AN ANALYSIS OF SOCIAL REFERENCING STIMULUS CLASSES AMONG CHILDREN WITH AUTISM

by

JAIME ANN DEQUINZIO

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree Doctor of Philosophy, The City University of New York

2009

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Claire L. Poulson_____________________

_____________________________ ____

Date Chair of Examining Committee

Dr. Maureen O’Connor_____________________

_____________________________ ____

Date Executive Officer

Dr. Claire L. Poulson____________________________
Dr. Nancy S. Hemmes___________________________
Dr. Bruce L. Brown_____________________________
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

AN ANALYSIS OF SOCIAL REFERENCING STIMULUS CLASSES

AMONG CHILDREN WITH AUTISM

by

Jaime Ann DeQuinzio

Adviser: Professor Claire L. Poulson

Social referencing consists of a child looking to the affective responses of an adult, which serve as discriminative stimuli for subsequent responding in the context of ambiguity or novelty. In this study, social referencing was defined as discriminative responding under a two-link chain. The discriminative stimulus for the first link was the presentation of experimental stimuli in the presence of which an observing response was required. Link 2 consisted of a conditional discrimination. The discriminative stimulus for the second link was an affective stimulus from one of two sets presented by the experimenter. Two experiments were conducted to teach children with autism to respond differentially to affective stimuli within the social referencing response chain, and to determine if differential responding generalized to similar stimuli. Experiment 1 attempted to evaluate discriminative responding to two sets of six affective stimuli in Link 2 of social referencing while participants encountered stimuli representing three types of tasks pictured in their activity schedules (i.e., handwriting, retrieving objects, and scripted social interaction). Because discriminative responding was not acquired by any of the three participants under that training paradigm, Experiment 2 was conducted. During this experiment, participants were seated at desks and were presented with stimuli that signaled social referencing. One affective stimulus from each of the two sets was used as the training stimulus. The remaining affective stimuli from the two sets were presented as probe stimuli to determine the extent to which each was part of an already established stimulus class. Participants were taught to engage in differential responding using manual guidance, differential reinforcement, and error correction. In the presence of an affective display from set 1 (e.g., smiling and nodding head), the correct response was a keep response in which the participants placed the stimuli in a bin on the desk. In the presence of an affective display from set 2 (e.g., shake head with eyebrows turned down), the correct response was a discard response in which the participants placed the stimuli in a garbage bin on the floor. Correct responding on training trials increased above baseline levels for all three participants with the systematic introduction of conditional discrimination training. Probe responding was inconsistent across the three participants, obviating analysis of stimulus class formation.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to the many people who supported me along this journey. I would like to acknowledge and thank my advisor Dr. Claire L. Poulson for her endless dedication and commitment to shaping my intellectual, scientific, clinical, and professional repertoires in a supportive and kind manner. I am also deeply indebted to Dr. Nancy S. Hemmes and Dr. Bruce L. Brown for their unvarying guidance and the prominent roles they played in developing my skills as a behavior analyst. I am profoundly grateful for the generosity that all three members of my committee have shown to me over the past six years. My sincerest thanks to Dr. Dawn B. Townsend who was the one who inspired me to start this journey, who has continually supported me throughout this process, and whom I persistently attempt to emulate in my academic and professional career. I am grateful to my parents for always stressing the importance of an education and for their compassionate support. Finally, to my husband for his thoughtful and confident expressions of encouragement and for never doubting that I could accomplish this goal.

Table of Contents

List of Tables

List of Figures

AN ANALYSIS OF SOCIAL REFERENCING STIMULUS CLASSES

AMONG CHILDREN WITH AUTISM

By

Jaime Ann DeQuinzio

Young children may learn about environmental events and to regulate their own behavior by observing the emotional or affective responses of others (Campos & Sternberg, 1981). *Social referencing* is an example of such an interaction in which children appear to observe the emotional reactions of others as a means of determining how to respond in ambiguous or novel situations. According to Feinman (1992) social referencing "(is) a process in which one person utilizes another person's interpretation of the situation to formulate his or her own interpretation of it…" (p. 4). Stated more plainly, social referencing involves the child looking to the facial expression and bodily gestures of another person to determine whether to approach or avoid the ambiguous stimulus conditions that initially set the occasion for the child to reference the adult.

Social referencing in infants has been studied by exposing infants to novel or ambiguous situations, such as a "visual cliff" (i.e., infants are placed on a Plexiglas surface that provides invisible support over an apparent drop), animated toys, or strangers. Studies generally contrast the effects of "positive emotional messages" and "negative emotional messages," provided by the mother on the behavior of the infant. When provided with positive emotional messages (e.g., joy), infants are *more* likely to cross the visual cliff, reach for toys, and approach strangers. When provided with negative emotional messages (e.g., fear), infants are *less* likely to cross the visual cliff, reach for toys, and approach strangers (Feinman, 1992).

Behavior Analysis of Social Referencing

Social referencing may be explained as a behavioral chain of stimulus-response interactions (Schlinger, 1995). Each component response in a behavioral chain produces stimulus conditions that function as both a conditioned reinforcer for the previous response and as a discriminative stimulus for the following response (Leslie, 1996). The effectiveness of those stimuli as conditioned reinforcers and discriminative stimuli is dependent upon their predictive reinforcement value (Schuster, 1969).

The social referencing response chain may be described as a two-link chain. The first link includes exposure to an ambiguous or novel stimulus, which evokes an observing response defined as the child orienting towards or looking at an adult. In the social referencing literature, the stimuli that are presented to infants as a means of evoking an observing response to initiate the social referencing chain are characterized as ambiguous or novel (Feinman, 1992). These labels alone do not provide a means of understanding the function of these stimuli within the social referencing response chain. Alternatively, a behavior-analytic account may assist in understanding how such stimuli function to evoke a social referencing response chain. According to such a perspective, stimuli may function by predicting the availability of reinforcement conditions for observing (S+ stimuli) or by predicting extinction or punishment conditions for observing (S- stimuli) (Leslie, 1996). Therefore, stimuli that have a history of being predictive of all consequential possibilities (reinforcement, and extinction or punishment) are ambiguous with respect to function and might occasion social referencing. It should be noted that novel stimuli without a history of predicting any type of consequences might occasion social referencing, as well (Brim, Townsend, DeQuinzio, & Poulson, in press).

An observing response is a response that produces a discriminative stimulus (Catania, 1998). In the case of social referencing, the observing response of the child sets the occasion for an affective discriminative stimulus (i.e., a response, such as smiling, frowning, head nodding, head shaking) to be performed by the adult. The adult's response constitutes link 2 stimulus of the response chain. The affective stimulus serves two functions in the behavioral chain of events. It serves as the conditioned reinforcer that maintains the observing response, as well as the discriminative stimulus that evokes the subsequent differentiated response of either approach or avoidance in the presence of the ambiguous stimulus conditions. The affective stimuli become discriminative for approach or avoidance responding because they are predictive of reinforcement, extinction, or punishment.

Responding within the social referencing chain can be analyzed as differentiated responding. Differentiated responding is acquired through a procedure whereby S+ and S- stimuli are presented in random order, and responding in the presence of S+ stimuli is reinforced while responding in the presences of S- stimuli is extinguished. This procedure produces more frequent responding in the presence of S+ stimuli and less frequent responding in presence of S- stimuli (Terrace, 1963).

One differential response learned in social referencing is the observing response. In the presence of ambiguous stimuli, observing is reinforced by the production of an affective discriminative stimulus. Hence, ambiguous stimuli may serve S+ functions because they are predictive of reinforcement contingencies for observing. In the presence of non-ambiguous or standard stimuli, observing may not be reinforced by the production of affective discriminative stimuli. For this reason, non-ambiguous stimuli may serve S- functions because they are predictive of extinction or non-reinforcement for engaging in an observing response. Once observing occurs more frequently in the presence of ambiguous (S+) stimuli and less frequently in the presence of non-ambiguous (S-) stimuli, differential observing is demonstrated.

Another discrimination learned during social referencing occurs among affective stimuli. As mentioned previously, observing produces affective discriminative stimuli that are predictive of reinforcement, extinction, or punishment contingencies for subsequent differential responding in the form of approach or avoidance. Therefore, certain classes of adult-affective stimuli, such as smiling, head nodding, and pointing, may serve S+ functions for approach responding if such responding produces reinforcement. These same stimuli may simultaneously function as S- stimuli for avoidance responding if extinction or punishment conditions are produced for such responding. Using similar logic, classes of affective stimuli, such as frowning, a fearful expression, or head shaking, may serve S+ functions for avoidance responding and S- functions for approach responding. Differential approach and avoidance responding is demonstrated when both of the following conditions occur: a) the frequency of approach responding is higher in the presence of approach-affective stimuli (S+) and less frequent in the presence of avoidance-affective stimuli (S-), and b) the frequency of avoidance responding is higher in the presence of avoidance-affective stimuli (S+) and less frequent in the presence of approach-affective stimuli (S-).

In addition to differential responding, another outcome of discrimination training procedures is the formation of stimulus classes (Fields and Reeve, 2000). When different sets of stimuli are associated with differential contingencies that produce differential responding, stimulus classes may be formed. A stimulus class is defined by a set of stimuli that come to occasion a common response class (Haring, Breen, & Laitinen, 1989). There are a number of different stimulus classes that may be formed within the social referencing paradigm. One stimulus class may be formed by the different contingencies that operate in the presence of ambiguous stimuli that occasion observing. The class of ambiguous stimuli that evoke an observing response may share some common physical features or may be physically dissimilar. For example, ambiguous stimuli, such as the presence of a stranger, a loud noise, or the presence of a novel toy may all be members of the class of stimuli that evoke an observing response. A second class of stimuli is the affective stimuli that occasion reinforcement for engaging in approach responding. These stimuli may be comprised of affective responses such as, smiling, head nodding, and pointing. A third class of stimuli is the affective stimuli that occasion reinforcement for the occurrence of avoidance responses. These stimuli may be comprised of affective responses such as, frowning, gasping, or head shaking.

Conceptual Theories of Social Referencing

Theories constructed to explain social referencing present opposing views of its development. One model explains social referencing as a pre-programmed ability in infants to perceive and respond to their mother's emotions (Feinman, 1992). In contrast, Gewirtz and Pelaez-Nogueras (1992) explained social referencing as a learned process whereby maternal facial displays come to reliably predict or serve as discriminative stimuli for contingencies of approach or avoidance responses within ambiguous contexts.

Gewirtz & Pelaez-Nogueras (1992) detailed a study that supported the operant learning explanation of social referencing. Twenty pairs of mothers and infants, age 9-12 months, participated. Infants sat in booster chairs facing a puppet theatre and their mothers stood next to them. Experimenters presented infants with toys covered by white cloths through the puppet theatre to serve as ambiguous stimuli. Mothers presented initially neutral or meaningless maternal facial displays, such as palms-to-cheeks and fist-to-nose, when infants engaged in orienting responses following the presentation of the covered toys. For ten of the participant pairs, a fist-to-nose facial display signaled that reaching for a hidden toy would be followed by reinforcement (i.e., musical baby melody). Reaching for the toy in the presence of a palms-to-cheeks facial display resulted in punishment (i.e., loud, harsh sound). For the other half of the participant pairs, the function of the facial displays was reversed so that the fist-to-nose display signaled punishment and the palms-to-cheeks display signaled reinforcement. Infants reached for the covered toy in the presence of the maternal facial display that signaled reinforcement for this response, and not in the presence of the maternal facial display that signaled punishment for this response. The results demonstrated that infant reaching behavior was instrumental and had come under the control of the emotionally meaningless facial displays. The authors argued that a pre-wired ability to interpret maternal emotions was unnecessary in explaining social referencing because, in this case, arbitrary facial displays came to control approach or avoidance responses of infants within the social referencing paradigm.

Social Referencing and Autism

One deficit of children with autism is their failure to engage in social referencing (Sigman, Kasari, Kwon, & Yirmiya, 1992; Warreyn, Roeyers, & Groote, 2005). A possible contributing factor for these social-affective deficits is the failure of children with autism to orient toward or to observe naturally occurring stimuli (Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998). Affective stimuli, such as facial expression and bodily gestures, are stimuli that occur often within social interactions. The failure to orient toward affective stimuli may pose two possible problems for learning social responses. In the absence of an orienting response, the child has missed an opportunity to observe the affective stimuli of others and for these stimuli to function as both conditioned reinforcers for orienting toward the social partner, and discriminative stimuli for subsequent responding. Affective stimuli may serve as conditioned reinforcers for orienting responses within the response chain of a social interaction and in turn affect the future occurrence of orienting. If opportunities for observing affective displays of others are scarce, opportunities for learning how to respond in the presence of these stimuli are limited. Affective stimuli may serve S+ and S- functions by being predictive of reinforcement or extinction for subsequent responding. Children with autism who do not observe affective discriminative stimuli may not accordingly experience these reinforcement or extinction contingencies for discriminating affective stimuli. Consequently, responding to affective stimuli does not come under discriminative control.

Another variable that may contribute to the failure of children with autism to develop social-affective behavior is known as stimulus overselectivity. Children with autism may show stimulus overselectivity, in which they respond to only one component of a complex stimulus (Lovaas, Koegel, & Schreibman 1979). A social interaction is a complex stimulus in that many component stimuli are present, such as the verbal and non-verbal behavior of the individual, as well as the many physical dimensions of the individual (e.g., hair color, height, color of clothing, etc.) and the environment. If children with autism orient towards affective stimuli, but they respond to a component of that complex stimulus that is socially irrelevant (e.g., the color of the person's eyes), then they will not access reinforcement contingencies that promote orienting toward or observing social-affective responses in the future.

There has been little research focusing on the amelioration of social referencing deficits in children with autism. Perhaps the interpretation of social referencing as learned behavior provides a direction for developing treatments for these deficits in children with autism using operant learning interpretations and procedures similar to those used by Gewirtz and Pelaez-Nogueras (1992).

In an empirical analysis of social referencing, Brim (2002, unpublished dissertation), demonstrated that children with autism could be taught to engage in the components of a social referencing response chain, thus further supporting an operant learning paradigm of social referencing. A multiple-baseline-across- responses-experimental design was used to evaluate social referencing training (verbal prompts, manual guidance, and reinforcement) on the occurrence of social referencing responses across handwriting, verbal imitation, and gross motor imitation tasks. Participants were first taught to engage in an observing response (i.e., to orient toward and to look at the face of the experimenter) in the presence of ambiguous stimuli only. Ambiguous stimuli were variants of standard task materials and included items such as, a piece of chalk as a writing utensil and a paper bag as a writing surface presented for the handwriting task, two-syllable words presented on auditory recordings with strange noises in the background (e.g., the word “orange” played with a cough in the background) as the verbal imitation task, and gross motor video models presented with the model wearing animal masks or filmed upside down as the motor imitation task. Following observing in the presence of the ambiguous stimuli, the experimenter presented one of two affective discriminative stimuli: a smile and head nod or a frown and head-shake. In the presence of the smile and head nod (S1), participants were reinforced for engaging in a task completion response by either demonstrating a handwriting response, a verbal imitation response, or a gross motor imitation response. In the presence of the frown and head-shake (S2), participants were reinforced for engaging in a task termination response by either putting away the handwriting materials, placing the auditory device into a bin without engaging in the verbal imitation response, and by sitting down without engaging in the gross motor imitation response.

Results showed that all of the participants learned to engage in the observing response in the presence of the ambiguous stimuli and the subsequent conditional discrimination in the presence of the S1 and S2 affective displays. The experimenter then assessed the extent to which the participants showed *differential observing* in the presence of standard task materials by conducting a test in which presentations of standard task materials and previously established ambiguous task materials were interspersed. Only one of the participants demonstrated differential observing under these conditions. The other three participants were subsequently taught to discriminate the ambiguous and standard task materials using a discrimination training procedure in which they were reinforced for observing only in the presence of the ambiguous task materials and not in the presence of the standard task materials. This study suggests that children with autism can be taught social referencing using behavior analytic procedures.

Further research may attempt to answer some additional questions regarding social referencing training with children with autism. For instance, it may be useful to examine the extent to which children with autism may learn to respond to more than two affective stimuli within a social referencing paradigm, considering there are numerous affective topographies that may serve as discriminative stimuli for approach and avoidance responding. The current research consisted of two experiments attempting to determine how children with autism respond to two sets of affective stimuli within the social referencing response chain. Experiment 1 attempted to evaluate discriminative responding to two sets of six affective stimuli in link 2 of social referencing while participants encountered stimuli pictured in their activity schedules.

EXPERIMENT 1

In Experiment 1, activity schedules were used to allow participants to engage in typical academic and social tasks independently so that they would encounter ambiguous stimuli naturally as they moved from task to task. Activity schedules are a set of pictures or words that serve as discriminative stimuli for the engagement in a sequence of tasks (MacDuff, Krantz, & McClannahan, 1993). Component steps of engaging in activity schedules typically include turning pages of the schedule, obtaining materials, completing tasks or going to a designated work area, putting materials away, and returning to the schedule.

The purpose of Experiment 1 was to evaluate the extent to which social referencing training (i.e., manual prompts, verbal prompts, modeling, and reinforcement) was effective at teaching children with autism to a) engage in social referencing skills during independent tasks signaled by activity schedules; b) show differential responding in the form of task completion or task termination; c) establish conditional control by many affective stimuli from the two different sets; and d) demonstrate generalized differential responding in the form of task completion or task termination by responding to affective probe stimuli from the two sets. Two experimental designs were used in Experiment 1 to evaluate the conditional discrimination training procedure. A multiple-baseline-experimental design across three response categories was used, as well as a multiple-baseline-across-subjects experimental design.

Although there are two links in the social referencing response chain, this study focused only on link 2 (i.e. discrimination of affective stimuli). As discussed previously, observing during link 1 of the social referencing response chain comes under the control of ambiguous stimuli because of the predictive reinforcement value of those stimuli. This study did not attempt to establish stimulus control over observing by ambiguous stimuli. Again, the focal point of this experiment was to establish conditional control of task completion and task termination responding by affective stimuli during link 2 of the response chain.

Method

Participants

Three children with autism between the ages of 5 and 10 who attend the Institute for Educational Achievement (IEA) participated (Jeremy, Troy, and Adam). All have previously received diagnoses of autism from independent agencies. During the school day, the children received one-to-one instruction and small-group instruction by therapists and teachers trained in applied behavior analysis. All of the participants had attended IEA since the age of 3 and had learned to follow activity schedules. Following an activity schedule involves skills such as discriminating among pictures of task materials by independently obtaining those materials depicted in the schedule or by going to a location or person depicted in the schedule and engaging in a social interaction. Participants also displayed some oral language including requesting preferred items, labeling common objects, and initiating simple greetings with others. Participants imitated verbal models presented by others or presented on auditory devices (e.g., button-activated voice recorders or Language Masters). Participants had learned to respond to token reinforcement systems, to engage in eye contact when presented with a “Look” instruction or when their names were called, and to follow teacher directions. In addition, participants had instruction in various handwriting tasks such as connecting dots, tracing shapes, and tracing letters. Unfamiliar objects, events, or people did not evoke engagement in the components of the social referencing response chain.

Setting and Stimuli

Sessions took place in a classroom that was smaller than the participants' regular classroom where daily programming was conducted. The classroom was approximately 12'x 6' and contained 2 desks and 2 chairs, a small bookshelf, a large shelf holding various curriculum items, a coat hook where the coats and backpacks were hung, and cubby containing the student's lunch box. The experimenter and a research assistant were present in the classroom during all experimental sessions. The research assistant collected interobserver agreement data and served as the social interaction partner.

Activity schedules. Participants engaged in three separate tasks signaled by pictures in activity schedules. The tasks included small motor responses (i.e., handwriting), gross motor responses (i.e., retrieving items from different locations), and vocal responses (i.e., initiating a verbal interaction using an auditory script). Activity schedules were created from 5 x 8 inch photo albums and contained pictures of the tasks inserted into each page. Thirty-six pictures (and thus thirty-six trials) were included in the activity schedules including 12 pictures from each task type. The pictures were placed into the schedules in a randomized order so that the same trial type from a single response category did occur consecutively more than twice.

The following is a description of the pictures that were included in the activity schedules. The picture that signaled the handwriting task was of a closed plastic bin that contained the handwriting materials. The pictures that signaled the retrieving items response were of various common items in their typical locations throughout the classroom (e.g., picture of a lunchbox in a cubby or a book on the bookshelf). The pictures that cued the vocal response were of familiar teachers who served as the partner for the scripted verbal interaction response. Examples of the scripted interactions include statements such as, “How are you?” “Hug me” and “High five.”

School Stimuli. School stimuli were used during the pre-training condition only and consisted of stimuli that were part of the participants’ regular programming and thus had consistent reinforcement histories. School stimuli for the handwriting task were a yellow # 2 pencil as the writing utensil, and an 8 ½ x 11 inch piece of white paper as the writing surface. The handwriting trials included connecting two dots separated by 5 inches. For the retrieving objects task, common objects around the school building such as a tissue box, books, and a lunch box were used. For the scripted verbal interaction task, statements were recorded on button activated voice recorders in a conversational volume and tone, similar to that used in the participants’ current language programs. Table 1 provides a list of school stimuli used for each response type.

Experimental Stimuli. Experimental stimuli were used during baseline and social referencing training. The experimental stimuli were used to occasion an observing response that produced experimenter affective stimuli and included materials that were different from the participants’ ordinary school stimuli. For the handwriting task, materials other than the standard # 2 pencil and sheet of paper were placed in the bin. For example, the plastic bin contained a jumbo poster marker and a piece of aluminum foil. For the retrieval task, additional stimuli were added to ordinary school stimuli. For example, a feather boa was placed around the soap dispenser, and metallic streamers were placed around the participants’ lunch boxes. For the vocal task, the scripted verbal statements were recorded in a strange voice (e.g., high pitch) or with a strange background noise, such as a fire engine or repetitive sneezing. Table 2 lists the experimental stimuli that were used for each response type.

