample-stimulus discriminability and sensitivity to reinforcement in delayed matching to sample. *Journal of the Experimental Analysis of Behavior, 58*, 159-172.
- Kangas, B. & Branch, M. (2008). Empirical validation of a procedure to correct position and stimulus biases in matching-to-sample. *Journal of the Experimental Analysis of Behavior, 90*, 103-112.

- Kodak, T., Campbell, V., Bergmann, S., LeBlanc, B., Kurtz-Nelson, E., Cariveau, T., ... Mahon, J. (2016) Examination of efficacious, efficient, and socially valid error-correction procedures to teach sight words and prepositions to children with autism spectrum disorders. *Journal of Applied Behavior Analysis, 49*, 532-547.
- Lionello-DeNolf, K., Dube, W., & McIlvane, W. (2007). Exclusive preference on concurrent schedules in children with autism spectrum disorder. *Shriver Center Publications and Presentations. Paper 28*.
- Lionello-DeNolf, K., Silva Barros, McIlvane, W. (2008). A novel method for teaching the first instances of simple discrimination to nonverbal children with autism in a laboratory environment. *Psychological Record, 58(2)*, 229-244.
- Lovaas, O. (1987). Behavioral treatment and normal education and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3-9.
- MacDuff, G.S., Krantz, P.J., & McClannahan, L.E. (2001). Prompts and prompt-fading strategies for people with autism. In C. Maurice, G. Green, & R. M. Foxx (Eds.), *Making a difference: Behavioral intervention for autism* (pp. 37-50). Austin, TX: Pro-Ed.

- McCarthy, D. & Davison, M. (1980). Independence of sensitivity to relative reinforcement rate and discriminability in signal detection. *Journal of the Experimental Analysis of Behavior*, 34, 273-284.
- McGhan, A., & Lerman, D. (2013). An assessment of error-correction procedures for learners with autism. *Journal of Applied Behavior Analysis*, 46, 626-639.
- McIlvane, W., & Dube, W. (2003). Stimulus control topography coherence theory: foundations and extensions. *The Behavior Analyst*, 26(2), 195-213.
- Northrup, J., Wacker, D., Sasso, G., Steege, M., Cigrand, K., Cook, J., & DeRaad, A. (1991). A brief functional analysis of aggression and alternative behavior in an outclinic setting. *Journal of Applied Behavior Analysis*, 24(3), 509-522.
- Peterson, G., Wheeler, R., & Trapold, M. (1980). Enhancement of pigeons' conditional discrimination performance by expectancies of reinforcement and nonreinforcement. *Animal Learning and Behavior*, 8(1), 22-30.
- Petersdottir, A. I., & Aguilar, G. (2016). Order of stimulus presentation influences children's acquisition in receptive identification tasks. *Journal of Applied Behavior Analysis*, 49, 58-68.

- Proctor, D., Williamson, R. A., Latzman, R., de Waal, B. M. F., & Brosnan, S. F. (2014). Gambling primates: reactions to a modified Iowa gambling task in humans, chimpanzees, and capuchin monkeys. *Animal Cognition, 17*, 983-995.
- Rodgers, T. A., & Iwata, B. A. (1991). An analysis of error-correction procedures during discrimination training. *Journal of Applied Behavior Analysis, 24*, 775-781.
- Sidman, M. (1960). *In Tactics of scientific research: Evaluating experimental data in psychology*. Oxford, England: Basic Books.
- Sidman, M., Cresson, O., Willson-Morris, M. (1974). Acquisition of matching to sample via mediated transfer. *Journal of the Experimental Analysis of Behavior, 22*, 261-273.
- Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discrimination of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior, 37 (1)*, 23-44.
- Shahan, T., & Podlesnik, C. (2006). Divided attention performance and the matching law. *Learning & Behavior, 34(3)*, 255-261.
- Smith, T., Mruzek, D., Wheat, L.A., & Hughes, C. (2006). Error correction in discrimination training for children with autism. *Behavioral Interventions, 21*, 245-263.