Affective stimuli. Following observing during social referencing training, the experimenter presented affective stimuli from one of two stimulus sets (see Table 3). One set of affective stimuli was discriminative for task completion (approach) responses. These stimuli will be referred to as Set 1 (S1) and included the following six responses performed by the experimenter: nodding head while smiling with teeth exposed, raising fists with thumbs extended up while smiling with mouth closed, extending arm toward the stimulus while pointing and moving the index finger, repetitively tilting head toward the stimulus while raising eyebrows, leaning the body toward the stimulus with palms up moving in a "push along" motion, and moving head back and forth from the stimulus to the learner with mouth open. Another set of affective stimuli was discriminative for task termination (avoidance) responses. These stimuli will be referred to as Set 2 (S2) and included the following six responses performed by the experimenter: shaking head left to right with eyebrows turned down, placing hands over mouth, waving hands in a criss-cross motion with eyes wide open, turning head and upper body away from the stimulus with eyes closed, gasping and stepping away from the stimulus, and frowning with bottom lip protruding out.

Stimuli from S1 and stimuli from S2 were further divided into training and probe stimuli. Training affective stimuli were followed by the use of the training package. Probe affective stimuli were not followed by the use of the training package. S1 probe stimuli were used to assess the extent to which the task completion response generalized from the S1 training stimuli to the S1 probe stimuli. In addition, S2 probe stimuli were used to assess the extent to which task termination generalized from the S2 training stimuli to the S2 probe stimuli. Probe stimuli for S1 and S2 were assigned to each participant by randomly drawing the stimuli, without replacement, from a pool containing all six stimuli from each set.

Dependent Measures and Response Definitions

In this study, social referencing was composed of a two-link chain. The discriminative stimulus for the first link was the presence of experimental stimuli. In the presence of experimental stimuli, the required response was an observing response (defined below), after which link 2 of the chain came into effect. The discriminative stimulus for the second link was an affective stimulus presented by the experimenter. In the presence of stimuli from S1, the correct response was task completion. In the presence of stimuli from S2, the correct response was task termination.

Observing was defined as orienting towards, and looking at the experimenter's face within 2s of the presentation of the experimental stimuli. Because observing had to occur on every trial in order to produce the discriminative stimulus for link 2, the experimenter used manual guidance to prompt the observing response if it did not occur within 2 s during both baseline and treatment conditions. Therefore, an evaluation of the acquisition of observing from the baseline condition to the treatment condition as a function of the independent variable was not possible in this study.

The occurrence of the observing response in the presence of the experimental stimuli resulted in the presentation of affective discriminative stimuli from one of the two stimulus sets described above. In the presence of an affective display from S1, participants were required to engage in a task completion response by either a) completing the handwriting task by connecting the dots, b) locating and retrieving the items depicted in the picture and delivering the item to a bin on the other side of the room, and c) locating the social interaction partner depicted in the picture and engaging in the verbal interaction by imitating the scripted statement produced from the auditory device. In the presence of an affective display from S2, participants were required to engage in a task termination response by either a) closing the folder containing the handwriting materials without connecting the dots, b) placing the item back on the shelf without delivering it to the bin on the other side of the room and c) locating the social interaction partner depicted in the picture, but returning to the activity schedule without imitating the auditory statement and engaging in the verbal interaction. These responses were evaluated during baseline and social referencing training by measuring the percentage of trials with a correct conditional discrimination. To evaluate the component responses of the conditional discrimination, percentage correct task completion and percentage correct task termination were also summarized separately.

Experimental Design

Two experimental designs were used in Experiment 1. A multiple-baseline-across-responses design was used to evaluate the effectiveness of the training across the three response types for each participant. A multiple-baseline-across-participants design was also used to evaluate the effectiveness of the training across the three participants for the handwriting response category.

Sessions consisted of 36 trials. Experimental stimuli from each response type were randomly paired with stimuli from both sets of affective stimuli for each trial so that no one experimental stimulus was repeatedly paired with any one affective stimulus. Following this random pairing of link 1 and link 2 stimuli, four orders of the 36 trial presentations were prepared using a random number generator computer program. This was to assure that no pattern of responding occurred based on the order of stimulus presentations and that no consecutive presentations of probe stimuli occurred. Probe stimuli were chosen for each participant by randomly drawing from the 10 total stimuli, without replacement.

Procedure

Pre-baseline Assessment of Affect Discrimination. A pre-baseline assessment was conducted to determine the extent to which participants could first discriminate among the affective stimuli used during conditional discrimination training in link 2 of the social referencing response chain. This assessment was conducted in one 15-minute session for each participant. During this assessment, participants were instructed to imitate all twelve of the affective stimuli presented in Table 3. Three trials of each of the twelve stimuli were presented during the assessment in random order for a total of 36 trials. The experimenter presented an instruction, (i.e., "Do this") followed by one of the affective stimuli. Imitating the affective models provided evidence that participants could discriminate among them, bolstering an argument for the occurrence of generalized task termination or task completion responding. If participants did not imitate the affective stimuli, then imitation training would have been provided. However, all of the participants imitated all twelve of the affective displays.

General Procedure. Session durations were approximately 30-40 minutes and were conducted once each day, four days per week. At the start of each pre-training, baseline, and training session, participants were instructed to "Follow the activity schedule." Activity schedules were placed on a desk in front of the participant. A plastic bin containing folders with the handwriting materials for each trial were placed on a shelf next to the desk. The experimenter stood approximately 1-2 feet behind and slightly to the left or right side of the participant. Research assistants served as recipients of the verbal interactions and also collected interobserver agreement data. The auditory devices were kept on another table in the classroom so that the experimenter could access each one easily for the social interaction trials. The experimenter picked up the auditory device and activated the script by pressing the button once the participant located the interaction partner. Items for the retrieval task were located in various locations in the classroom (e.g., on shelves, hanging on hooks, placed in cubbies). The experimenter held the token board so that the participant could see it, or placed the token board on the desk in front of the participant when delivering tokens. Once the specified number of tokens was earned for correct responding, participants were provided with edibles or with a short break to access preferred activities.

Pre-training. Pre-training consisted of two sessions each with 36 trials (12 trials from each response category) in which participants followed activity schedules including standard school stimuli only (see Table 1). The participants already had established learning histories with the handwriting, retrieving objects, and social interaction tasks presented using school stimuli. Nevertheless, because the experimenter did not have a history of instruction with the participants, pre-training allowed the experimenter to establish instructional control, and to establish the format within which the participants would work throughout the course of Experiment 1. Participants were instructed to follow the activity schedule. Manual guidance was used to prompt participants to engage in responses related to following activity schedules, such as turning the pages in the schedule, returning to the schedule following the completion of the task, and completing tasks cued by pictures in the schedule. For the handwriting task, the experimenter used manual guidance to assist the participant in walking to the shelf, removing a folder from the plastic bin, placing it on the desk, opening the folder, engaging in the handwriting task, putting the materials back in the folder, and placing the folder in the desk. For the retrieval task, the experimenter used manual guidance to assist the learner in walking to and locating the item in the classroom picking it up and delivering it to another location in the room. For the verbal interaction task, the experimenter used manual guidance to assist the learner in locating the interaction partner and activated the auditory device once the learner approached the partner. The experimenter also provided verbal prompts to assist the learner in imitating the auditory prompt if needed. Reinforcement with tokens was provided for responding throughout these sessions.

Baseline. During baseline, a total of 36 trials was conducted. These trials consisted only of experimental stimulus materials listed in Table 2, as opposed to the standard school stimuli used during pre-training. Twelve trials from each of the three response categories were conducted. Participants were instructed to follow activity schedules. Similar to procedures used during pre-training described above, during baseline the experimenter manually prompted responses related to following activity schedules if needed. As opposed to pre-training however, the experimenter did not prompt engagement in the handwriting, retrieving objects, and social interaction tasks during baseline. For example, the experimenter used manual guidance to assist participants in turning the pages of the schedule, obtaining the handwriting folders from the shelf, or walking to the location of the item or person pictured in activity schedule, and returning to the schedule between trials, but did not prompt the participant to connect dots on the handwriting worksheets, deliver items to locations, or imitate the scripted statements. Reinforcement was provided with tokens for responding related to following the activity schedule only. If participants engaged in the handwriting, retrieving objects, or verbal interaction tasks during baseline, no programmed contingencies were used. The experimenter removed the stimuli to terminate the trial after a 5-s interval and manually prompted the participant back to the activity schedule to begin the next trial. Reinforcement was not provided for engaging in the handwriting, retrieving objects, or social interaction tasks.

Verbal prompts and manual guidance were used to prompt an observing response in the presence of experimental stimuli during this condition following a 2-s delay so that the opportunity for link 2 responding was always available. Following observing, the experimenter presented an affective discriminative stimulus from either S1 or S2, but did not provide additional prompts to assist in the performance of the subsequent task, completion response, or task termination response. If a task completion or task termination response occurred following the affective display, no programmed contingencies were used.

Conditional Discrimination Training. During conditional discrimination training, 36 trials were conducted using only experimental stimuli (Table 2). Twelve trials from each of the three response categories were conducted. The same procedure used in baseline was used to assist participants in completing responses related to following activity schedules if needed. Manual guidance was used to prompt an observing response on each trial.

Following the observing response, the experimenter presented affective discriminative stimuli from either S1 or S2. Six trials from each affective stimulus set were presented including four training and two probe trial from each set. If the experimenter presented an affective discriminative stimulus from S1 during a training trial, she used manual guidance and modeling (0-s delay) to assist the participant in engaging in a task completion response. Reinforcement in the form of a token and praise (e.g., “Good writing when I smile at you”) was provided. The 0-s delay was used for the first three treatment sessions for each response type (handwriting, retrieving items, and verbal interaction). Starting on the fourth session of treatment for each response type, a 3-s delay was used.

If the experimenter presented an affective discriminative stimulus from S2 during a training trial, she used manual guidance and modeling (0-s delay) to assist the participant in engaging in a task termination response. Reinforcement in the form of a token and praise (e.g., "Good putting it away when I frown") was provided. The 0-s delay was used for the first three treatment sessions for each response type (handwriting, retrieving items, and verbal interaction). Starting on the fourth session of treatment for each response type, a 3-s delay was used. These conditional discrimination trials consisted of 4 training and 2 probe trials. See Figure 1 for a display of the trial configurations used during baseline and conditional discrimination training in Experiment 1.

For Jeremy, two variations of the above training procedure were attempted for the first response category during training (i.e., handwriting). For sessions 30-34, the handwriting trials were presented in a 12 trial block at the beginning of the session including six completion trials and six termination trials (Saunders & Spradlin, 1989). Probe trials were eliminated. The number of affective discriminative stimuli was reduced to two; one serving as the discriminative stimulus for the completion trials (i.e., arm extended while repetitively moving index finger toward the experimental stimulus) and one serving as the discriminative stimulus for the termination trials (i.e., waving hands in a criss-cross motion with eyes wide open). The training procedure was implemented using a 0s delay for the first three completion or termination trials, followed by the use of a 3s delay for the remaining three completion or termination trials. Following this procedural variation, a 3-s delay was used for sessions 35 and 36.

Interobserver Agreement

Research assistants, who are employed as teachers at the IEA and trained in Applied Behavior Analysis independently scored the responses of the participants for both training and probe trials during a minimum of 50% of baseline and training sessions. Interobserver agreement (IOA) was obtained on the percentage of trials with a correct conditional discrimination. Interobserver agreement was calculated on a point-by-point basis and by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100 to obtain a percentage of IOA. For Jeremy, the mean percentage of IOA for the small motor response category during baseline was 100%, and during training was 99%; for the gross motor response category the mean percentage of IOA during baseline was 96%; for the vocal response category, mean percentage of IOA during baseline was 99%. For Troy, the mean percentage of IOA for the small motor response category during baseline was 97%, and during training was 96%; for the gross motor response category the mean percentage of IOA during baseline was 95%; for the vocal response category, mean percentage IOA during baseline was 98%. For Adam, the mean percentage of IOA for the vocal response category during baseline was 94%, and during training was 100%; for the gross motor response category the mean percentage IOA during baseline was 97%; for the small motor category, mean percentage IOA during baseline was 99%.

Research assistants also independently scored the responses of the experimenter to determine the integrity of the independent variable. Interobserver agreement was calculated for the percentage of trials with an accurate presentation of treatment components during baseline and conditional discrimination training. Interobserver agreement was calculated on a point-by-point basis and by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100 to obtain a percentage of IOA. For Jeremy, mean IOA was 100% during baseline and 100% during training for the small motor response category; mean IOA during baseline was 100% for the gross motor response category and 100% for the vocal response category. For Troy, mean IOA was 100% during baseline and 100% during training for the small motor response category; mean IOA during baseline was 100% for the gross motor response category and 100% for the vocal response category. For Adam, mean IOA was 100% during baseline and 100% during training for the vocal response category; mean IOA during baseline was 100% for the gross motor response category and 100% for the small motor response category.

Results and Discussion

Figures 2-11 display the results for Experiment 1. Figure 2 displays the percentage of handwriting trials with a correct conditional discrimination across the three participants. Because the conditional discrimination training procedure used was not effective at teaching Jeremy to respond differentially (i.e., engage in correct task completion and task termination responses) in the presence of affective stimuli, it was not implemented across the remaining two participants in the across-subjects design. Therefore, the across-subjects design was terminated and the training procedure was instead implemented for all three participants in a multiple-baseline-across-responses-experimental design.

Because of Jeremy's difficulty learning the discrimination with the original training procedure, variations of the training procedure were attempted for this participant. These variations included a reduction in the number of affective stimuli used as discriminative stimuli for differential task completion and task termination responding, the elimination of probe trials, and the handwriting training trials presented in blocks at the beginning of each session. In spite of these modifications, Figures 3-5 show that they were not effective in teaching Jeremy to respond differentially to affective stimuli. Figure 3 shows the percentage of training and probe trials with a correct conditional discrimination across the small motor, vocal, and gross motor response categories. Responding remained variable during the baseline and the first three training conditions, demonstrating no change in trend with the implementation of training. Responding increased to 75% in the final training condition. Nevertheless, correct responding occurred mainly on the trials that required a completion response; a response already in Jeremy's repertoire. Figures 4 and 5 demonstrate how those responses were distributed between the completion and termination trials. Although correct responding on completion trials increased to 100%, correct responding on the termination trials remained similar to responding in the original training phase prior to the modifications. One possible explanation for the failure of acquisition of the termination response is the stimulus control that activity schedules may have exerted over completion responding. That is, it may have been extremely difficult for Jeremy to engage in termination responding in the presence of stimuli that have a strong history of reinforcement for completion responding.

Similar results were obtained for Troy (see Figures 6-8) during training on the small motor response category and for Adam (see Figures 8-11) during training on the vocal response category. In addition, for two of the three participants, prompting the termination response, or restricting access to materials pictured in the activity schedules, served a discriminative function for disruptive behavior. Experiment 2 was conducted in an attempt to eliminate control by activity schedules, to establish conditional control by affective stimuli, and to allow participants access to experimental stimuli on all trials.

EXPERIMENT 2

The results of Experiment 1 indicated that the participants did not learn to respond to the two sets of affective stimuli using that training paradigm. Therefore, Experiment 2 was conducted as a means of evaluating participants' responding to affective stimuli using a simpler paradigm. Participants did not follow activity schedules during this experiment. Instead, the conditional discrimination of affective stimuli was taught using experimenter-presented trials within a choice paradigm. It was expected this analysis might be useful in determining the extent to which relations among the two sets of affective stimuli had existed prior to the experiment. If the affective stimuli were not functioning as separate classes, that is, if the participants did not respond differentially to the affective stimuli from S1 and S2 during baseline, then class formation might still be demonstrated experimentally, using conditional discrimination training procedures. Determining the extent to which relations already existed among the affective stimuli, or the extent to which stimulus relations might be formed, may provide implications for teaching such relations among affective stimuli within complex social paradigms such as social referencing to children with autism. Finally teaching the discrimination within a choice paradigm might eliminate disruptive behavior evoked by restricted access to items by allowing participants to manipulate items presented on each trial.

Hence, the purpose of Experiment 2 was to evaluate the extent to which conditional discrimination training conducted using manual guidance, differential reinforcement, and error correction within a choice paradigm would be effective in teaching children with autism to a) demonstrate differential responding in the form of a keep response or a discard response; b) establish conditional control by two affective training stimuli; and c) demonstrate generalized differential responding to the remaining affective probe stimuli from the two sets. The conditional discrimination training procedure was evaluated using an across- subjects-experimental design.

Method

Participants and Setting

The same participants from Experiment 1 participated in Experiment 2 and sessions took place in the same classroom used in Experiment 1.

Stimuli

The stimuli used to occasion social referencing in Experiment 2 during baseline and treatment conditions consisted of various types of common objects, such as small ceramic figures, various types of containers (i.e., gift bags, small boxes, jars), a disposable camera, a wallet, various types of picture frames and photo albums, key chains, and other common household objects (e.g., bowls, sponges, brushes, etc) . Items such as these were considered common objects that participants may have encountered in the past, but might have different contingencies associated with handling them in different environments (e.g., it may be acceptable to handle a picture frame in one's own house, but not in a neighbor's house). The total number of objects was 240. Because conditional control by experimenter affective stimuli was of most importance, a different object was used on each trial so that no single object inadvertently acquired stimulus control over responding. Once all 240 objects were used (i.e., following 10 sessions), the previously used objects were used again. At the start of each trial, the experimenter non-systematically removed an object from a bag on the floor and placed it on the desk in front of the participant. The objects were used to occasion social referencing that, in turn, produced experimenter affective stimuli for link 2 of the response chain.

Affective Stimuli. Following observing, the experimenter presented affective stimuli from one of two stimulus sets (see Table 4). Two affective stimuli that were used in Experiment 1 were eliminated from Experiment 2 (i.e., reduced from 12 to 10). This was done to reduce the total number of trials from 36 in Experiment 1 to 24 in Experiment 2, thus shortening session length. One set of affective stimuli was discriminative for the keep response (i.e., placing the object in a bin on the desk). These stimuli will be referred to as Set 1 (S1) and included the following five affective responses performed by the experimenter: nodding head while smiling with teeth exposed, raising fists with thumbs extended up while smiling with mouth closed, extending arm toward the stimulus while pointing and moving the index finger, repetitively tilting head toward the stimulus while raising eyebrows, and leaning the body toward the stimulus with palms up moving in a "push along" motion.

Another set of affective stimuli was discriminative for the discard response (i.e., placing the object in a garbage bin located to the left of the desk on the floor). These stimuli will be referred to as Set 2 (S2) and included the following five responses performed by the experimenter: shaking head left to right with eyebrows turned down, placing hands over mouth, waving hands in a criss-cross motion with eyes wide open, turning head and upper body away from the stimulus with eyes closed, and gasping and stepping away from the stimulus.

Stimuli from S1 and stimuli from S2 were further divided into training and probe stimuli. The presentation of two training stimuli (i.e., one from S1 and one from S2) were followed by the use of the conditional discrimination training procedure during the treatment condition. Nodding head with an open smile was chosen as the training stimulus from set S1 and shaking head left to right with eyebrows down was chosen as the training stimulus from set S2. The remaining 8 stimuli were used as probe stimuli. The presentation of the probe stimuli was not followed by the conditional discrimination training procedure during the treatment condition. S1 probe stimuli were used to assess the extent to which the keep response generalized from the S1 training stimulus to the S1 probe stimuli. S2 probe stimuli were used to assess the extent to which the discard response generalized from the S2 training stimulus to the S2 probe stimuli.

Dependent Measures and Response Definitions

During Experiment 2, the experimenter placed an object on the desk in front of the participant. This was the discriminative stimulus for the first link (i.e., observing). In the presence of the object, the required response was an observing response (defined below), after which link 2 of the chain came into effect. The discriminative stimulus for the second link was an affective stimulus presented by the experimenter. In the presence of S1, the appropriate response was a keep response. In the presence of S2, the appropriate response was a discard response.

Observing was defined as orienting towards, and looking at the experimenter's face within 2s of the presentation of the experimental stimuli. Because the observing response had to occur on every trial in order to produce the discriminative stimulus for link 2, the experimenter used manual guidance to prompt the observing response if it did not occur within 2s during both baseline and training conditions. Therefore, an evaluation of the acquisition of the observing response from the baseline condition to the treatment condition as a function of the independent variable was not conducted.

The occurrence of the observing response in the presence of the object resulted in the presentation of affective discriminative stimuli from one of the two stimulus sets described above. In the presence of S1, participants were required to engage in a keep response by placing the object in a bin on the desk within 3s of the presentation of the affective display. In the presence of S2, participants were required to engage in a discard response by placing the object in a garbage bin on the floor next to the desk within 3s of the presentation of the affective display. Differential responding to S1 and S2 stimuli was evaluated by measuring the percentage of training and probe trials with a correct conditional discrimination. The keep and discard responses were also evaluated separately. Acquisition of the keep response was evaluated by measuring the percentage of training and probe trials during which a correct response occurred in the presence of affective stimuli from S1. Moreover, acquisition of the discard response was evaluated by measuring the percentage of training and probe trials during which a correct response occurred in the presence of affective stimuli from S2. If a response other than a keep or discard response occurred within 3s of the affective display, it was scored as “neither” and those data were summarized as the percentage of neither responding that occurred on both S1 and S2 trials. Finally, conditional probabilities were calculated to determine the percentage of correct responses that occurred given a response during the treatment condition.

Experimental Design. A multiple-baseline-across-subjects-experimental design was used to determine the extent to which conditional discrimination training, consisting of manual guidance, differential reinforcement, and error correction was effective at teaching children with autism to engage in differential keep and discard responses within link 2 of the social referencing paradigm. The training procedure was introduced successively across subjects after a baseline measure was obtained for each subject.

Procedure

General Procedure. Session durations were approximately 20-30 minutes and sessions were conducted once each day four days per week. A total of 24 trials per session were conducted, consisting of 16 training trials (i.e., 8 S1 trials and 8 S2 trials) and 8 probe trials (i.e., 4 S1 trials and 4 S2 trials). Participants were seated at a desk. The conditional discrimination was taught using a choice paradigm. A bin for the keep response was placed in a corner of the desk where the participants were seated. A garbage bin for the discard response was placed on the floor next to the desk. The experimenter sat on the opposite side of the desk facing the participant. At the start of each trial, the experimenter non-systematically removed an object from a large bag on the floor and placed it on the desk in front of the participant. If the learner engaged in an observing response within 2s of the object being placed on the desk, the experimenter presented either the S1 or S2 affective display. If the learner did not engage in an observing response within 2s of the object being placed on the desk, the experimenter used manual guidance to assist the participant in engaging in an observing response. Following this, the experimenter presented either the S1 or S2 affective display. The trial was terminated 3s following the presentation of S1 or S2 and a 3-5-s inter-trial interval commenced.

Baseline. During baseline, the conditional discrimination training procedure was not used. The experimenter placed an object on the desk to begin the trial. Following observing, the experimenter presented either the S1 or S2 affective display. If a keep or discard response occurred within 3s of S1 or S2, the experimenter provided feedback to the participant (i.e., “You kept the keychain” or “You put the keychain in the garbage”) regardless of whether it was a correct response, and the 3-5s inter-trial interval commenced. If the participant engaged in a neither response (i.e., did not respond or engaged in some other response such as manipulating the object) the object was removed 3s after the affective display was presented, the trial was terminated, and the 3-5-s inter-trial interval commenced. Participants earned tokens during the inter-trial interval for on-task behavior.

Conditional Discrimination Training. The conditional discrimination training procedure, consisting of manual guidance, differential reinforcement, and error correction was used in the treatment condition following the presentation of the S1 and S2 training stimuli (i.e., nodding head with an open smile and shaking head with eyebrows down). Following observing on the training trials, the experimenter presented either the S1 or S2 affective discriminative stimulus for 3s. Then, the experimenter waited 3s. If a correct keep response or discard response occurred, the experimenter provided token reinforcement and praise while pointing to the keep bin on the desk (e.g., "Good keeping it here when I smile and nod my head") or the garbage bin on the floor (e.g., "Great putting it in the garbage when I shake my head"). If an incorrect response occurred, or a neither response occurred within the 3s, the experimenter used manual guidance to prompt the correct response (i.e., place the object in the bin on the desk in the presence of S1 or place the object in the garbage bin in the presence of S2). Following this correction procedure, the trial was rehearsed until the participant engaged in the correct response in the absence of prompts. For the first few sessions of treatment, these rehearsed trials were followed by token reinforcement and praise. For the remainder of the treatment condition, only correct responses that occurred following the first presentation of S1 or S2 were followed by token reinforcement and praise.

Additional Manipulations Following Termination of Experiment 2

Modified Procedure for Jeremy and Troy. After the experimental design was implemented, a modified procedure was implemented for Jeremy and Troy as an attempt to reduce variability in responding and to improve performance in the few remaining sessions (i.e., sessions 47-50). The modified procedure was implemented because Jeremy and Troy did not reach a level of responding consistent with Adam's after the implementation of the of the conditional discrimination training procedure. To keep the overall rate of reinforcement high for the sessions, all probe trials were eliminated and only training trials were presented. The same prompting and correction procedure described above was used with the addition of modeling and instructions. At the beginning of the session, the experimenter modeled the keep and discard response and provided a verbal instruction. Specifically the experimenter said, "When I smile and nod my head (presented the S1 affective display), do this (modeled the keep response). When I shake my head (presented the S2 affective display), do this (modeled the discard response)."

Reinforcement for Responding on Keep Probe Trials for Adam. Additional analyses were also conducted for Adam following the termination of Experiment 2. Following the conditional discrimination training condition for Adam, the experimenter reinforced responding in the presence of one affective stimulus from the S1 category (i.e., palms up, moving hands in a push along motion). The three remaining probes in the keep category were not reinforced and continued to be presented as a means of evaluating generalized responding. During this condition, the conditional discrimination training procedure continued to be implemented on training trials.

Interobserver Agreement

Research assistants, who were employed as teachers at the IEA and trained in Applied Behavior Analysis independently, scored the responses of the participants for both training and probe trials during a minimum of 40% of baseline and conditional discrimination training sessions. Interobserver agreement (IOA) was obtained on the number of trials with a correct conditional discrimination. Interobserver agreement was calculated on a point-by-point basis and by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100 to obtain a percentage of agreement. For Adam, the mean percentage of IOA during baseline was 100%, and during training was 97%. For Jeremy, the mean percentage of IOA during baseline was 100%, and during training was 99%. For Troy, the mean percentage of IOA during baseline was 100%, and during training was 100%.

Research assistants also independently scored the responses of the experimenter to determine the integrity of the independent variable. Interobserver agreement was calculated for the percentage of trials with an accurate presentation of treatment components during baseline and conditional discrimination training. Interobserver agreement was calculated on a point-by-point basis and by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100 to obtain a percentage of IOA. For Adam, mean IOA was 100% during baseline and 100% during training. For Jeremy, mean IOA was 100% during baseline and 100% during training. For Troy, mean IOA was 100% during baseline and 100% during training.

Results and Discussion

Figures 12-18 display the results for acquisition of the conditional discrimination for all three participants during Experiment 2. A description and explanation of responding on training trials is first provided followed by a description and explanation of responding on probe trials for all three participants. The additional analysis of Adam's S1 probe responding is displayed in Figures 19 and 20.

Analysis of Responding on Training Trials for Adam, Jeremy, and Troy

Figure 12 displays the percentage of trials with a correct conditional discrimination in the presence of both training and probe S1 and S2 stimuli across the three participants. With respect to training trials (closed circles), none of the participants responded differentially to the S1 and S2 affective displays during baseline. Responding by all three participants increased systematically above baseline levels with the successive introduction of the treatment condition (i.e., manual guidance, differential reinforcement, and error correction).

For Adam, responding on training trials increased to 100% by session 14 and remained high (i.e., between 75% and 100%) for the remainder of that condition. For Jeremy, responding on training trials was variable, although the data path does demonstrate a change in level and an increasing trend from the first day of treatment to the last day of treatment. Jeremy's highest level of responding was reached during session 34 when he responded correctly on 63% of the training trials presented. Jeremy's data remained variable after session 34 and stabilized during the last 4 sessions of this condition at 56%. Troy's responding on training trials during the treatment condition was also variable. Troy's highest level of responding during this condition was reached during session 46 when he responded correctly on 81% of the training trials presented. Unfortunately, responding dropped back to 44% on the final day of treatment.

It may be concluded that the treatment procedure consisting of manual guidance, differential reinforcement, and error correction was effective for teaching Adam to respond differentially to S1 and S2 affective stimuli, but was not effective for teaching differential responding to Jeremy and Troy. Sessions 47-50 illustrate that a modified procedure (i.e., the elimination of probe trials and the addition of modeling and instructions) might have been helpful in increasing correct responding on training trials for Troy, but not for Jeremy.

Evaluating how responding was distributed among the three response possibilities (i.e., keep, discard, and neither) provides a more detailed analysis of the component responses of the conditional discrimination. Figure 13 shows the percentage of training and probe trials during which the keep response occurred across the three participants during each session and Figure 14 displays the percentage of training and probe trials during which the discard response occurred across the three participants during each session. The training data in these figures show that Adam learned to engage in the keep response in the presence of S1 and the discard response in the presence of S2 during treatment. Acquisition of the discrimination is further supported by the data in Figures 15 and 16 that display the percentage of trials with a neither response for Adam during baseline and treatment. The percentage of both S1 and S2 training trials with a neither response was at 100% during all baseline sessions and then decreased to 0% by the fourth session of treatment. Neither responding remained low throughout the course of treatment, representing good discriminative control by the S1 and S2 affective stimuli. Additionally, Figures 17 and 18 illustrate that when Adam did engage in a response during treatment, the probability that the response was correct was between 80% and 100% throughout the treatment condition on both S1 and S2 trials.

Unfortunately, the same acquisition pattern cannot be reported for Jeremy and Troy. For Jeremy, responses were distributed unequally among keep, discard, and neither. Specifically, keep responding remained fairly low throughout the treatment condition with only one session during which correct responding was higher than 50% (Figure 13). On the other hand, correct responding for Jeremy on discard training trials reached 100% on sessions 38, 40, 42, and 45 (Figure 14). Furthermore, the probability that Jeremy would emit a correct response given a response on S2 trials reached 100%, as opposed to the low and variable probability that of a correct response on S1 trials demonstrated during treatment (Figures 17 and 18). These data provide evidence that Jeremy learned to engage in the discard response in the presence of the S2 affective display, but did not learn to engage in the keep response in the presence of the S1 affective display. Also for Jeremy, neither responding, represented in Figures 15 and 16, occurred on 100% of the S1 and S2 training trials in baseline. The percentage of trials with a neither response progressively decreased on the discard trials, providing evidence of good discriminative control by the S2 affective display (Figure 16). Nevertheless, the percentage of keep trials with a neither response remained above 50% during the majority of the treatment sessions, indicating poor discriminative control by the S1 affective display (Figure 15).

Troy’s keep, discard, and neither responses were also distributed unequally. Specifically, Troy responded correctly with a keep response in the presence of S1 on 100% of the training trials on sessions 34, 35, and 38 (Figure 13). Correct discard responding only reached above 50% during one session throughout the entire treatment condition (Figure 14). Furthermore, the probability that Troy would emit a correct response given a response on S2 trials decreased throughout the course of treatment, as opposed to the high, but variable probability of a correct response on S1 trials demonstrated during treatment (Figures 17 and 18). These data provide evidence that Troy learned to engage in the keep response in the presence of the S1 affective display, but did not learn to engage in the discard response in the presence of the S2 affective display. Also for Troy, neither responding, represented in Figures 15 and 16, progressively decreased to 0% on both the S1 and S2 trial types during the treatment condition. In fact, Tory displayed the keep response more often than the other two response possibilities on both S1 and S2 trials. Therefore, conclusions regarding conditional control by affective stimuli are not possible in this case.

To summarize the training data, both Jeremy and Troy did not fully learn to discriminate between the S1 and S2 affective training stimuli as did Adam. Nevertheless, during the modified treatment condition (i.e., probe trials removed, modeling added to treatment package) Troy's correct keep and discard responding increased to 100% on session 48.

Analysis of Probe Responding For Adam, Jeremy, and Troy

Responding on probe trials is represented by open circles in Figures 12-14. For Adam, responding on the probe trials increased following the introduction of treatment along with responding on the training trials. While the percentage of responding on training trials remained high for Adam throughout treatment, there was an abrupt drop in responding on probe trials starting at Session 25, where responding to all probe stimuli decreased to 0%. Although responding on many of the discard probe trials recovered following session 25 (see Figure 14), responding on keep probe trials did not (see Figure 13).

A more detailed examination of probe responding was conducted for Adam to determine the pattern of responding to each of the four affective probe stimuli from S1 (i.e., the keep category). Figure 19 displays Adam's responding to the four individual keep probe stimuli from S1 (i.e., raising fists with thumbs extended up while smiling with mouth closed, leaning the body toward the stimulus with palms up moving in a "push along" motion, tilting head toward the stimulus while raising eyebrows, and extending arm toward the stimulus while pointing and moving the index finger repetitively) beginning with Session 25. As depicted in this figure, responding to only one of the keep probe stimuli (i.e., thumbs up, closed smile) recovered following Session 25 and remained high and relatively stable. On the contrary, responding to the remaining three probe stimuli (i.e., palms up push along, tilt head raise eyebrows, and extend arm move index finger) remained low and relatively stable following Session 25. Considering these response patterns, it was of interest to determine the effects of reinforcing one of the S1 keep probes in the set (i.e., palms up) on responding to the remaining two in the set. Thumbs up was not included in this analysis because responding to this stimulus recovered following Session 25.

An across-stimuli-experimental-design was used to successively introduce reinforcement across the three remaining probe stimuli from the keep category. Figure 20 displays percentage correct responding for each session across the three keep stimuli (i.e., palms up push along, tilt head raise eyebrows, and extend arm move index finger). The data from Sessions 25 through 41 of treatment from the first phase of this experiment were used as the baseline comparison for this evaluation. It was proposed that when reinforcement was applied following responding to the first stimulus (i.e., palms up push along), such responding would increase. This occurred during sessions 43 and 44 when reinforcement was provided following a correct keep response in the presence of the palms up push along affective display. If an increase in responding to the remaining three stimuli occurred concurrently with the introduction of reinforcement for responding to the first stimulus, then we may have concluded that the stimuli were part of the same stimulus class, in that reinforcing responding in the presence of one member of the class produces responding in the presence of other members of the class (Haring, Breen, Laitinen, 1989). This pattern of responding was indeed demonstrated immediately following the introduction of reinforcement in sessions 44 through 48, nevertheless, responding then dropped to 0% and was variable through the last sessions. Even more concerning was the variability in responding that continued in the presence of the palms up affective display, notwithstanding the reinforcement contingency that was applied to responding in its presence. Because of this variability in responding, it was not possible to draw conclusions regarding the existence of affective stimulus classes.

Jeremy and Troy's responding to probe stimuli is also of interest. The open circles in Figure 12 demonstrate that overall changes in responding to affective probe stimuli from S1 and S2 increased with changes in responding to affective training stimuli upon the introduction of treatment (Figure 12). For Jeremy, the percentage of probe trials with a correct keep response remained low throughout the treatment condition fluctuating between 0% and 25% (Figure 13). The percentage of probe trials with a correct discard response, however, increased to 100% on sessions 29 and 41, although responding remained variable throughout the treatment condition (Figure 14). For Troy the opposite occurred. The percentage of probe trials with a correct keep response increased to 100%

by the 6th session of treatment, and remained high, along with responding on training trials (Figure 13). The percentage of probe trials with a correct discard response, increased at first to 100% on the 5th session of treatment, but did soon decreased fluctuating between 0% and 25% for the remainder of the condition (Figure 14).

Given these data, it may be tempting to conclude that Jeremy engaged in generalized discard responding from the training S2 stimulus (i.e., shake head with eyebrows turned down) to some of the S2 probe stimuli. Furthermore, one may conclude that these data are a demonstration that the affective stimuli in S2 are part of a stimulus class. The data may support these conclusions in part; although given the extreme variability in responding throughout the treatment condition to both training and probe affective stimuli, the evidence seems lacking.

One other event is worth noting with respect to Troy's probe responding. Similar to the responding that occurred during training trials, Troy's probe responding consisted of mostly keep responses. Regardless of whether a training or probe affective stimulus was presented, Troy, more often than not engaged in a keep response in the presence of both S1 and S2 affective probe stimuli.

General Discussion

All three participants in this study learned to respond to affective displays presented by the experimenter within the social referencing paradigm. Experimental control was demonstrated in Experiment 2, when all three participants' responding to affective stimuli increased with the systematic introduction of the training procedures.

Although experimental control was demonstrated in Experiment 2, Jeremy and Troy did not reach optimal levels of responding. Essentially, the procedure used in Experiment 2 was a trial-and-error procedure. The participants were given the opportunity to respond following the presentation of the affective display, which simultaneously provided them with the opportunity to make errors. It is possible that minimizing errors early in training using an errorless learning paradigm in place of the error correction procedure may have accelerated acquisition in the initial phases of the treatment condition for Jeremy and Troy (Terrace, 1963). For example, the addition of a manual prompt immediately following the affective display, as opposed to waiting 3s, would have assisted the participants in the more immediate production of the correct response and increased the rate of reinforcement. Initially, this additional stimulus (i.e., manual prompt) would acquire stimulus control over the keep and discard response. These prompts would then be reduced over successive trials so that stimulus control could be transferred to the affective display (MacDuff, G.S., Krantz, P.J., & McClannahan, L.E., 2001).

One possible problematic factor in Experiment 1 was that the activity schedules could have exerted stimulus control over completion responding resulting in the failure of all three participants to acquire the termination response, and thus differential responding in the presence of S1 and S2 affective stimuli. Because of this possible competing source of stimulus control, the affective stimuli did not acquire control over differential completion and termination responding. In Experiment 2, it was anticipated that participants would learn to respond differentially to the affective stimuli from S1 and S2 because the response requirements (i.e., keep or discard) did not have a strong history of reinforcement associated with the stimuli used to occasion social referencing. Also, the choice paradigm allowed the participants to manipulate the objects presented, so restricted access to the objects was not a source for concern or discriminative for disruptive behavior as it was in Experiment 1.

Additionally, in Experiment 2, only one affective stimulus from S1 and one affective stimulus from S2 were trained, while responding to the other S1 and S2 stimuli was measured on probe trials. It was hoped that this reduction in the number of affective training stimuli in Experiment 2 would have made acquisition easier for participants. It may be the case that in Experiment 1, there were too many S1 and S2 stimuli included in the training set at one time allowing only one trial of training per affective stimulus. This may have interfered with acquisition of the conditional discrimination because this trial configuration limited participants' exposure to the training procedure and reinforcement contingencies. With all of the aforementioned modifications, participants continued to display various difficulties with learning the discrimination.

Conclusions regarding affective stimulus class formation are difficult to form from the data produced in this series of experiments, and evaluation of probe responding must be done on an individual basis for each of the learners. For Adam, because responding to both S1 and S2 affective probe stimuli increased to criterion levels with the introduction of treatment along with changes in responding to the training stimuli, we may conclude that generalization occurred during sessions 8-25. Nevertheless, responding to probe stimuli did not maintain throughout the course of the treatment condition. The extent to which learners respond similarly when presented with slightly physically different nonreinforced probe stimuli is of primary interest when making inferences about stimulus generalization. A major challenge for learners is the discriminability of reinforcement contingencies that are associated with the presentation of S^+ (training) and S (probe) stimuli. Specifically, to demonstrate stimulus generalization, responding must be controlled by the physical properties of the probe stimuli and not by those aspects of the probe stimuli that are discriminative for nonreinforcement. Thus, the discriminative stimuli may serve two different functions. Stimuli may signal the response topography to be emitted by the learner *and* they may signal the availability of reinforcement. To evaluate stimulus generalization, learners must discriminate the former rather than the latter and not respond differentially to changes in reinforcement contingencies associated with S^+ and S^- stimuli (i.e., training and probe stimuli). We hypothesized that the predictive reinforcement value of the affective probe stimuli interfered with Adam's maintenance of probe responding throughout the course of the treatment condition.

We tested this hypothesis by introducing reinforcement for responding on one of the S1 affective stimuli from the keep category while continuing to measure responding on the remaining two in the absence of reinforcement. As we expected, responding increased in the presence of the stimulus that was predictive of reinforcement and in the presence of the stimuli that were not immediately following the introduction of reinforcement. These results support the case of discriminability of reinforcement contingencies described above. In addition, this type of generalized responding usually supports conclusions regarding the existence of a stimulus class. Nevertheless, because responding did not maintain and progressed variably, such conclusions may not be fully supported.

If anything, it is clear that to facilitate the learning of affect discrimination by children with autism, it may be necessary to create specifically individualized training procedures for each learner. For instance, in Experiment 2, it was determined that the modified procedure (i.e., providing modeling and instructions at the beginning of the session) produced a sharp increase in responding differentially to the two affective displays presented for Troy. Although this was not demonstrated experimentally, it does provide some evidence of the extent to which truly individualized instruction may be needed to teach the discrimination of affective stimuli by children with autism. Future research may explore the effectiveness of this procedure experimentally.

In addition, it may be useful to identify the prerequisite skills necessary for learning social referencing and to conduct a pretest of these skills for each individual learner before attempting to train link 1 and link 2 responses to occur consecutively within the social referencing response chain. Because the sources of stimulus control may be difficult to predict for any one learner when choosing participants to include in experimental analyses, the pretest measure may assist in the development of individualized training procedures that can later be implemented across subjects grouped according to similar skills. This would prevent the loss of experimental control that results when different learners require different types of training procedures to acquire the same skill.

With respect to differential responding in the presence of affective stimuli (i.e., link 2) it may be useful to first establish the sets of affective stimuli as classes using a basic matching to sample paradigm, as opposed to attempting to teach differential responding within the context of social referencing. Under such training conditions, participants would learn to touch a green comparison stimulus in the presence of affective stimuli from S1 and a red comparison stimulus in the presence of affective stimuli from S2 establishing conditional control by affective stimuli from each set. Subsequent discrimination of affect within the social referencing paradigm may be facilitated because the affective stimuli have already been established as classes of stimuli that occasion common response classes and have acquired conditional discriminative control over responding.

Furthermore, for some learners who fail to acquire conditional discriminations, it also may be necessary to first teach the simple discrimination of one affective stimulus use a blocking procedure before attempting to teach the simultaneous discrimination of two or more different affective stimuli. Using a blocking procedure (Saunders & Spradlin, 1989) the experimenter would present the same affective stimulus from S1 as the sample stimulus for a specified number of consecutive trials until the learner responds correctly according to some predetermined criterion. Following this, the experimenter would present another sample affective stimulus from S2 until a criterion level of responding is met. Following the acquisition of these simple discriminations, the experimenter would then intersperse affective stimuli from S1 and S2 as the sample stimuli using a simultaneous discrimination training procedure. Additional pairs of affective stimuli from S1 and S2 may also be trained in this manner in an attempt to increase the size of the stimulus classes. Although in the current experiment the blocking procedure failed to assist in the learning of the discrimination for Jeremy during Experiment 1, this may not be the case for other learners. Again, a pretest may determine which learners have a history of failing to acquire conditional discriminations.

Future research may also focus on additional questions related to responding under link 1 of the social referencing response chain by attempting to establish conditional control over observing by ambiguous or novel stimuli. This may be done by reinforcing observing in the presence of ambiguous or novel stimuli and by withholding reinforcement in their absence. This type of training would first require the identification of stimuli that should and should not come to control differential observing in the natural environment.