- Sundberg, M. L. (2008). VB-MAPP: *Verbal behavior milestones assessment and placement program*. Concord, CA: AVB Press.
- Tomonaga, M. (1993). Tests for control by exclusion and negative stimulus relations of arbitrary matching to sample in “symmetry-emergent” chimpanzee. *Journal of the Experimental Analysis of Behavior*, 59, 215-229.
- Townley-Cochran, D., Leaf, J., Leaf, R., Taubman, M., & McEachin, J. (2017). Comparing error correction procedures for children diagnosed with autism. *Education and Teaching in Autism and Developmental Disabilities*, 52, 91-101.

Appendix

The quantitative framework that categorizes the errors according to their sources of bias or discriminability in Davison and Tustin (1978) describes how accuracy occurs as a function of discriminability, relative rates of reinforcement, and bias (Alsop & Rowley, 1996; Jones & White, 1992). Davison and Tustin's quantitative framework is based on the matching law originally developed to describe choice in concurrent schedules of reinforcement. The generalized matching law accounts for operant responding between two concurrently available alternatives as a function of relative rates of reinforcement (Baum, 1974):

$$\log\left(\frac{B_1}{B_2}\right) = a \log\left(\frac{r_1}{r_2}\right) + \log c \quad (1A),$$

where B_1 and B_2 are the frequency (or rate) of responding to each of the alternatives and r_1 and r_2 are the number (or rate) of reinforcers obtained for responding on each of the alternatives. Parameter $\log c$ refers to inherent bias, the tendency to allocate responding to one alternative more than the other irrespective of reinforcer allocation. For example, responding might be allocated to the dog comparison stimulus more than the cat comparison stimulus even though reinforcer rates are equal for selecting each of those after their matching sample. Parameter a describes the sensitivity of $\frac{B_1}{B_2}$ to changes in $\frac{r_1}{r_2}$. As a becomes closer to 1, the more the

response ratio matches the reinforcement ratio. As a approaches zero, the response ratio becomes more insensitive to changes in the reinforcement ratio.

Because the generalized matching law relates the ratio of responses on the two alternatives $\left(\frac{B_1}{B_2}\right)$ to the ratio of reinforcement for those responses $\left(\frac{r_1}{r_2}\right)$, it can also describe the ratio of responses controlled by the reinforcement ratio in a typical matching-to-sample experiment. In a typical matching-to-sample experiment in which all correct responses are reinforced, the number of reinforcement deliveries matches the number of accurate comparison responses following both sample stimuli. However, the generalized matching law analyzes only response ratios without regard to the sample stimulus. Specifically, the generalized matching law only analyzes the ratio of responses on each of the comparisons in relation to the ratio of reinforcers delivered for responding on those comparisons. However, both stimulus discriminability and bias for one of the stimuli can modulate the effect the reinforcer ratio has on the response ratio (Jones & White, 1992). Because the generalized matching law does not take stimulus bias and discriminability into account, Davison and Tustin (1978) quantified discriminability and bias to derive their framework for quantifying errors in conditional discriminations. Their relation is described by two equations, one in the presence of one sample,

$$\log\left(\frac{B_1}{B_2}\right) = a \log\left(\frac{R_1}{R_2}\right) + \log b + \log d \quad (2A),$$

and another in the presence of the other sample:

$$\log\left(\frac{B_1}{B_2}\right) = a \log\left(\frac{R_1}{R_2}\right) + \log b - \log d \quad (3A),$$

This framework describes the distribution of choices following presentation of a target stimulus where B_1 and B_2 are responses to the two comparison stimuli and $\frac{R_1}{R_2}$ is the overall reinforcer distribution obtained for correct responses to those comparison stimuli. The sensitivity parameter a is the same as in the generalized matching law and scales the extent to which changes in the reinforcer distribution $\left(\frac{R_1}{R_2}\right)$ produce changes in the response allocation $\left(\frac{B_1}{B_2}\right)$. $\log d$ is negative in Equation 3A because this equation shows responding in the absence of the first stimulus, whereas $\log d$ is positive in Equation 2A because Equation 2A shows responding in the presence of the first stimulus. $\log b$ refers to bias to one alternative irrespective of the reinforcement schedules or the discriminability of the sample stimuli. $\log d$ refers to the discriminability between the sample stimuli. $\log d$ and $\log b$ as defined here are portrayed above to account for errors due to discriminability and bias.