Finally, it is possible that once learners' responding is under conditional control of the Link 1 and Link 2 stimuli in separate training contexts they may more easily acquire the entire response chain within the context of social referencing. Adding to the complexity of the training context may include the final step of attempting to transfer control of well established link 1 and link 2 responses to stimuli related to everyday tasks and even perhaps while independently following ac*ti*vity schedules at school or even in the home or community.

Table 1

List of school stimuli used during the pre-training sessions in Experiment 1.

Handwriting	Retrieving Objects	Scripted Verbal Interaction
Standard # 2 pencil/paper	Back pack on hook	"You look nice"
	Spoon in utensil caddy	"How are you?"
	Fork in utensil caddy	"What's up?"
	Book on bookshelf	"Hug me"
	Game on shelf	"What's new?"
	Napkin in holder on shelf	"High five"
	Paper towel roll on shelf	"Nice shirt"
	Tissue box on shelf	"Let's dance"
	Soap on shelf	"Tickle me"
	Lunch box in cubby	"Sing a song"
	Glue in art caddy	"Let's jump"
	Scissors in art caddy	"Tell a secret"

Table 2

List of experimental stimuli used during baseline and conditional discrimination training during Experiment 1.

Handwriting	Retrieving Objects	Scripted Verbal Interaction
Mouse pen/ bubble wrap	Back pack on hook/jumbo party hat attached	“You look nice” /high pitch
Fiber optic pen/paper plate	Spoon in utensil caddy/pipe cleaner wrapped around it	“How are you?”/low pitch
Koosh pen/ cardboard	Fork in utensil caddy /crazy straws wrapped around it	“What’s up?”/ fast
Jumbo poster marker/ paper bag	Book on bookshelf /fuzzy necklace hanging from it	“Hug me”/shout
Jumbo chalk/yellow envelope	Game on shelf/ wrapped in orange pillow case	“What’s new?”/slow
Fabric marker/fabric	Napkin in holder/fake rose sticking out	“High five”/sneeze
Pink highlighter pen/party streamer	Paper towel roll /feathers sticking out of it	“Nice shirt.”/baby crying
Marker/aluminum foil	Tissue box/ bead necklace wrapped around it	“Let’s dance.”/static
Crayon/file folder	Soap dispenser/feather boa wrapped around it	“Tickle me”/cough
Feather pen/gift wrap	Lunch box /streamer on it	“Sing a song”/ banging

Table 2 *(continued).*

Handwriting	Retrieving Objects	Scripted Verbal Interaction
Skinny white chalk/black construction paper	Scissors in art caddy/ Styrofoam cone in it	“Let’s jump”/ stammering
Black Marker/cork	Glue in art caddy/metallic party horn in it	“Tell a secret”/ background radio

Table 3

List of affective stimuli from each set presented following an observing response during baseline and conditional discrimination training in Experiment 1.

Approach- Task Completion (S1)	Avoidance- Task Termination (S2)
Nodding head while smiling with teeth exposed	Shaking head left to right with eyebrows turned down
Raising fists with thumbs extended up while smiling with mouth closed	Scared face with hands over mouth
Extending arm toward the stimulus while pointing and moving the index finger	Waving hands in a criss-cross motion with eyes wide open
Repetitively tilting head toward the stimulus while raising eyebrows	Turning head and upper body away from the stimulus while closing eyes
Leaning the body toward the stimulus with palms up moving in a “push along” motion	Gasping and stepping away from the stimulus
Moving head back and forth from the stimulus to the learner with mouth open.	Frowning with the bottom lip protruding out.

Table 4.

List of affective stimuli from each set presented following an observing response during baseline and conditional discrimination training in Experiment 2.

***T** indicates the training stimuli and P indicates the probe stimuli.*

S1- KEEP (Keep)	S2-DISCARD (Discard)
Nodding head while smiling with teeth exposed (T)	**Shaking head left to right with eyebrows turned down (T)**
Raising fists with thumbs extended up while smiling with mouth closed (P)	Scared face with hands over mouth (P)
Extending arm toward the stimulus while pointing and moving the index finger (P)	Waving hands in a criss-cross motion with eyes wide open (P)
Repetitively tilting head toward the stimulus while raising eyebrows (P)	Turning head and upper body away from the stimulus while closing eyes (P)
Leaning the body toward the stimulus with palms up moving in a “push along” motion (P)	Gasping and stepping away from the stimulus (P)

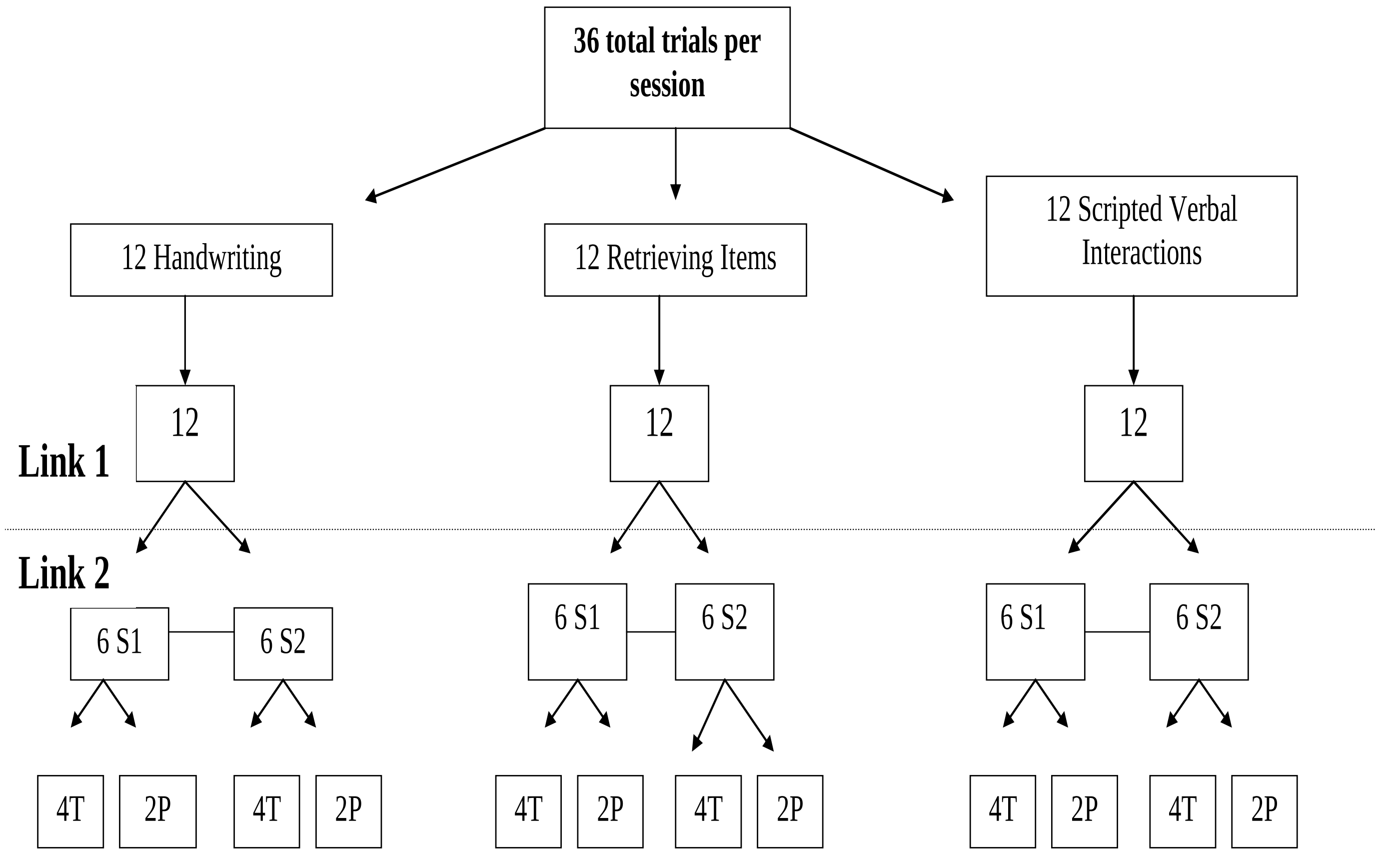

Figure 1. Trial configurations that were used during baseline and training conditions in Experiment 1.

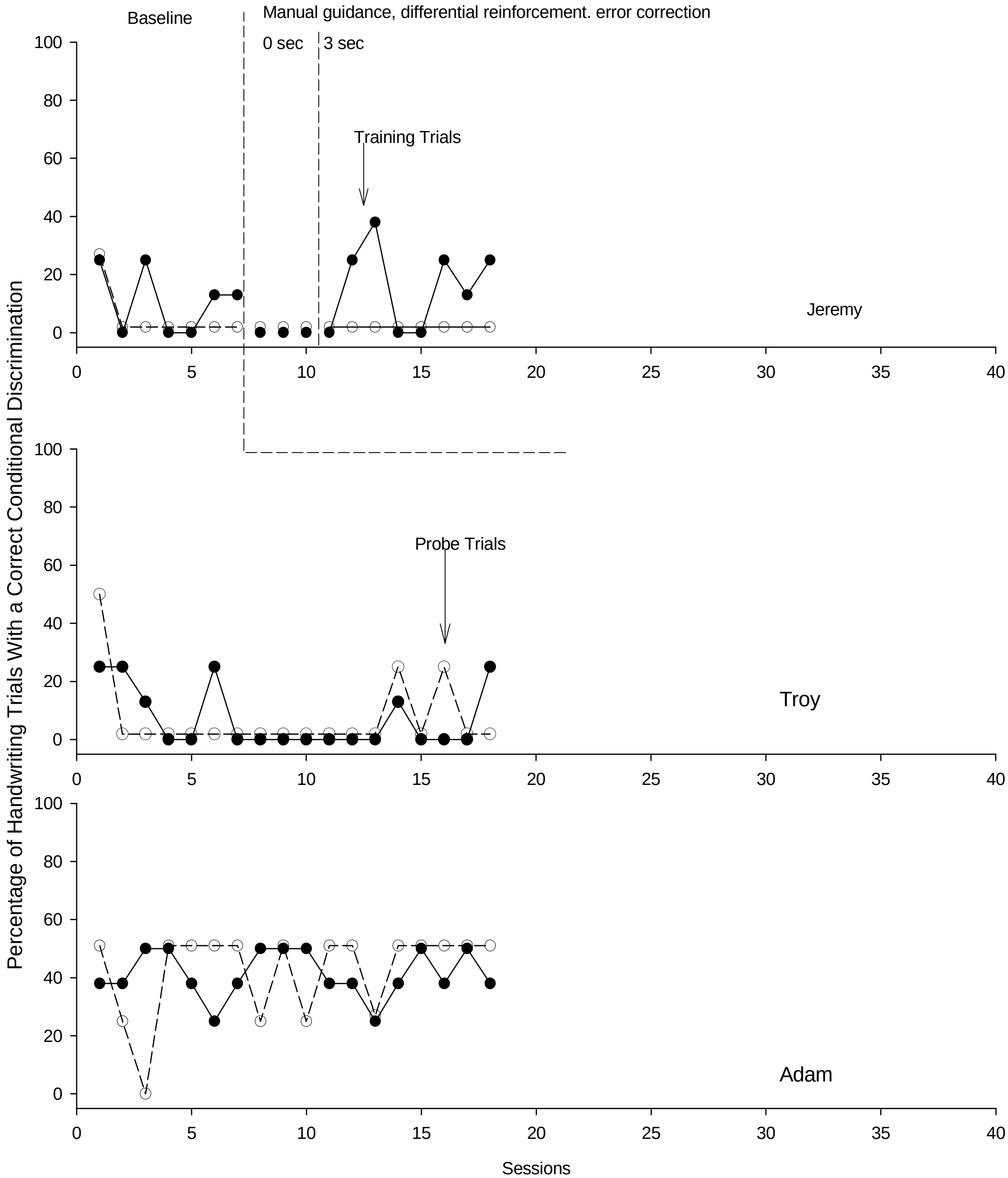

Figure 2. Percentage of training and probe handwriting trials with a correct conditional discrimination across participants Jeremy, Troy, and Adam (Experiment 1).

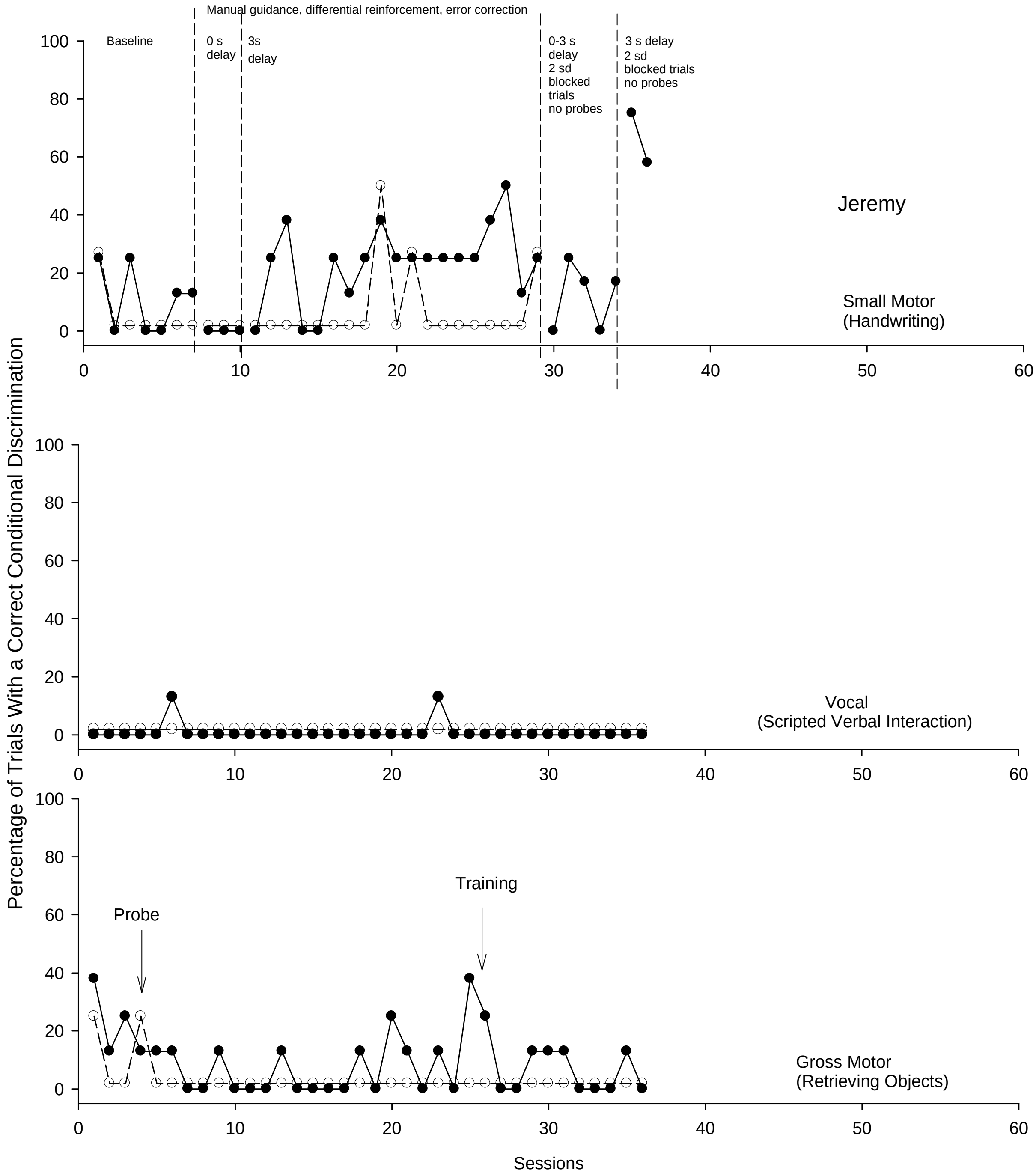

Figure 3. Percentage of training and probe trials with a correct conditional discrimination across small motor, vocal, and gross motor responses for Jeremy (Experiment 1).

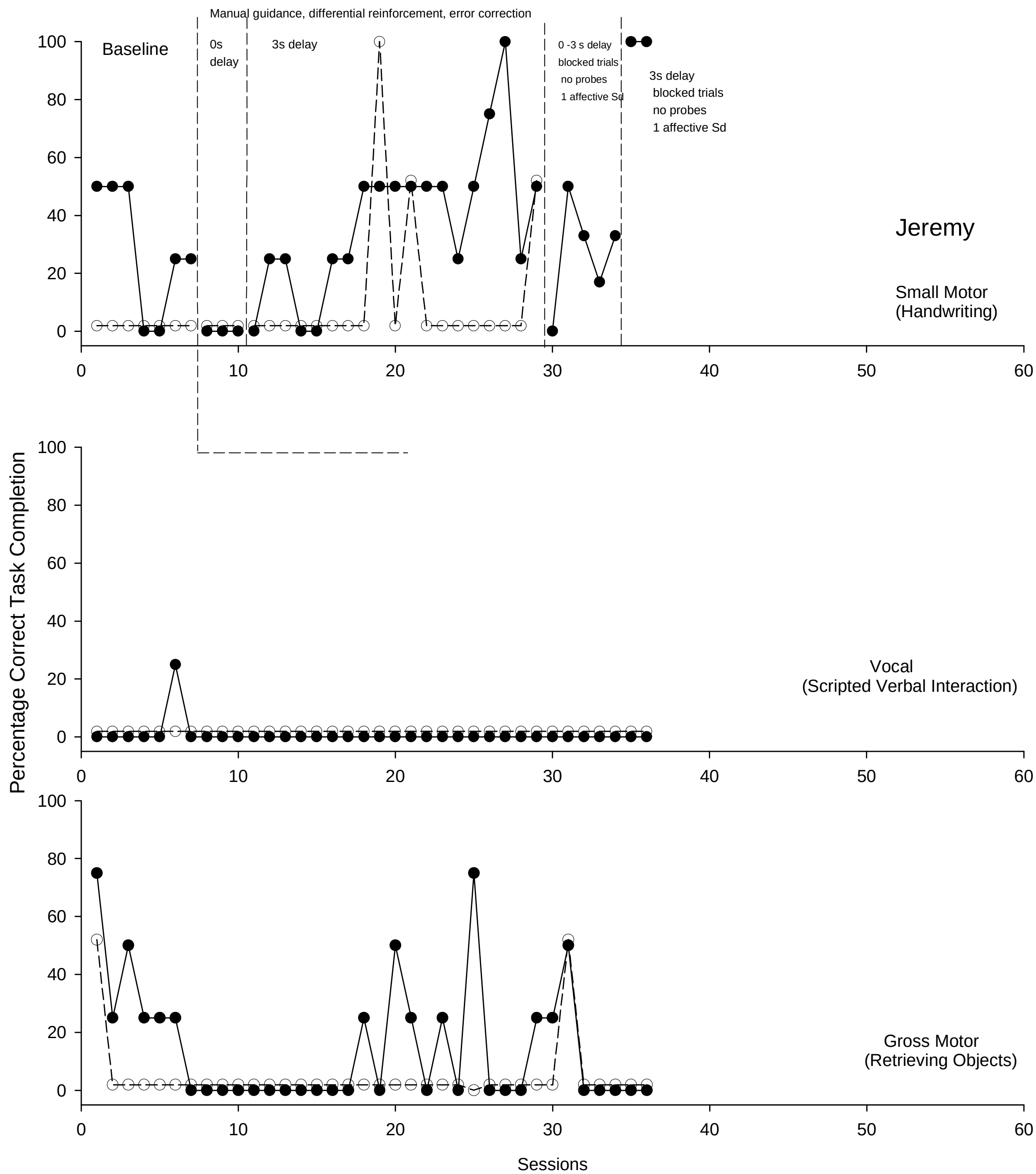

Figure 4. Percentage of training and probe trials with a correct completion response across small motor, vocal, and gross motor responses for Jeremy (Experiment 1).

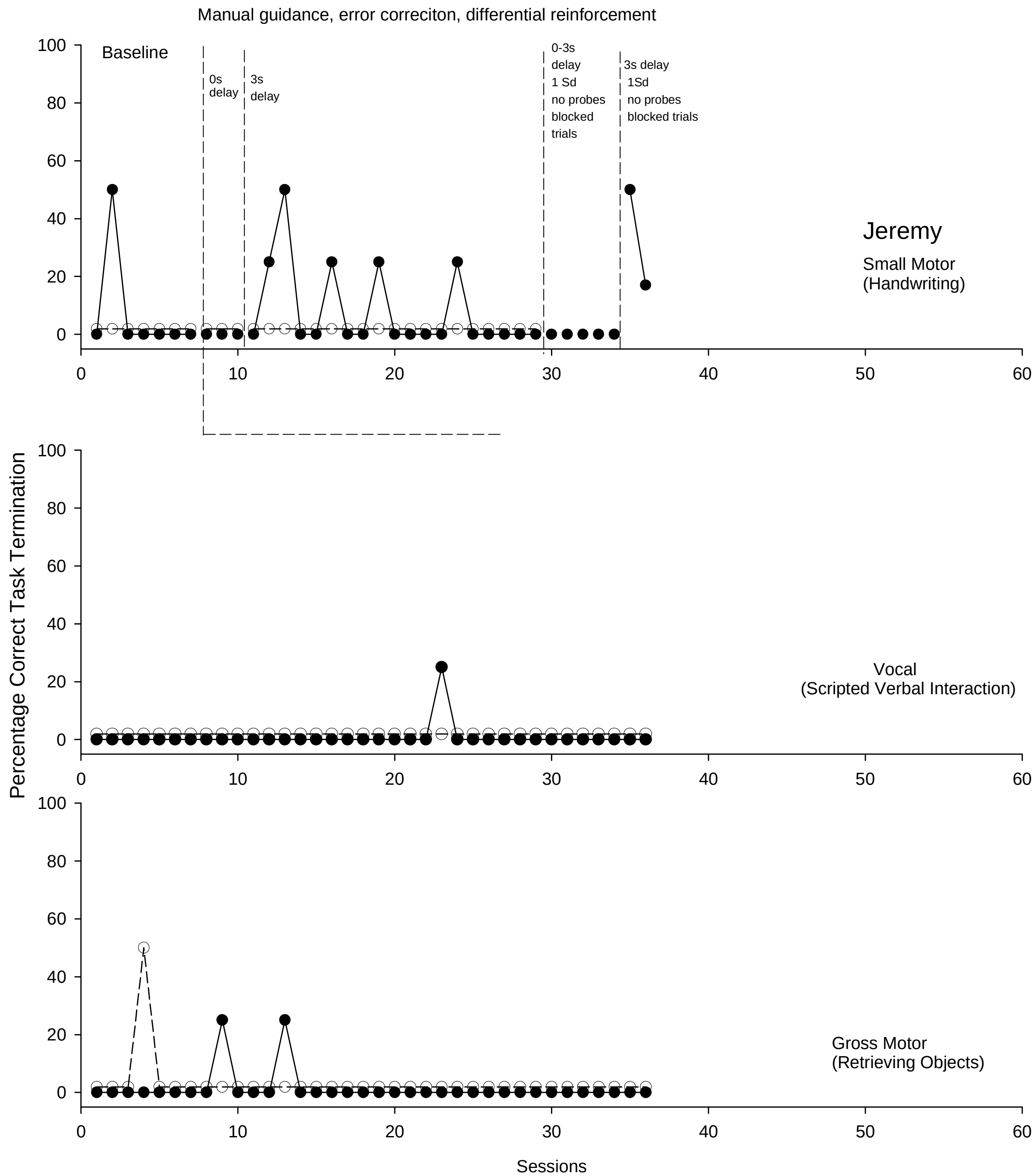

Figure 5. Percentage of training and probe trials with a correct termination response across small motor, vocal, and gross motor responses for Jeremy (Experiment 1).