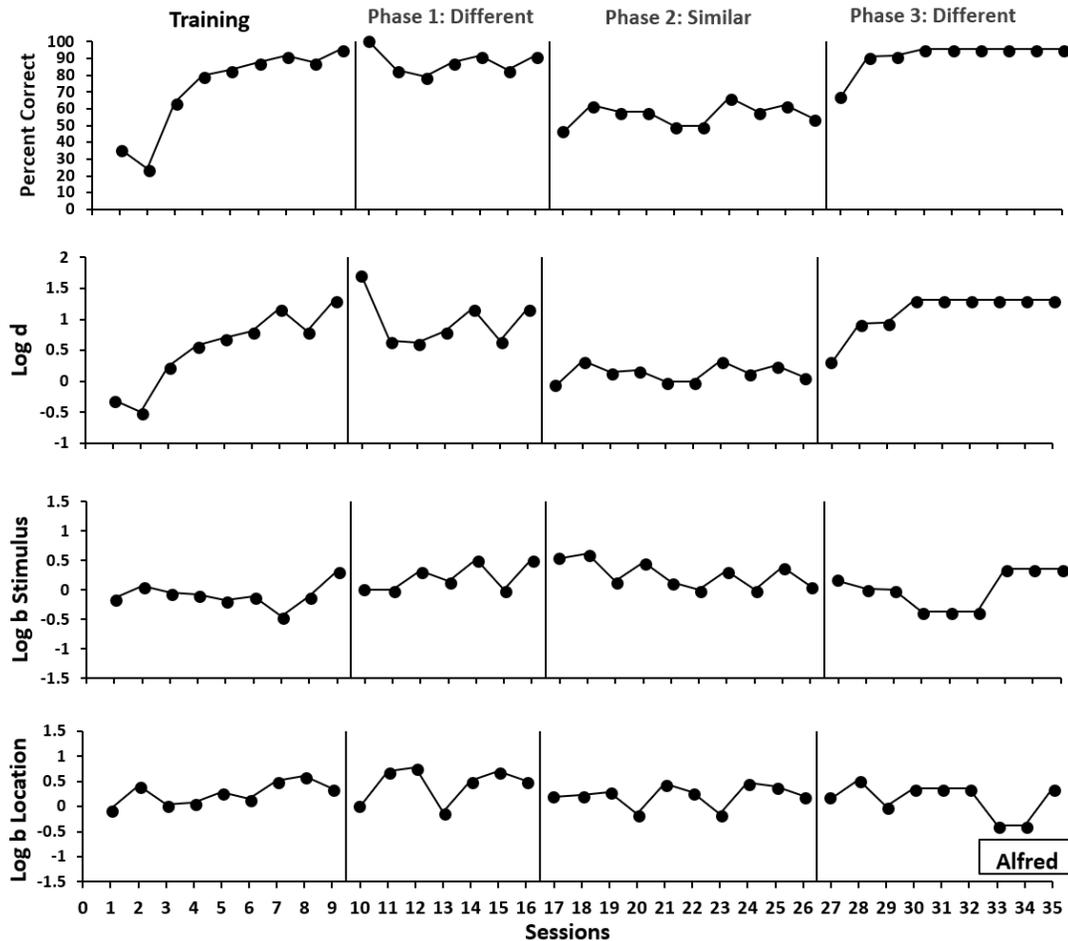


Figure 1. Results for one participant, Alfred. As more similar stimuli were introduced during Phase 2, **log d** and percent correct decreased while both stimulus and location bias remained relatively the same compared to Phase 1. A return to the highly different stimuli in Phase 3 increased both **log d** and percent correct. Some variability in stimulus and location bias in Phase 3.

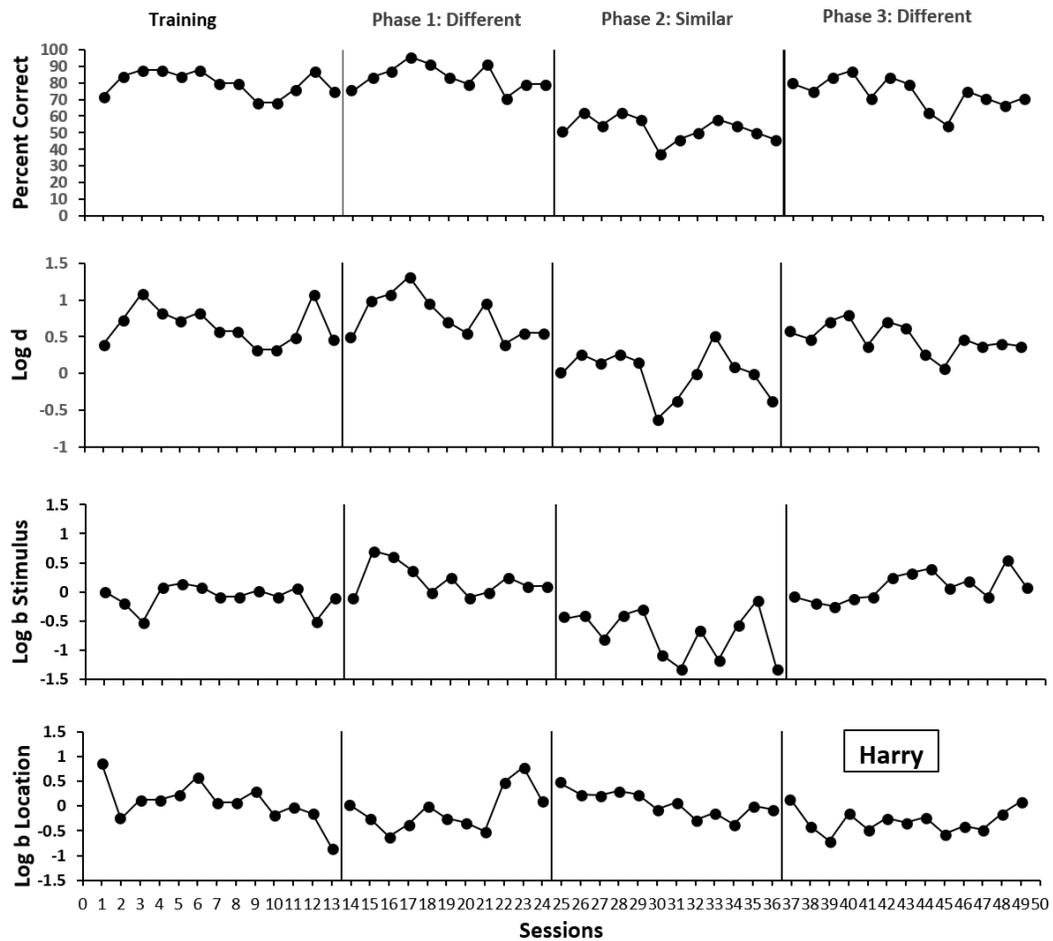


Figure 2. Results for one participant, Harry. As more similar stimuli were introduced during Phase 2, $\log d$ and percent correct decreased. Location bias hovers near zero, but stimulus bias displays more variability and a decrease in level. Returning to highly different stimuli in Phase 3 displayed increases in $\log d$ and percent correct with location and stimulus biases hovering near zero.