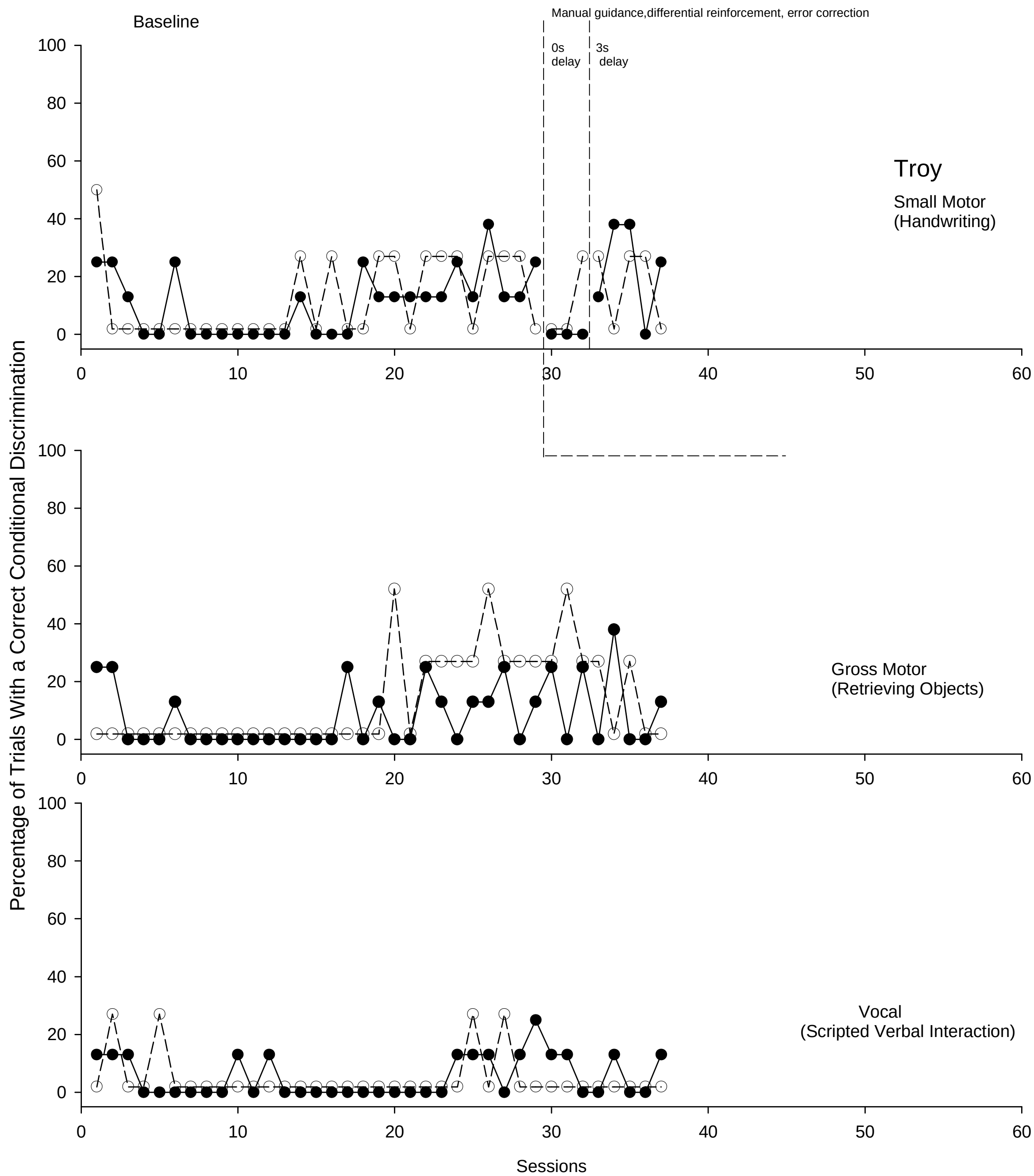

Figure 6. Percentage of training and probe trials with a correct conditional discrimination across small motor, vocal, and gross motor responses for Troy (Experiment 1).

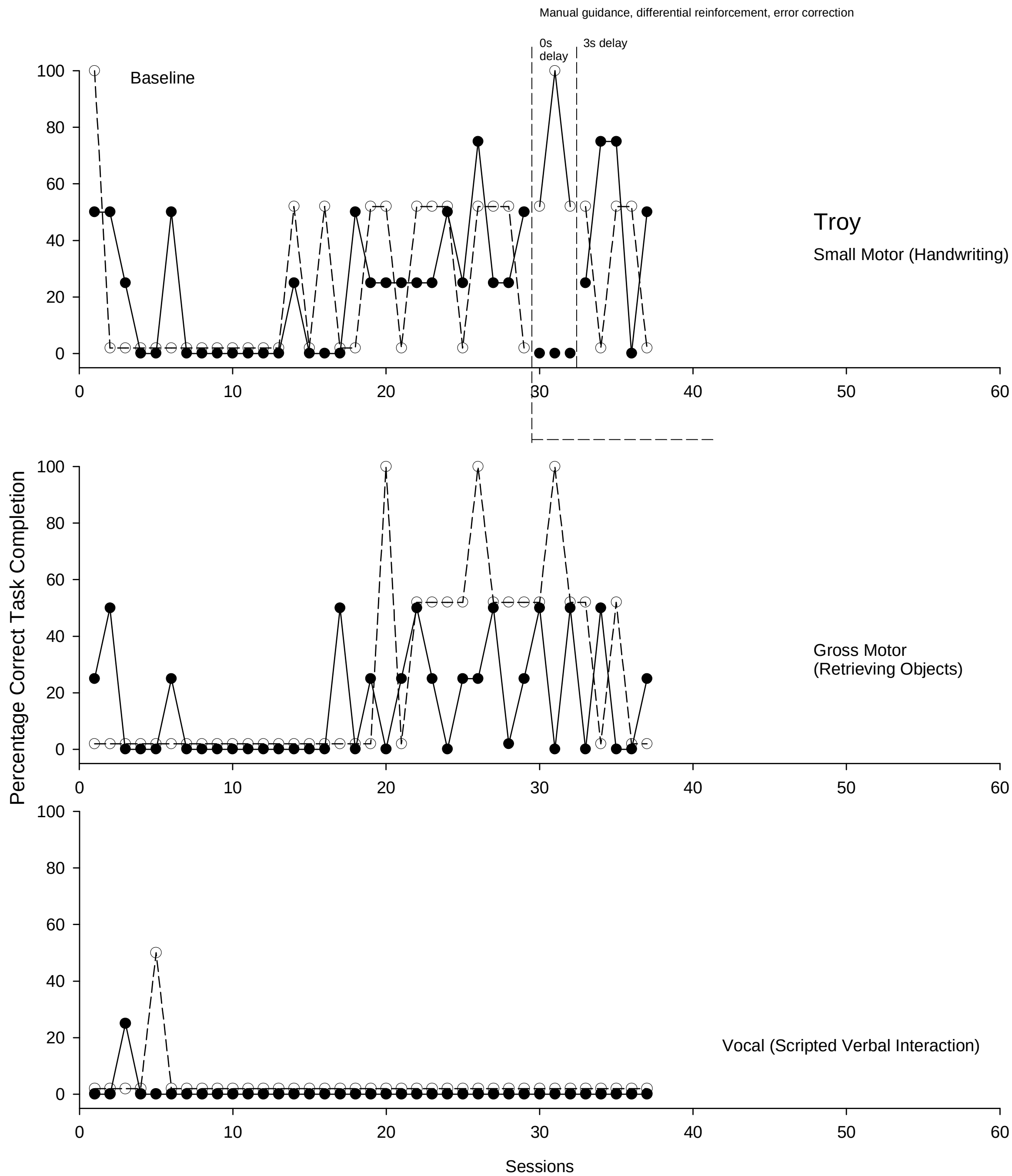

Figure 7. Percentage of training and probe trials with a correct completion response across small motor, vocal, and gross motor responses for Troy (Experiment 1).

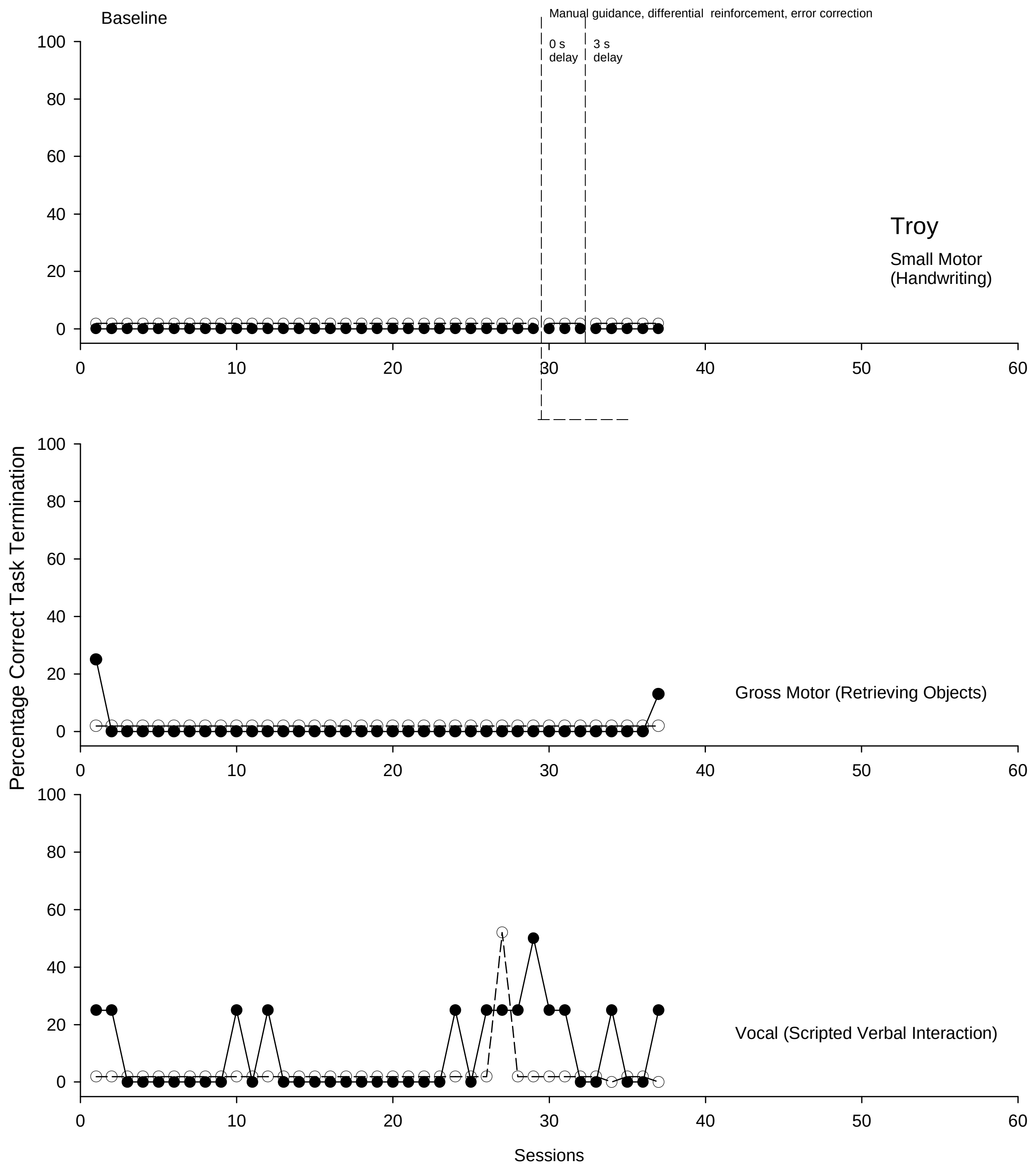

Figure 8. Percentage of training and probe trials with a correct termination response across small motor, vocal, and gross motor responses for Troy (Experiment 1).

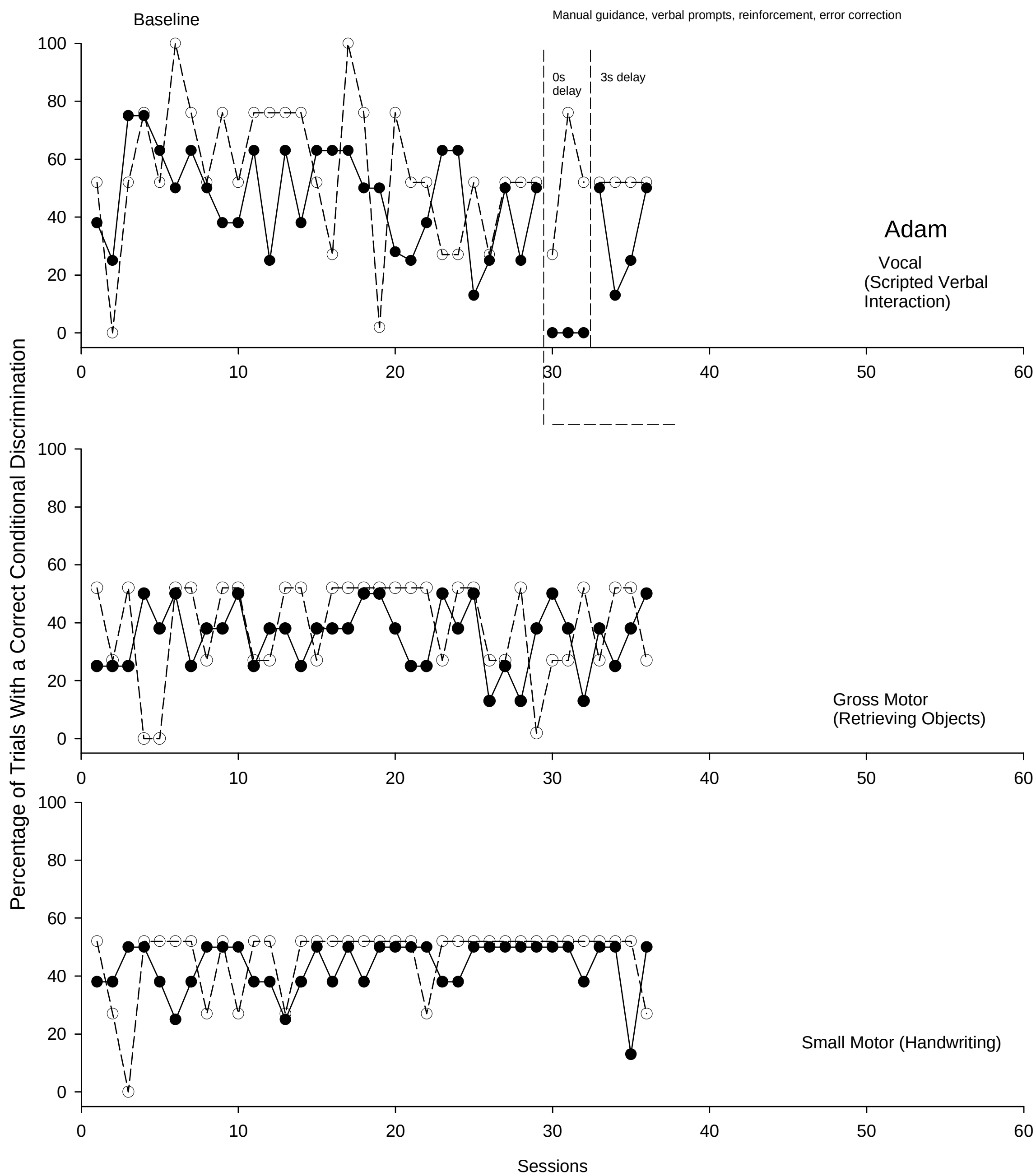

Figure 9. Percentage of training and probe trials with a correct conditional discrimination across small motor, vocal, and gross motor responses for Adam (Experiment 1) .

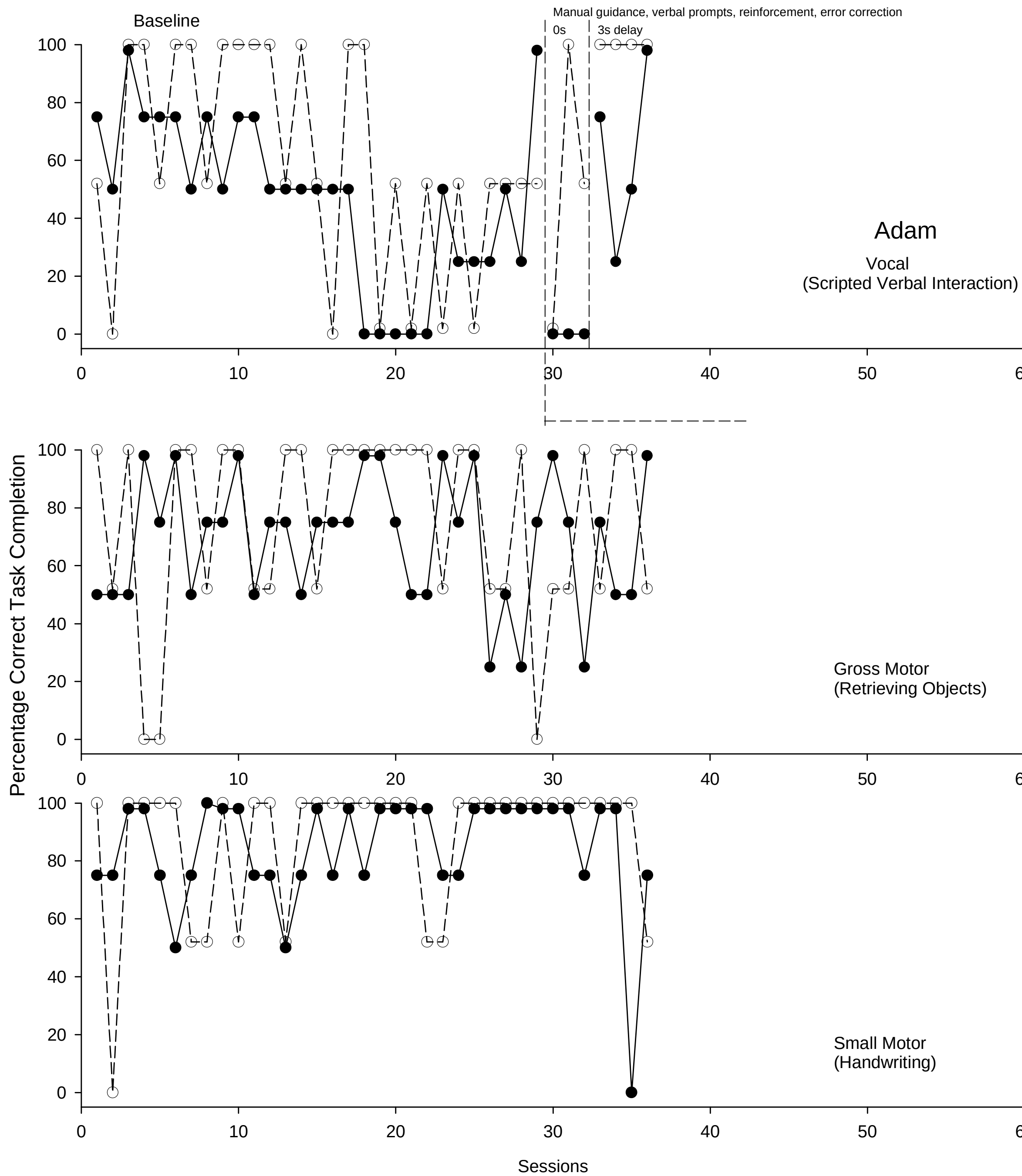

Figure 10. Percentage of training and probe trials with a correct completion response across small motor, vocal, and gross motor responses for Adam (Experiment 1).

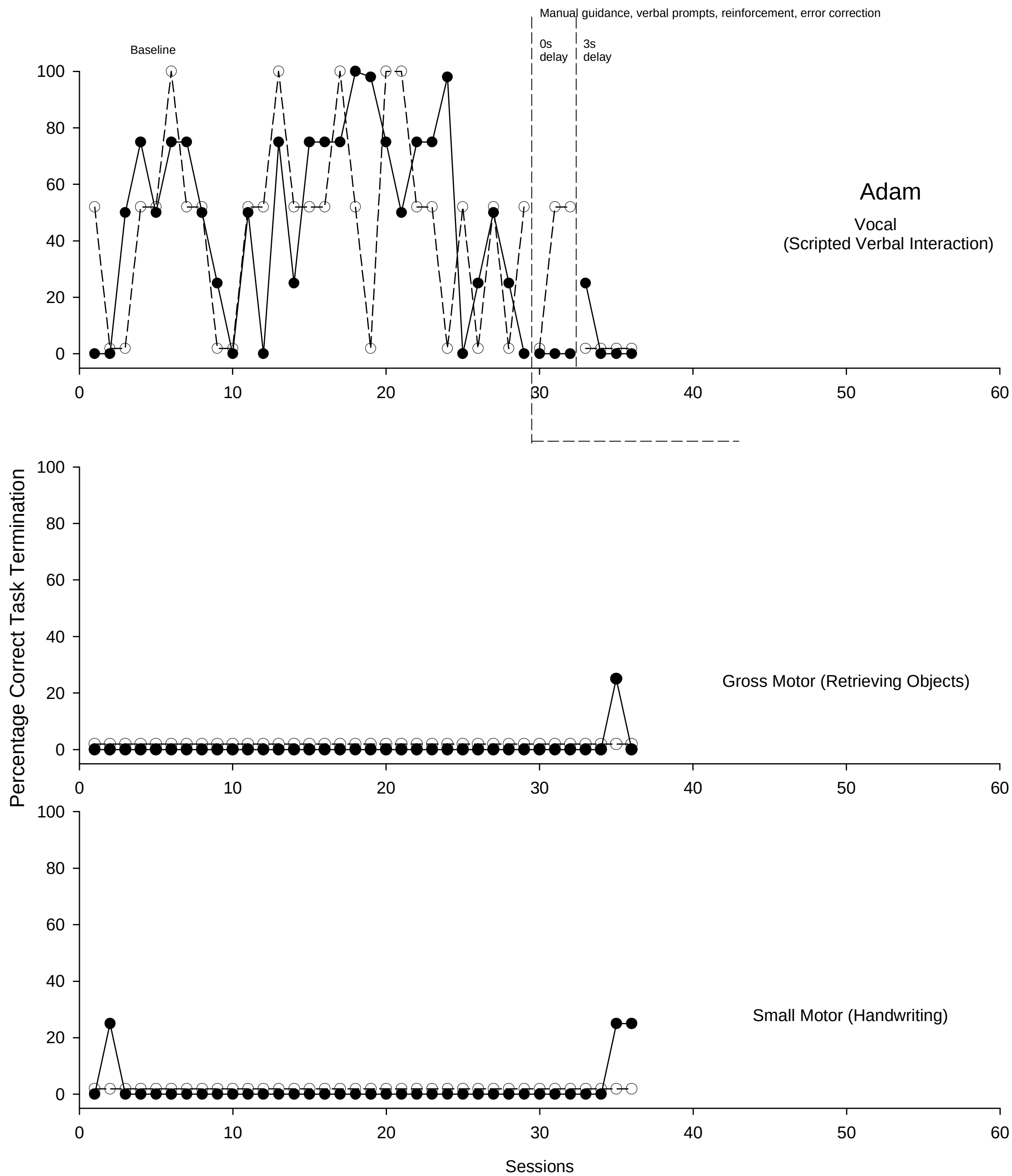

Figure 11. Percentage of training and probe trials with a correct termination response across small motor, vocal, and gross motor responses for Adam (Experiment 1).

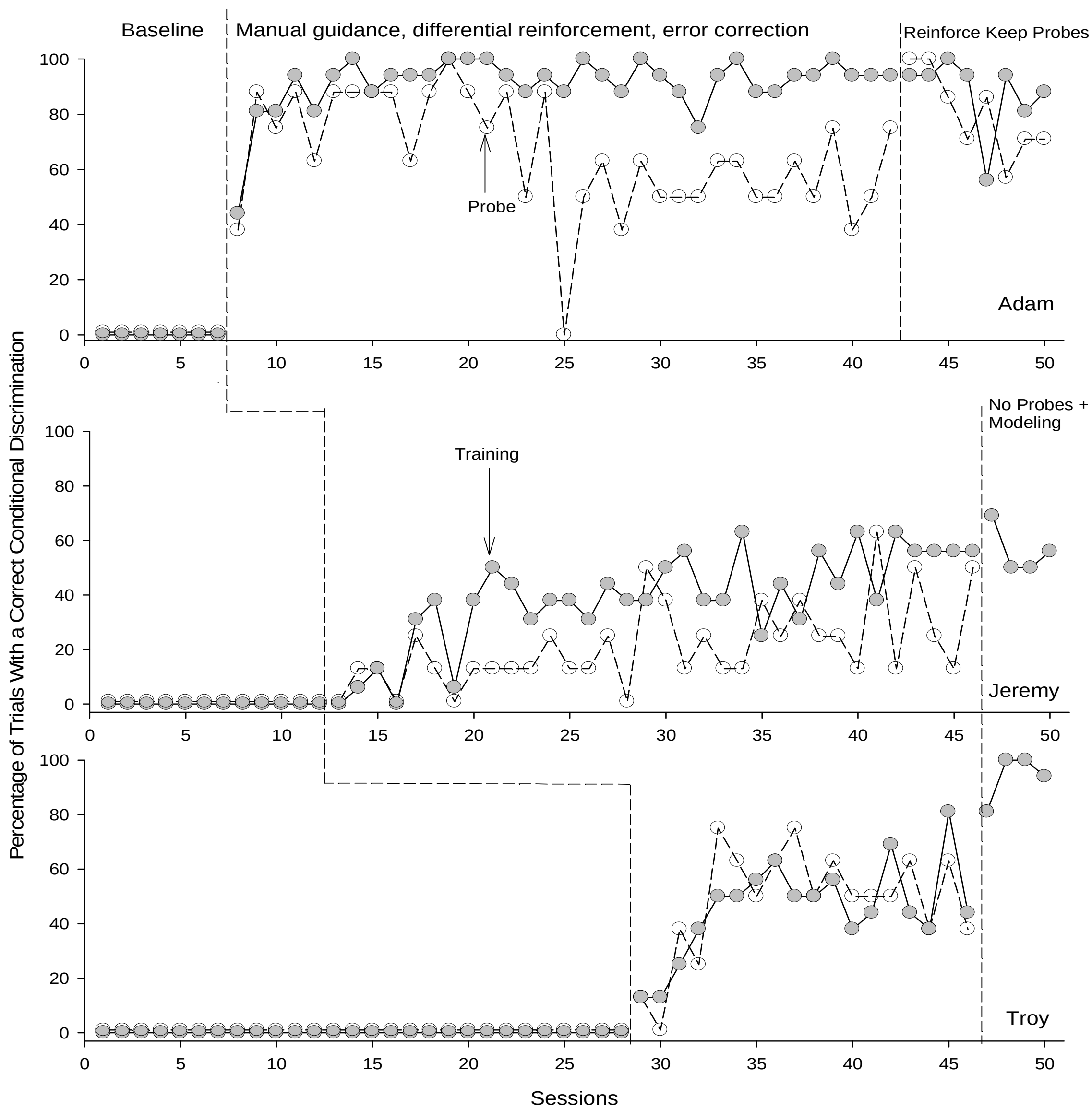

Figure 12. Percentage of training and probe trials with a correct conditional discrimination across participants (Experiment 2).

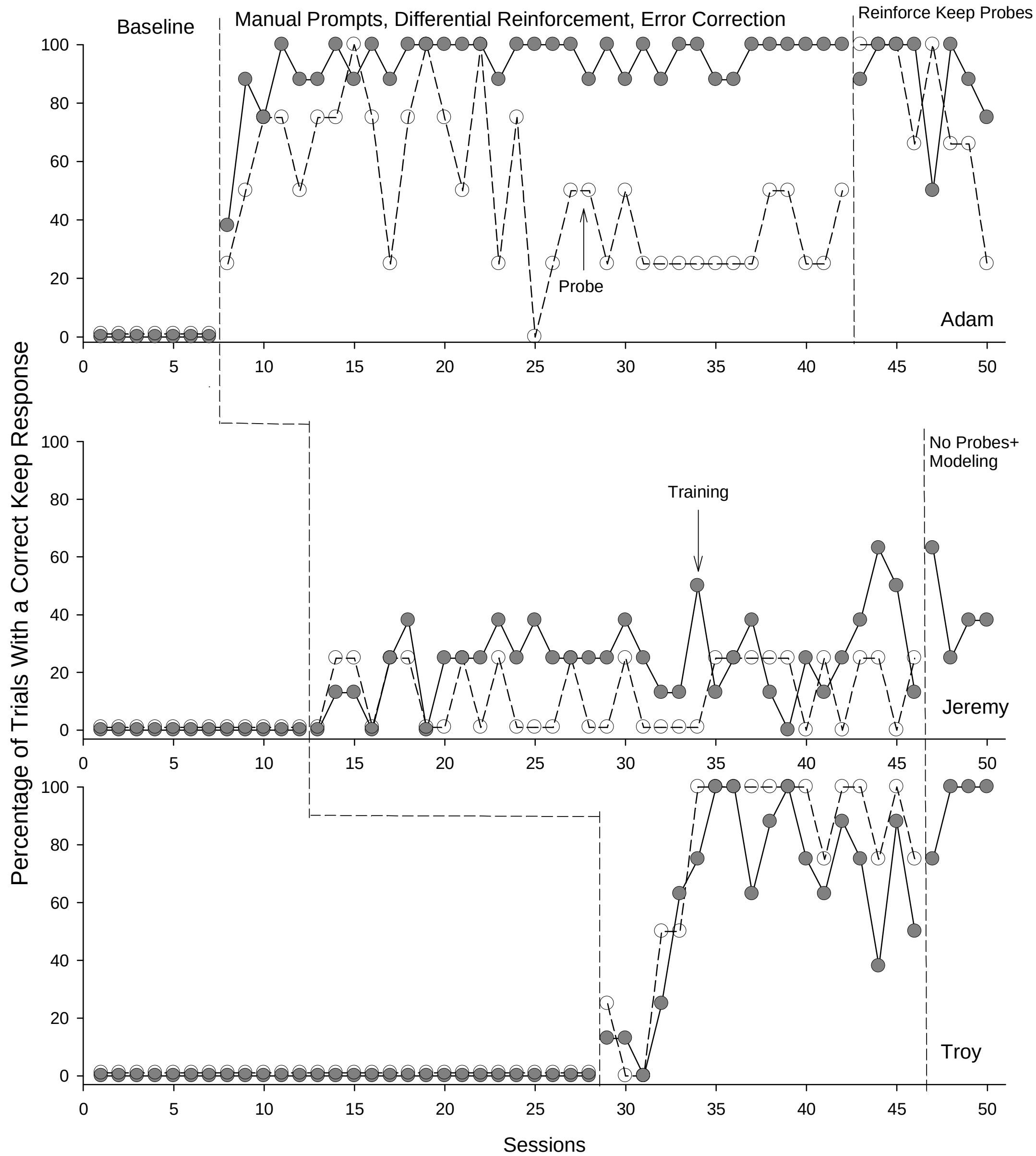

Figure 13. Percentage of training and probe trials with a correct keep response in the presence of S1 stimuli across participants (Experiment 2).

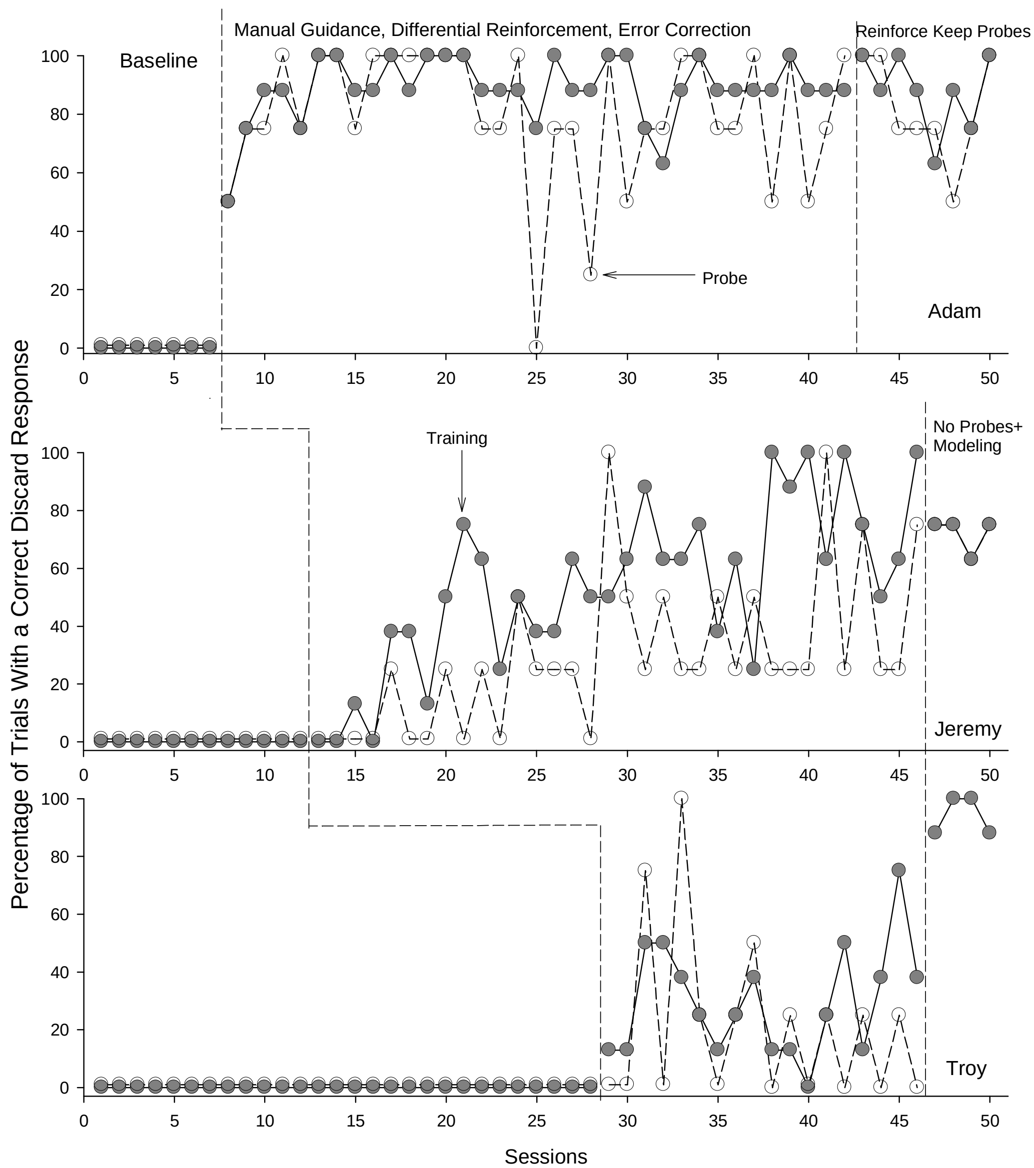

Figure 14. Percentage of training and probe trials with a correct discard response in the presence of S2 stimuli across participants (Experiment 2)

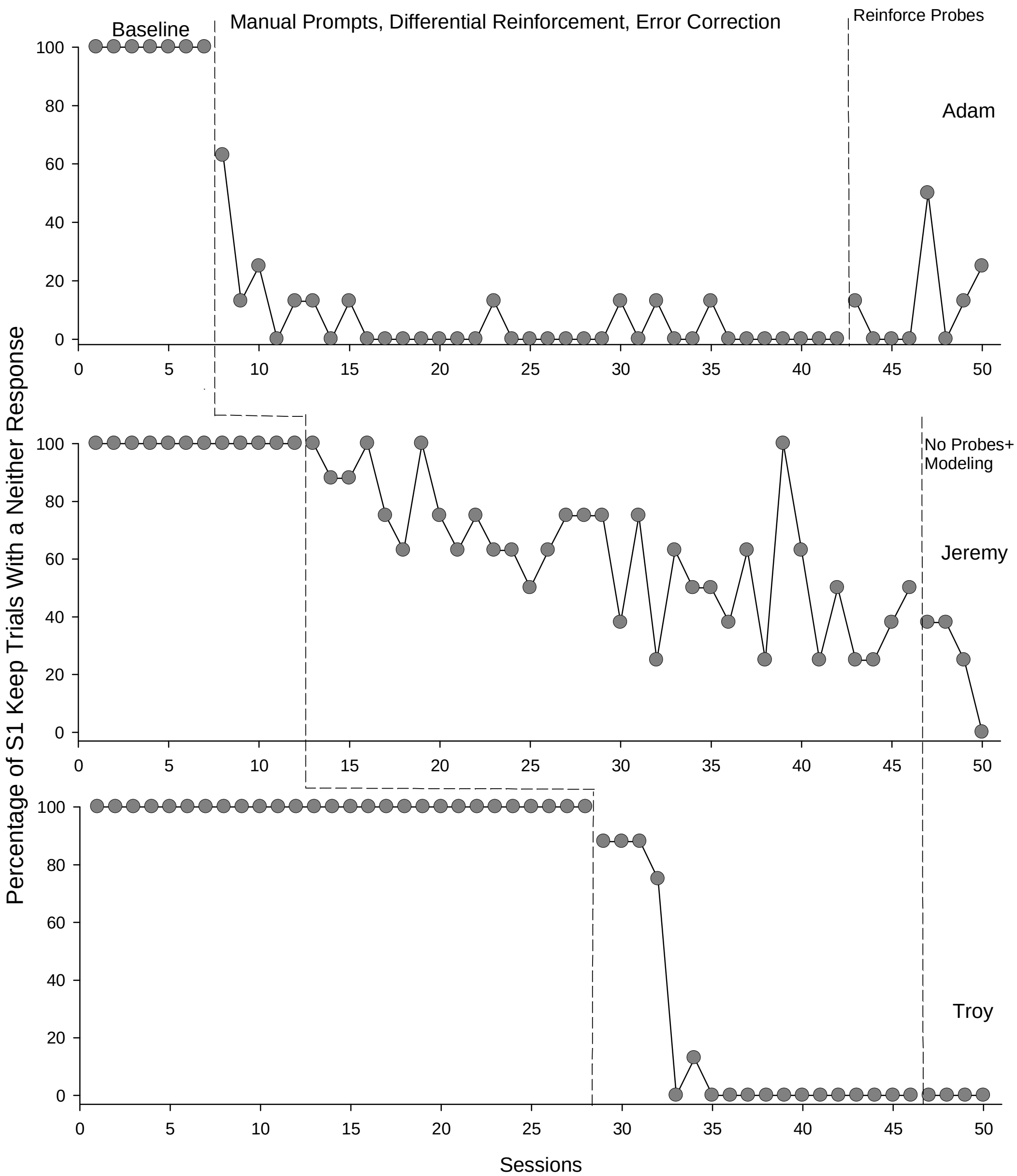

Figure 15. Percentage of S1 training trials with a neither response during baseline and treatment for Adam, Jeremy, and Troy.

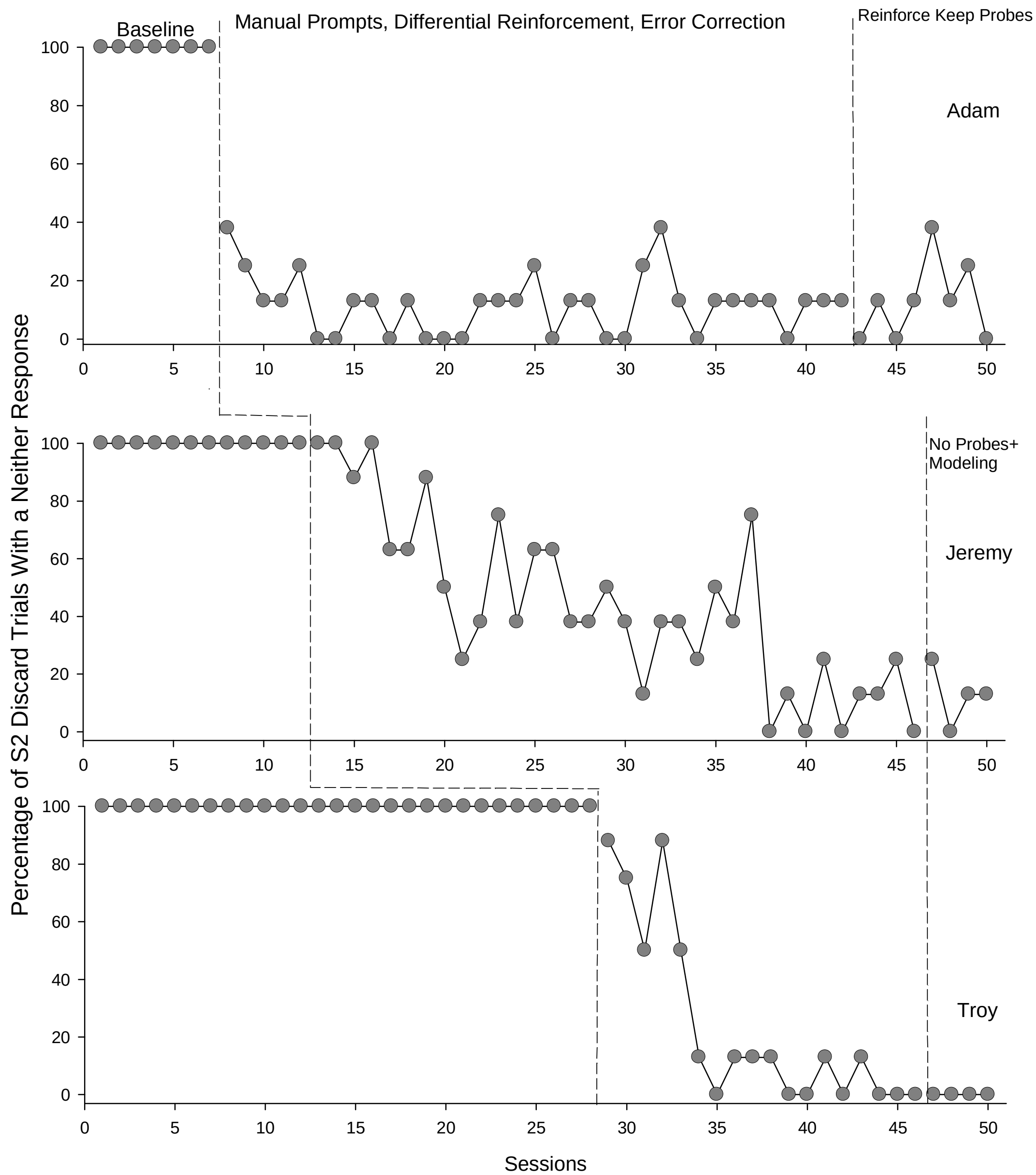

Figure 16. Percentage of S2 training trials with a neither response during baseline and treatment for Adam, Jeremy, and Troy.