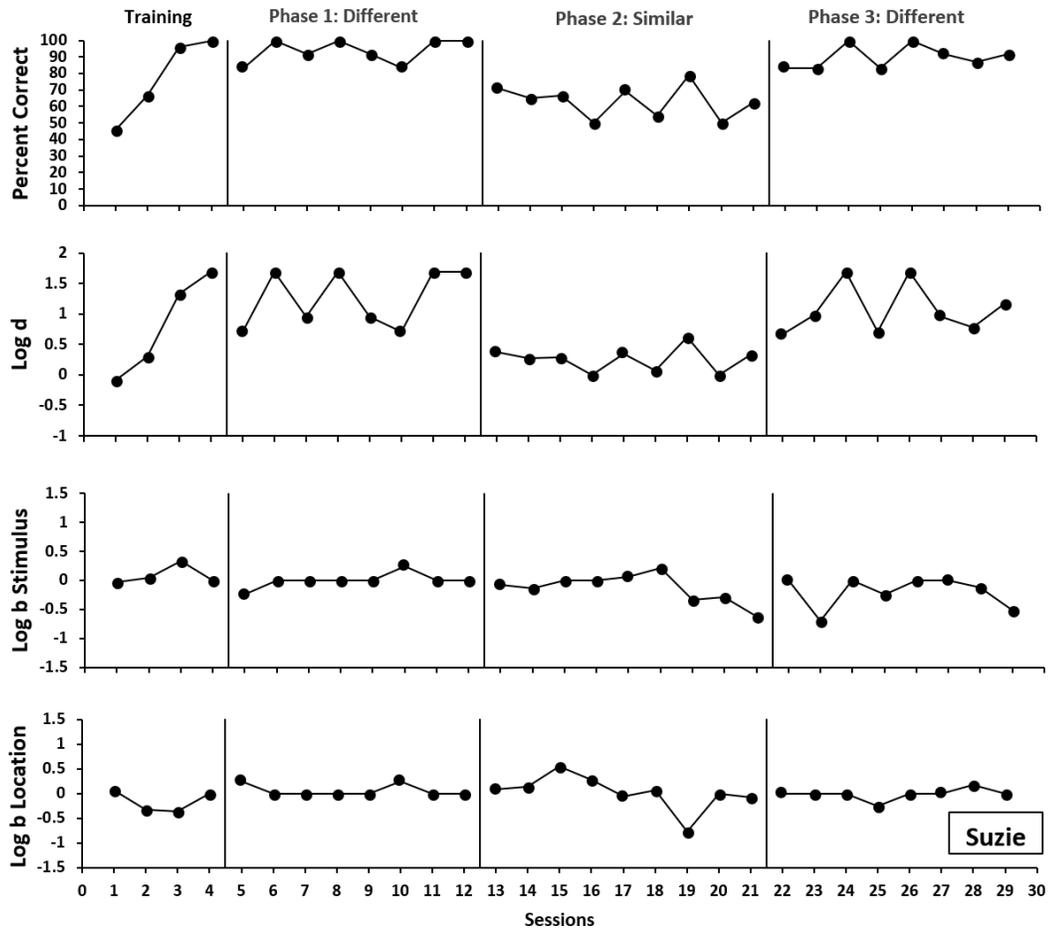


Figure 3. Results for one participant, Suzie. Presentation of more similar stimuli in Phase 2 displays a decrease in $\log d$ and percent correct. No major difference in stimulus bias compared to Phase 1, but some variability in location bias on a few sessions in Phase 2. Return to the highly different stimuli in Phase 3 shows increases in $\log d$ and percent correct to similar levels as in Phase 1. Both stimulus and location biases stay near zero.

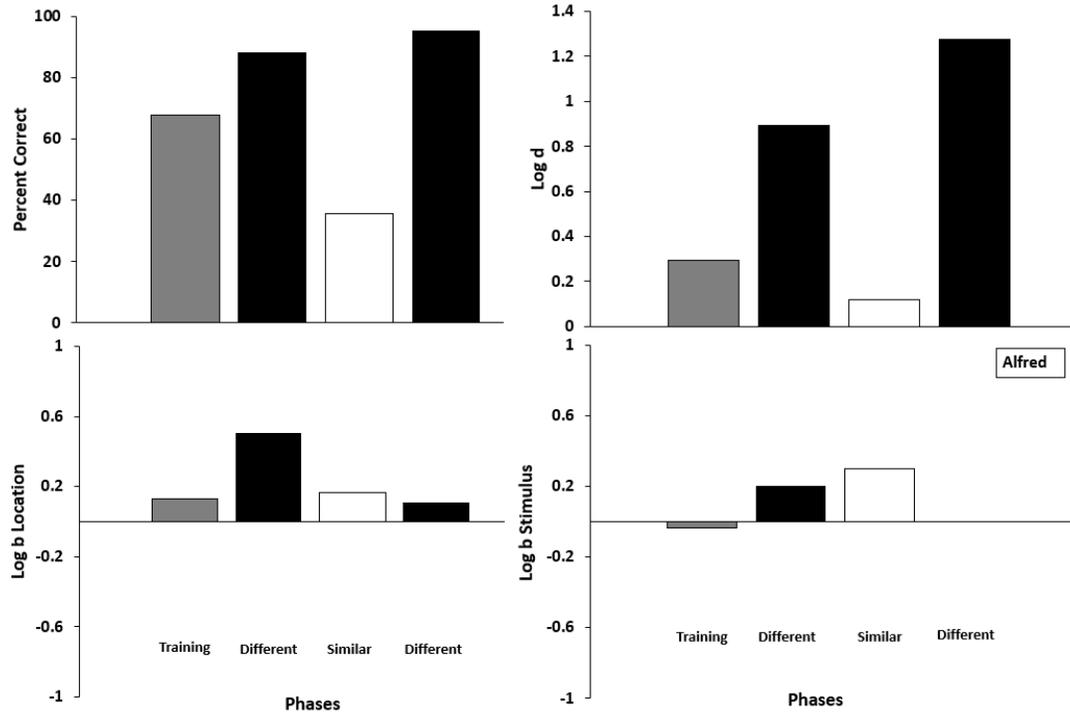


Figure 4. Depicts sum of percent correct, $\log d$, $\log b$ (stimulus), and $\log b$ (location) for the last seven sessions of each Phase, including Training, for Alfred. Gray bars depict Training. Black bars depict Phases 1 and 3 with highly different stimuli and white bars depict Phase 2 with more similar stimuli. Percent correct and $\log d$ reliably decrease when more similar stimuli are presented in Phase 2. There is some variability in both location and stimulus bias across sessions.