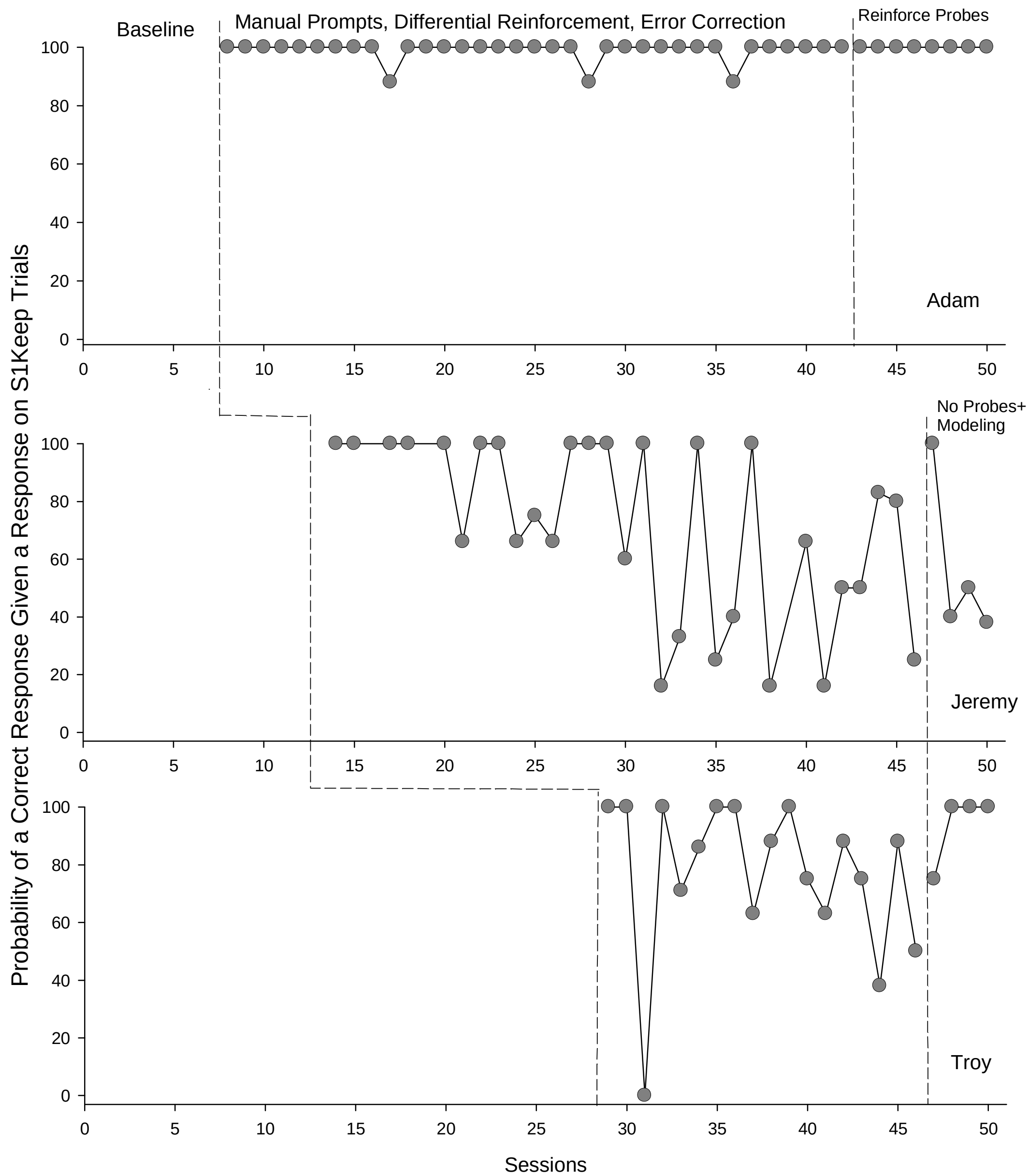

Figure 17. Probability of a correct response given a response on S1 training trials during treatment for Adam, Jeremy, and Troy.

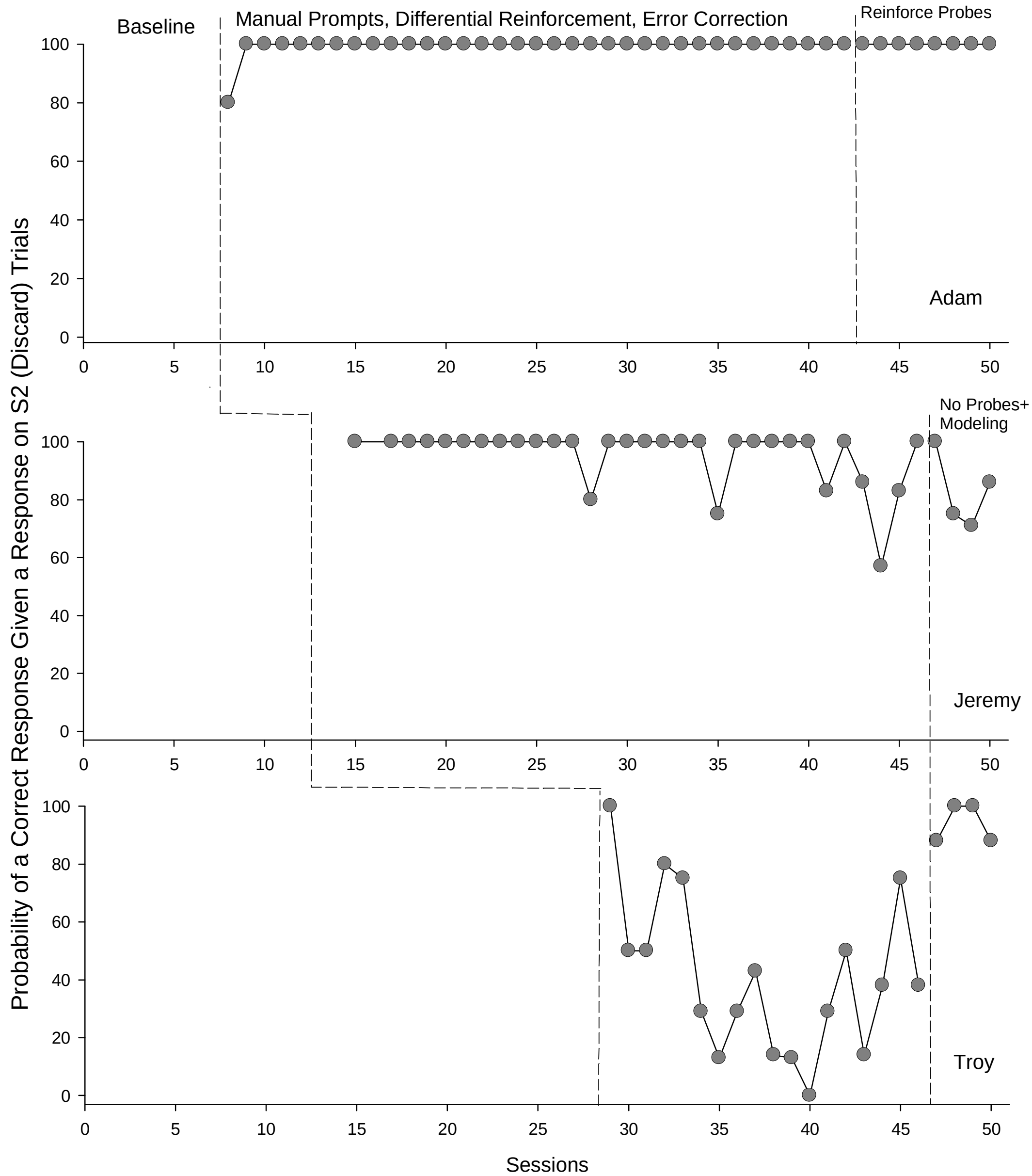

Figure 18. Probability of a correct response given a response on S2 training trials during treatment for Adam, Jeremy, and Troy.

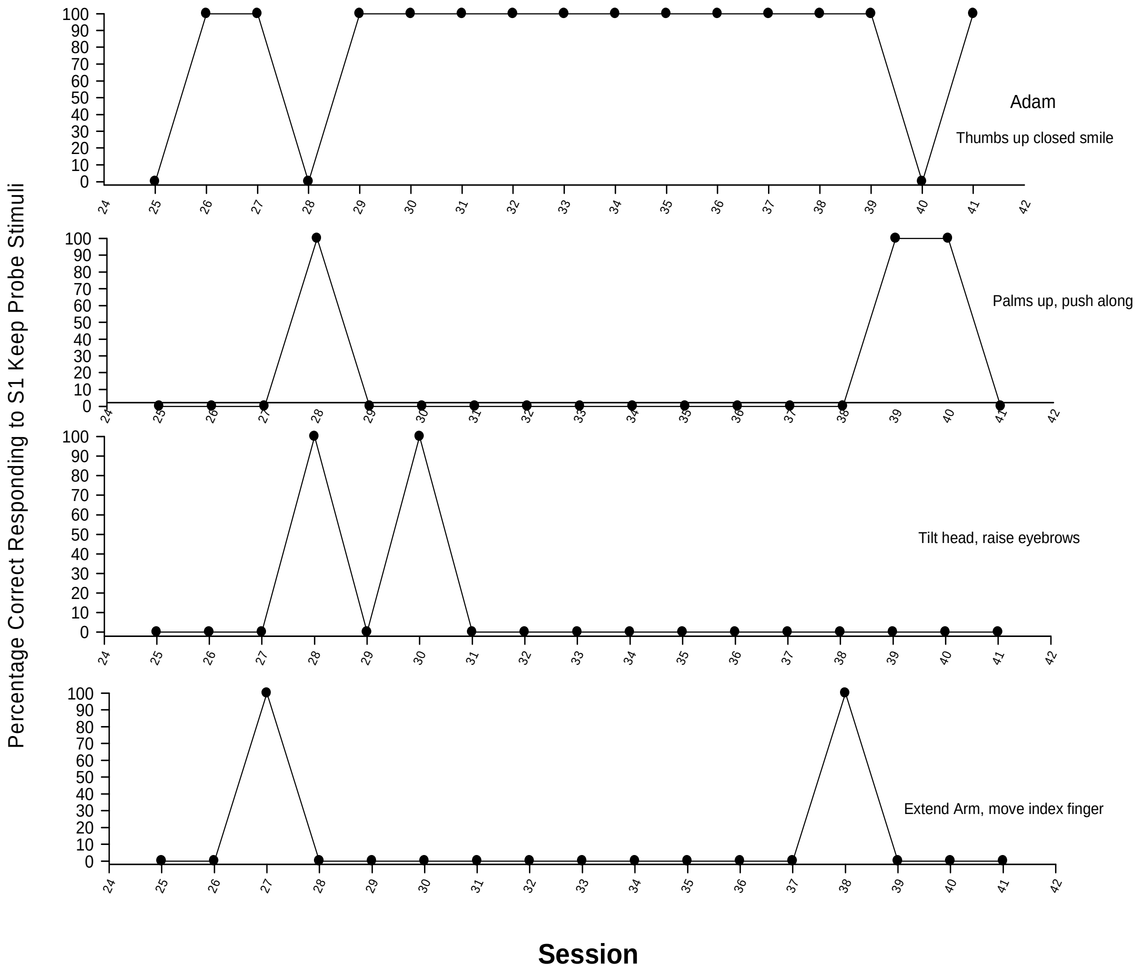

Figure 19. Percentage of S1 probe trials with a correct keep response for Adam during session 25-41 of treatment (Experiment 2).

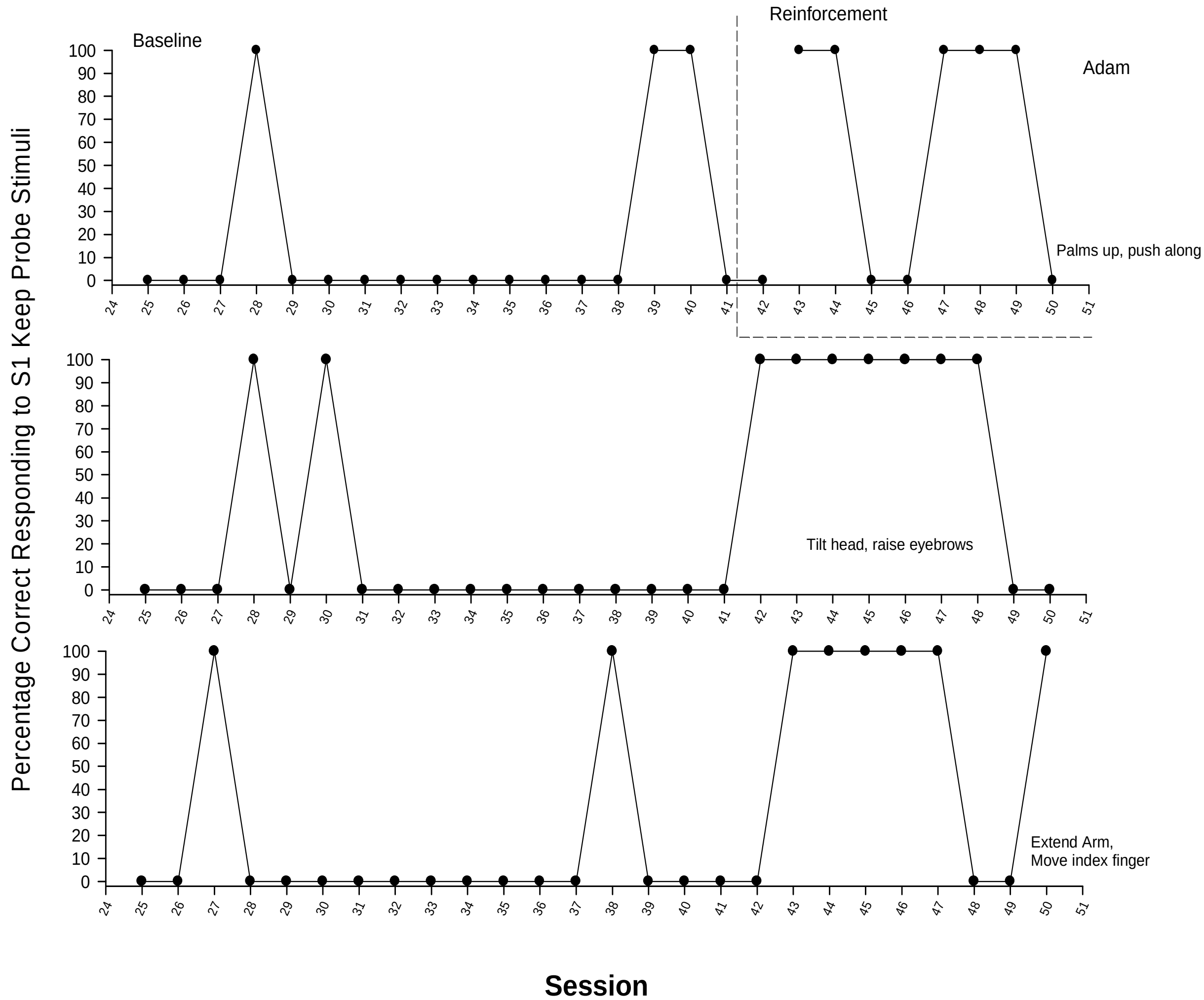

Figure 20. Percentage of S1 probe trials with a correct keep response for Adam across three affective stimuli from S1 (Experiment 2).

Appendix

Drawing Inferences Regarding Stimulus Generalization in Applied Behavior Analysis

In their seminal article, Stokes and Baer (1977) elaborated on procedures for producing generality of behavior change and were extremely successful in focusing the field on the production of nontrivial demonstrations of behavior change. Because of their influential work, generalization from training has become the cornerstone of applied behavior analysis. Nevertheless, the term *generalization* may sometimes be overused in ways that fail to make contact with its scientific definition. The present paper focuses on important aspects of drawing inferences about the occurrence of stimulus generalization in applied behavior analysis, factors that may limit such inferences, and methodological solutions from both the basic and applied research for overcoming such limitations.

Defining Stimulus Generalization

Stimulus generalization has been defined as the spread of effects of reinforcement occasioned by one stimulus (S^+) to other stimuli that differ along one or more dimensions and that do not occasion reinforcement (S^-) (Catania, 1998). Procedures for measuring stimulus generalization have involved presenting the S^+stimulus and many S^- stimuli that share some physical properties with the S^+stimulus. These studies have demonstrated that the highest response rates are observed in the presence of the S^+stimulus and that response rates progressively decline in the presence of stimuli that systematically vary from the S^+stimulus (Guttman and Kalish, 1956).

Drawing Inferences Regarding Stimulus Generalization

In demonstrations of stimulus generalization, an organism must be capable of showing discriminative responding to the physical properties associated with both the S^+ stimulus (also called the training stimulus) and the S^- stimuli (also called the testing or probe stimuli), otherwise, it may be argued that responding during tests of generalization may be attributed to the *failure* of the subject to discriminate the physical properties of the two sets of stimuli (Wasserman, Kiedinger, and Bhatt, 1988).

In addition to physical properties, stimuli have a predictive function with respect to reinforcement schedules. Learners may therefore respond to stimuli with similar physical properties, but fail to respond if the stimuli are predictive of non-reinforcement. Therefore, researchers must also be concerned with control that the stimuli may acquire with respect to their predictive reinforcement value when making inferences regarding stimulus generalization. Because stimulus generalization is usually measured under extinction conditions, the learner may well cease responding to the generalization testing stimuli well before an adequate test of generalization has been conducted.

A laboratory example with pigeons provides a means for examining issues related to measuring stimulus generalization. Jenkins and Harrison (1960) compared the extent to which stimulus generalization occurred in pigeons following discrimination training and in pigeons not exposed to discrimination training. For pigeons that did not experience discrimination training, key pecking was reinforced on a variable-interval schedule in the presence of a continuous tone of one frequency (i.e., a 1,000 cycle-per-second tone). Following this training, response rates in the presence of the generalization testing stimuli (i.e., tones of different frequencies ranging from 300-3,500 cycles per second) were measured under extinction conditions. Responding in the presence of the testing stimuli was similar to responding that occurred in the presence of the training stimulus and produced a flat generalization gradient. Even though all of the tone stimuli occasioned the same response, stimulus generalization could not be said to occur because the many exemplars of tone stimuli were not demonstrated to be physically discriminable from each other. Therefore, for this particular group, stimulus generalization could not be distinguished from a failure to discriminate testing stimuli from training stimuli. To determine the extent to which pigeons could discriminate the tone frequencies, differential responding of a second experimental group was evaluated.

The second group of pigeons was first exposed to discrimination training. Key-pecking was reinforced on a variable-interval schedule only in the presence of the 1,000 cycle-per-second tone (S+), whereas, key-pecking in the absence of the tone (S-) was not reinforced. Following discrimination training, response rates in the presence of tones of different frequencies ranging from 300-3,500 cycles per second and in the absence of any tone, were measured under extinction conditions. The highest rate of responding was observed in the presence of the S+ and the lowest rate of responding was observed in the presence of the S-, forming a peaked generalization gradient. Additionally, the rate of key pecking was higher in the presence of the stimuli that were more physically similar to the S+. The generalization gradient observed in this group in which pigeons showed differential responding across frequencies provided evidence that pigeons could discriminate among the frequencies. Therefore the results from the first group of pigeons could be taken as evidence of stimulus generalization, rather than a failure to discriminate.

Studies on generalized imitation with children also offer a means for measuring stimulus generalization (Baer & Sherman, 1964; Bear, Peterson, & Sherman, 1967; Steinman, 1970; Paone, 2006). In these studies, stimulus models that occasion reinforcement (S+ training stimuli) and that do not occasion reinforcement (S- probe stimuli) are successively presented to the subject. Differential responding to these training and probe stimuli occurs because the subject discriminates the topographical differences among the modeling stimuli, *and* fails to discriminate the reinforcement contingencies occasioned by each. For example, on one trial, clapping hands may be the modeling stimulus that occasions reinforcement and on the next trial, tapping table may be the modeling stimulus that is discriminative for extinction. The subject's differential response topography in the presence of these different modeling stimuli shows discrimination among the modeling stimuli from trial to trial. Therefore, flat generalization gradients may be observed during tests of generalized imitation, however such an outcome may not imply the failure of the subject to discriminate between the stimulus models. Nevertheless, non-differential probability of imitative responding on each of these trial types could provide a demonstration of the subject's failure to discriminate the different reinforcement contingencies in place for imitative responding in the presence of the discriminably different stimuli.

On a more general level, in addition to the use of demonstrably discriminable stimuli, drawing inferences regarding the occurrence of stimulus generalization requires a demonstration of a functional relation between the treatment procedure and responding in the presence of the testing stimuli. This demonstration is achieved only within the logic of an experimental design in which threats to internal validity, such as maturation, history, and testing are controlled (Kazdin, 1982). Multiple-baseline designs, reversal designs, multi-element designs, and changing criterion designs represent valid experimental designs in applied behavior analysis. Other response measures that are not functionally related to the treatment procedure, for example pre-post data, may provide information about learning, but such measures do not allow inferences about stimulus generalization, per se. One reason is that pre-post test measures fall outside the experimental manipulations associated with valid experimental designs.