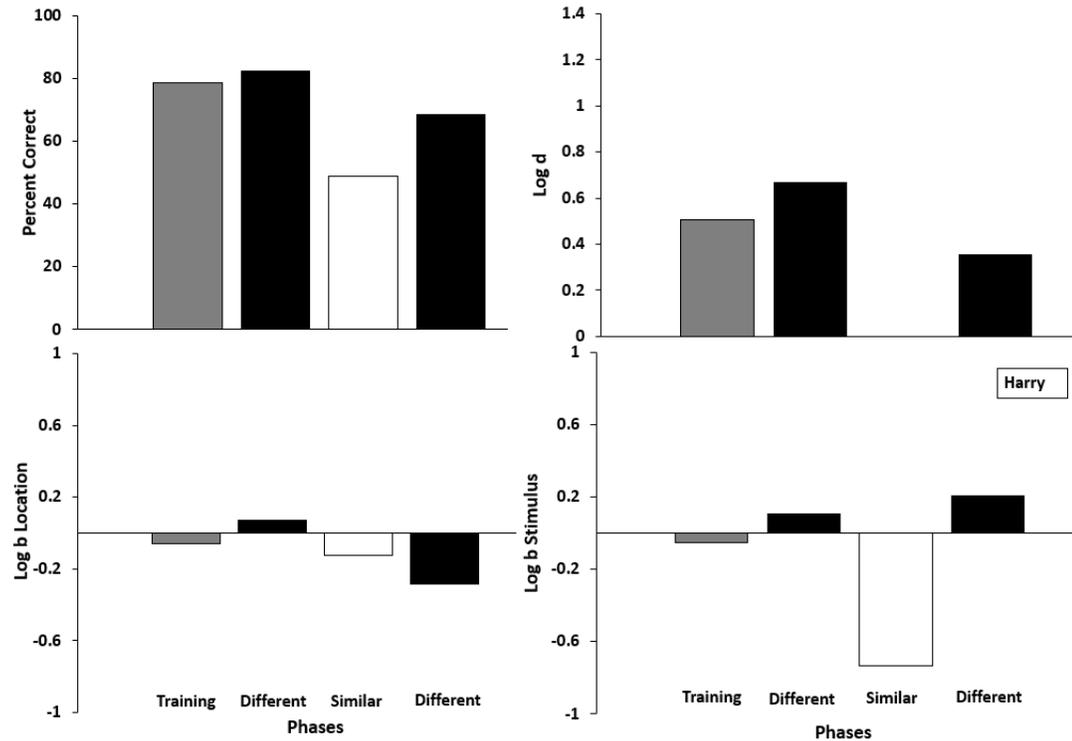


Figure 5. Depicts sum of percent correct, $\log d$, $\log b$ (stimulus), and $\log b$ (location) for the last seven sessions of each Phase, including Training, for Harry. Gray bars depict Training. Black bars depict Phases 1 and 3 with highly different stimuli and white bars depict Phase 2 with more similar stimuli. Percent correct and $\log d$ reliably decrease when more similar stimuli are presented in Phase 2. There is some variability in location bias, and a large difference in stimulus bias during Phase 2.

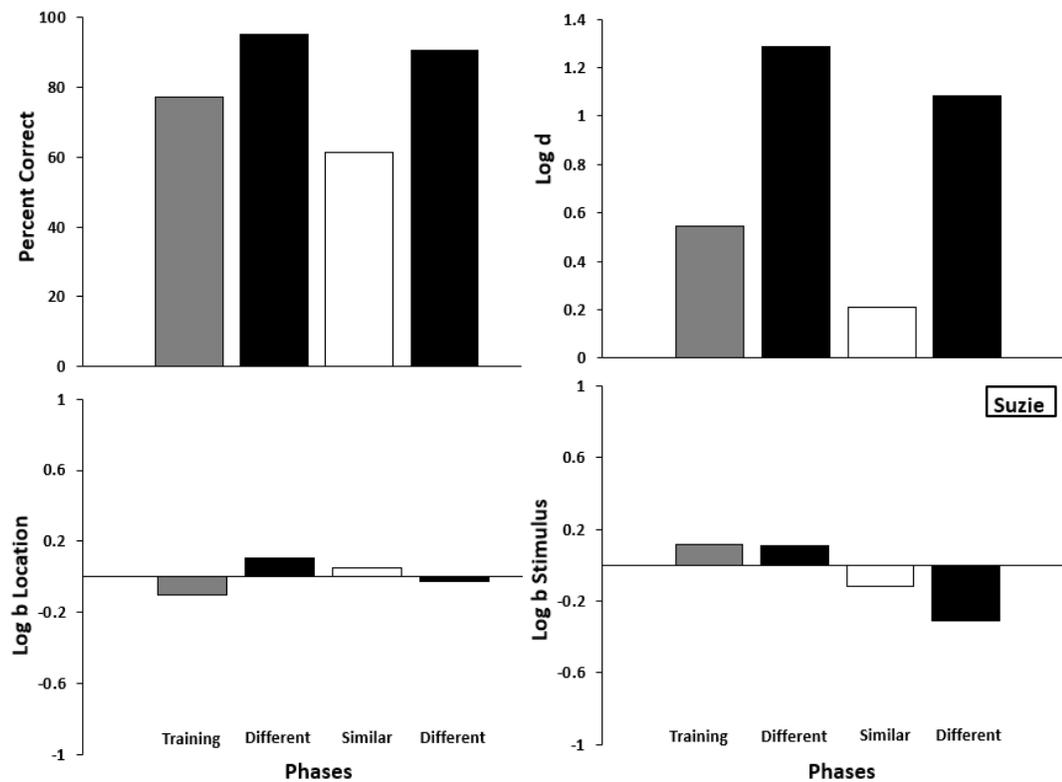


Figure 6. Depicts sum of percent correct, $\log d$, $\log b$ (stimulus), and $\log b$ (location) for the last seven sessions of each Phase, including Training, for Suzie. Gray bars depict Training. Black bars depict Phases 1 and 3 with highly different sample stimuli and white bars depict Phase 2 with more similar stimuli. Percent correct and $\log d$ reliably decrease when more similar stimuli are presented in Phase 2. There is some variability in stimulus bias, but no major differences in location bias across sessions.