Consider, for example, studies of generalized imitation using single-case experimental design. Baseline provides a measure of steady-state responding against which to compare responding with the introduction of treatment. When using a reversal design, for example, any change in responding from baseline to treatment may be replicated with sufficient regularity to allow the conclusion that the treatment produced the regular changes in responding. Threats to internal validity (e.g., history, maturation, testing, etc) are ruled out by the experimental design. Information regarding the occurrence of generalization from training can be obtained by measuring responding to different but physically similar non-reinforced probe stimuli, however, it is just as important to obtain steady-state baseline response measures in the presence of the probe stimuli. The logic of the experimental design requires measuring any changes in response to probe stimuli over baseline levels once treatment is introduced on the training stimuli. If this is done, researchers are in a good position to monitor changes in non-reinforced probe responding to determine the extent to which they reflect changes in responding to training stimuli.

In a successful single-case experimental design, probe responding that approximates or parallels changes in responding to the training stimuli provides clear evidence to support the inference that stimulus generalization has occurred and the extent to which it has occurred.

Such continuous and direct evaluation of stimulus generalization may not be practical in all situations. Perhaps because of limited resources, pre-post test measures have been used to infer that generalization from training has occurred (see Figure 1 for hypothetical example). For example, one may pre-test in the community, train in a treatment location, and post-test in the community again. In this case the pre-post measures do not occur during the experimental manipulations that constitute the experimental design, as described above, and therefore it is not possible to evaluate of the extent to which pre-post responding changes as a function of the independent variable. As a result, change in pre-post measures must be evaluated with the understanding that threats to internal validity may offer an alternative interpretation for its occurrence. One such alternative is that responding during the pre-post measures is evoked by stimuli that acquired stimulus control prior to the implementation of the independent variable. Such measures certainly imply that learning has occurred, but they do not clearly implicate that treatment procedure by itself as the cause for that learning. Therefore, pre-post test measures in and of themselves cannot allow conclusions to be drawn about stimulus generalization.

As a case in point, Jerome, Frantino, and Sturmey (2007) demonstrated an important change in behavior with developmentally disabled adults. The participants first learned a leisure skill (i.e., accessing Internet Web sites) in one condition, and later demonstrated that leisure skill in a non-trained environment during post-test measures. During the baseline condition, participants were provided access to Web sites for 5 min and the errorless learning and backward chaining procedures were not used. During the teaching condition, the experimenters used the backward chaining and errorless learning procedures to teach participants a 13-step response chain for accessing age appropriate Web sites on the Internet using the same computer as in the baseline condition. During post-teaching sessions, another computer in a different classroom became available and the experimenters measured the occurrence of the 13-step response chain in the presence of this new stimulus. The authors concluded that the skills generalized to a novel computer and termed these testing sessions “generalization probes.”

Because the post-test measures were not part of the across-subjects experimental design, it is difficult to determine the extent to which these responses represent stimulus generalization, because there was no measure of responding in the presence of the other computer as a function of the independent variables. Because of this, alternative explanations for the participants' performance in the presence of the other computer cannot be eliminated. Furthermore, there was no demonstration that the learners discriminated between the two computers in the first place. Nevertheless, it is important for service providers to know that the participants had learned to display these leisure skills using two different computers.

Eikeseth and Nesset (2003) used pre-post-test data to determine the extent to which improved articulation of sounds learned during vocal imitation training would result in an increase in accurate articulation of sounds under more natural speech conditions (i.e., picture-naming task and conversational speech). Because articulation of sounds improved for both participants from pre- to post-test measures, the authors inferred that generalization had occurred "from the contrived class of vocal imitation to conversational speech and picture-naming" (pgs. 334-335).

Inferring that the generalization of stimulus control across common stimulus features had occurred is imprecise. These measures were not part of the multiple-baseline design that was used to demonstrate experimental control by the vocal imitation training procedure, and therefore are subject to threats to internal validity. Instead of inferring stimulus generalization from these measures, we suggest that the demonstration of these responses in more natural contexts are attributable to other learning principles or other threats to internal validity, such as repeated testing or practice effects.

Interspersing generalization probe stimuli (S^- stimuli) throughout baseline and training within a multiple-baseline design or reversal design allows one to examine the extent to which the generalization probe data change, along with changes in the training data, as a function of changes in the independent variable (see Figure 2 for hypothetical example). Thus, it is safer to make inferences regarding the occurrence of stimulus generalization under these conditions. This more fine-grained analysis includes the generalization probe data in the experimental design itself, thus bolstering the argument that generalization from the training stimuli to generalization probe stimuli has occurred.

To illustrate the above point, Poulson, Kyparissos, Andreatos, Kymissis, and Parnes (2002) were able to conclude that generalized imitation occurred in three 12- to 14- month-old infants because of their use of non-reinforced generalization probe measures that were interspersed throughout the baseline and training conditions in an across-responses- multiple-baseline design. Mothers modeled motor-with-toy responses (e.g., banging a hammer), motor-without-toy responses (e.g., tap table), and vocal responses (e.g., *AH*) during baseline and treatment conditions. During baseline, only the model was presented and no praise was provided for imitative responding. During treatment, the mother presented the model from one of the three stimulus sets and praised the infant contingent upon responding that matched the model. Interspersed within the baseline and treatment sessions, were unreinforced probe stimuli. For each of the participants, the authors were able to infer the extent to which generalized imitation occurred within stimulus sets by analyzing how responding to non-reinforced probe stimuli increased systematically with the introduction of treatment for the reinforced training stimuli.

Similarly, Haring, Breen, and Laitinen (1989) used generalization probes that were interspersed every third or fourth training trial to assess generalization of the trained conditional discrimination between age-appropriate and age-inappropriate stimuli by adolescents with developmental disabilities. While these stimuli were being trained using a conditional discrimination training procedure, within-set generalization was being assessed by the interspersed presentation of untrained stimuli from that set. In addition, between-set generalization probes were also interspersed among the training trials to determine the extent to which the conditional discrimination training procedure promoted responding in the presence of stimuli from a different set.

Furthermore, concerning the certainty with which one may infer the presence of stimulus generalization, there are studies with less frequent measurement of responding to probe stimuli. Hence, there are inferences that may be made regarding generalization from training that fall between the extremes of pre-post measures and frequent probing throughout baseline and treatment. Late or delayed introduction of generalization probes during training following the interspersing of probe measures during baseline provides another analysis of stimulus generalization. Because the generalization probe data are not evaluated throughout the entire treatment condition (See Figure 3 for hypothetical example), inferences regarding stimulus generalization may be less strongly supported because there are fewer generalization data.

For example, Garcia, Baer, and Firestone (1971) evaluated the generalization of imitative responding by presenting generalization probe models that were similar to the trained small motor models, large motor models, and short vocal models. The generalization probe models were first presented in the pre-training session. Presentation of the generalization probes did not occur following training until acquisition and consistent demonstration of the trained imitative responses occurred. The authors concluded that a generalized imitative repertoire was established because of the imitative responding observed during the trials in which there were no programmed consequences for imitation. This conclusion is warranted, but note that it is based on less evidence than the above described studies with more frequent measures of probe responding.

Another variation in generalization testing procedures is the use of more frequent pre-post measures throughout baseline and training (see Figure 4 for hypothetical example). This type of procedure provides a limited amount of information needed to draw inferences regarding stimulus generalization. Reeve, Reeve, Townsend, and Poulson (2007) used frequent pre-and post-intervention measures to assess the extent to which helping responses learned under the training conditions, would be emitted in a novel room with a novel instructor by presenting both previously trained and novel helping scenarios. Specifically, they used 3 pre-test and 3 post-test measures. They inferred that the pre-post-test results showed generalization from training. Nevertheless, the use of frequent pre-post measures may not support a clear inference about generalization from training any more than does a single pre-post measure. The reason is that the post-test measures were obtained following, rather than during, the acquisition of the correct response. The increase in correct responding from baseline to treatment is systematically replicated in the experimental design, thus ruling out threats to internal validity, and allowing the conclusion that the treatment caused the change in the treatment data. Because the pre-post data were not included in that experimental demonstration, perhaps it should not be concluded that the treatment caused the change in the pre-post data. Thus, the authors may have been incorrect in their conclusion that the pre-post data demonstrated generalization from training.

Limiting Factors in the Inference of Stimulus Generalization

The extent to which learners respond similarly when presented with slightly physically different non-reinforced probe stimuli is a core interest when making inferences about stimulus generalization. A major challenge for learners is the discriminability of reinforcement contingencies that are associated with the presentation of S^+ and S^- stimuli. Specifically, to demonstrate stimulus generalization, responding must be controlled by the physical properties of the probe stimuli and not by those aspects of the probe stimuli that are discriminative for non-reinforcement. Thus, properties of stimuli may serve two different functions. Stimuli may signal the response topography to be emitted by the learner *and* they may signal the availability of reinforcement. To evaluate stimulus generalization, learners must discriminate the former rather than the latter and not respond differentially to changes in reinforcement contingencies associated with S^+ and S^- stimuli (i.e., training and probe stimuli).

There may be procedures that facilitate this complex discrimination problem for the learner. Interspersing generalization probes among training trials may better support the maintenance of responding when subjects are presented with non-reinforced probe trials, whereas massed testing under extinction conditions may reduce response rates. Massed testing involves the repeated presentation of probe stimuli that differ from training stimuli, following the acquisition of the trained response(s) (see Figure 5 for hypothetical example). Researchers must ask under what conditions massed testing will provide an effective measure of stimulus generalization. Arguably, this type of testing may be extremely intrusive to use in applied research because of the problems related to testing under extinction conditions. Irrespective of any experimental design issues, massed testing may be a better means of demonstrating stimulus generalization with laboratory populations such as pigeons that engage in conveniently high-rate responding (i.e., key-pecking) and that are exposed to deprivation procedures before training and high rates of reinforcement during training.

In a study on generalized imitation with children with developmental disabilities, Peterson (1968) demonstrated that massed testing of stimulus generalization following training was associated with reduced imitative responding, whereas interspersed testing of stimulus generalization among training trials was associated with maintained responding in the presence of testing stimuli. This work highlighted the importance of arranging contingencies during tests of stimulus generalization that are less discriminable. In this case, the contingencies occasioned by the testing stimuli presented were less discriminable from the contingencies occasioned by the training stimuli and did not interfere with response acquisition during their interspersed presentation.

Two important procedural questions may be derived from Peterson (1968). First, is there an optimal ratio of training stimuli to generalization probe stimuli when testing for stimulus generalization in applied research? We have found no experimental research bearing directly on this question. Second, how do we make the predictive value of the probe stimuli for reinforcement less discriminable? Poulson (2009) offered some practical suggestions for using interspersed generalization probes. Specifically, she suggested using probe stimuli and training stimuli from the same physical stimulus modality. For example, when training imitative responding in the presence of gross-motor stimulus models, present non-trained gross-motor stimuli, not verbal stimuli, to probe for stimulus generalization. Also, probe stimuli may be presented using a variable, rather than a fixed schedule of interspersed presentation. Instead of presenting a probe trial after a fixed number of training trials, probe trials may be presented following a random sequence of training trials (e.g., the first probe may be presented after seven training trials, the next probe may be presented after four training trials, the following probe may be presented after ten training trials, then following three training trials, and so on) (Baer, Peterson, & Sherman, 1967).

Ajar (2007) used an interesting procedure to make the reinforcement value of the generalization probes less discriminable when teaching youth with autism to engage in appropriate affective verbal and nonverbal responding. In this study, generalization probe trials were initially easily discriminable by the participants because, unlike the training trials, the written scripts were not used during probe trials to prompt target affective responses. Therefore, the presence of written scripts became discriminative for reinforcement, whereas the absence of scripts became discriminative for extinction. As a result, one participant did not respond during probe trials. To resolve the problem of such unwanted discrimination between training and probe stimuli, the experimenters removed the scripts for some of the training trials, but retained the correction procedure during these trials. Following this change, a breakdown in the discrimination between training and probe trials occurred such that there was an increase in responding during probe trials for that participant.

Another suggestion for making the reinforcement value of probe stimuli less discriminable from that of training stimuli was demonstrated by Garcia, Baer, and Firestone (1971) and was mentioned previously in this paper. The experimenters trained a set of imitative responses to a criterion level and later switched to an intermittent schedule of reinforcement before interspersing probe stimuli. This procedure was designed to increase resistance to extinction during the interspersed presentation of non-reinforced probe stimuli.

Shroder and Baer (1972) also reduced the probability of reinforcement for responding during training trials before introducing interspersed generalization probes. Additionally, during a comparison of two training procedures, concurrent and serial, the experimenters found that responding to probe stimuli had a higher probability following concurrent training (i.e., three responses simultaneously trained to criterion) than responding to probe stimuli following serial training (i.e., one response trained to criterion). Shroeder and Baer argued that probes following serial training might have been more readily discriminated as stimuli that did not occasion reinforcement than were probes following concurrent training. The authors called for future research to address the parameters of such procedures, specifically the discriminability of the reinforcement value of probe stimuli.

Each of these studies shows sensitivity to the issue that the early introduction of generalization probes, for whatever reason, may interfere with initial acquisition of target responding in addition to control that these stimuli may or may not have based on their physical or predictive properties. Terrace (1963) showed that early, as opposed to late, introduction of S- stimuli facilitated discrimination between S+ and S- stimuli. In that study, pigeons learned the discriminated operant of pecking in the presence of a red key (S+) and not pecking in the presence of a green key (S-). The effects of early versus the late introduction of the S- stimulus on responding in the presence of the S+ and S- stimuli were evaluated. Subjects acquired the discrimination between red and green with fewer responses to the S- (i.e., few errors) when the S- was presented early in discrimination training as opposed to late introduction. Terrace (1963) concluded that the manner in which the S- was introduced had a systematic effect on the acquisition of the discrimination. In a later study with adult humans, Fields (1981) also reported that the early or late introduction of probe stimuli that did not occasion reinforcement into the training protocol, similarly affected accurate naming of Braille stimuli.

The manner in which early or late introduction of probe stimuli intersects with the problem of discriminability between training (S+) and probe (S-) stimuli on the basis of reinforcement value is unknown. Nevertheless, if one waits past the acquisition phase on the training trials to introduce probe stimuli, the procedure more closely resemble pre-post- testing, because the probe stimuli cannot be shown to co-vary with the training stimuli in trend and in level. Because the introduction of probe stimuli may have a direct influence on the acquisition of the target response, one must take this into consideration when designing an optimal probe procedure to measure stimulus generalization.

Another means of making the reinforcement value of probe stimuli less discriminable may be to reinforce responding in their presence. This procedure requires the use of trial-unique stimuli, meaning that every probe trial consists of a stimulus that is only used once, before its association with reinforcement to measure generalization. Therefore, reinforcing responding on these trials may not interfere with the measurement of generalization and will help to prevent extinction of responding during generalization testing. The use of trial-unique stimuli to test for stimulus generalization within a matching-to-sample paradigm has proven effective in preventing problems that emerge when reinforcement is discontinued.

Brown, Brown, and Poulson (1995) used trial-unique stimuli to test for stimulus generalization by young children within a matching-to-sample paradigm. Initially, pre-training sessions were conducted to familiarize the children with the matching to sample procedure and consisted of three steps; presenting one sample and one comparison stimulus, presenting one sample stimulus and two comparison stimuli, and presenting one sample stimulus and three comparison stimuli. Once participants responded correctly on 80% of the final ten trials during the third pre-training step, generalization testing commenced. During generalization testing, one reinforced test trial was interspersed between presentations of three baseline-training trials (these trials were the same configurations as those from the three comparison pre-training step). Among the 150 test trials, no stimulus set appeared more than once. In fact, each configuration of sample and comparison stimuli was used only once allowing correct responding to be reinforced and preventing difficulties that emerge when reinforcement is discontinued. All children showed generalization of matching to sample to novel stimuli during the first block of generalization testing.

Comparison studies have demonstrated that in a matching-to-sample paradigm, use of multiple trial-unique choice stimuli results in higher response rates during stimulus generalization testing than does training with only two choice stimuli that are not trial unique. Wright, Cook, Rivera, Sands, and Delius (1988) trained matching-to-sample responses in one group of pigeons using the same two video picture stimuli and the trial-unique group of pigeons was trained using 152 different video picture stimuli daily. Following training, both groups were tested for generalization by interspersing ten testing trials using a ratio of 3 training trials to 1 generalization testing trial. The 2-stimulus group responded at chance levels during generalization testing. Pigeons in the trial-unique group matched correctly on at least 80% of the generalization test trials.

Williams, Johnston, and Saunders (2006) compared the use of multiple a trial-unique choice stimuli with a procedure in which only two choice stimuli were used to test delayed matching-to-sample in adults with mental retardation. In the trial-unique procedure, every trial in a session contained different stimuli. For the other procedure, the same two comparison stimuli were presented on each trial with their positions counterbalanced. The highest accuracy was observed with the trial-unique procedure. The authors concluded that responding during the procedure in which only two choice stimuli were used was controlled by the stimulus that occasioned reinforcement on the previous trials. So, instead of matching, the subject may have selected the comparison stimulus that was selected on the previous trial, especially if selection of that comparison stimulus occasioned reinforcement. This source of competing stimulus control is not present in the trial-unique procedure because the comparison stimuli are different on every trial. Additionally, this source of stimulus control could mask the extent to which generalization of the matching response would occur following training.

Stimulus Generalization Gradients and Variations of Generalization Tests in Basic Research

Massed testing, as discussed previously in this paper, has been used to measure generalization in the absence of reinforcement and to produce generalization gradients. Research in the experimental analysis of behavior with non-human subjects has suggested some additional workable solutions to the issues that are common when testing for generalization in the absence of programmed reinforcement. In their classic paper, Guttman and Kalish (1956) trained a discrimination using an intermittent schedule of reinforcement before introducing massed testing of stimulus generalization. The massed testing procedure included the presentation of many different stimulus values under extinction conditions. Decremental generalization gradients emerged during testing and responding in the presence of the test stimuli was analyzed as resistance to extinction.

In their review of the research on stimulus generalization, Honig and Urcuioli (1988) discussed variations from those procedures used originally by Guttman and Kalish (1956) and the effects of different types of measurements on the resulting generalization gradients. These variations included the single-stimulus test procedure, resistance to reinforcement procedure, and steady-state testing. In a single-stimulus testing procedure, each subject is presented with one of the stimulus values from the stimulus dimension being evaluated. Responding by each subject to the individual stimulus values is used to produce the generalization gradient. Generalization gradients emerge that are similar to those that emerge when testing involves the presentation of all of the dimensional values each subject. Although this provides some evidence that one stimulus may be sufficient to demonstrate stimulus generalization, it requires that the experimenter be sure that the organism has no past learning history with the stimulus chosen for testing. This is often possible in research with naïve laboratory animals, but it is typically not the case with human subjects. Using few stimuli to test for generalization is dangerous in that one can lose the generalization measure if the subject had a learning history with that stimulus that is unknown to the experimenter. The use of only two or three generalization stimuli is problematic for the same reason, but this procedure may be used to keep session length relatively short.

Stimulus generalization may be masked by low response rates caused by extinction procedures used during massed testing. Hearst, Besley, and Farthington (1970) developed the reverse of the generalization testing procedure used by Guttman and Kalish by using inhibitory generalization gradients to test for stimulus generalization. In their resistance to reinforcement procedure, instead of extinguishing behavior at all values during testing, reinforcement was provided for responding at all test values. If some of the S- stimuli were more inhibitory than others, decremental response rates would emerge in their presence resisting the effects of reinforcement and providing a measure of responding in the presence of testing stimuli.

Finally, Honig and Urcuioli (1981) discussed steady-state testing as a means of offsetting the undesirable effects of extinction during generalization testing. In a steady-state testing procedure, reinforcement is provided for responding to training stimuli as non-reinforced generalization testing trials are interspersed. Testing of this sort has been conducted in studies of generalized imitation since its inception (Baer, et al., 1967).

Competing Sources of Stimulus Control in Stimulus Generalization Testing

At this point, we should discuss some exceptions to issues we have raised regarding extinction during tests of stimulus generalization. Competing sources of stimulus control, specifically compliance with instructions, may be an example of this exception. Many studies have demonstrated that a decrement in responding many not always occur when non-reinforced probe stimuli are introduced to measure generalization. Peterson and Whitehurst (1971) demonstrated that imitative responding of children of typical development was especially resistant to extinction procedures. Imitation persisted for two of the three subjects when reinforcement was removed completely and when differential reinforcement was provided for non-imitative responding. In the final phase of the experiment, the experimenter was absent and another adult presented imitation trials. For all three subjects, correct imitative responding reduced to zero levels. Therefore, behavior of subjects with a learning history of compliance with adult instructions may not extinguish, whereas, others, may.

In another study on generalized imitation conducted with children with mental retardation, Burgess, Burgess, and Esveldt (1970) found that two of the three subjects imitated probe models when first presented and continued to imitate these models even when reinforcement was removed for imitating training models. The authors concluded that subject responding was under instructional control.

These studies highlight other possible sources of stimulus control, and by inference, reinforcer control, other than that by the programmed conditional stimuli. Specifically, both overt and covert instructions provided by adults may control responding in subjects with a history of reinforcement for compliance.

Other Demonstrations of Learning

Research and knowledge in the field of applied behavior analysis has progressed substantially in its recognition of the need to produce behavior change in various ways that extend beyond that strictly defined by treatment. The demonstration of stimulus generalization is one way such change is demonstrated, and the conditions under which such change may be evaluated has been the focus of this paper. At this point, it should be noted that there are other important behavior change phenomena that are not the product of stimulus generalization. In particular, learning may also have occurred when response generalization, maintenance of responding, and learning to learn is observed.

Response generalization provides information to researchers about the way in which behavior change extends beyond the boundaries of treatment. Catania (1998) defined response generalization as the spread of effects of reinforcement to responses outside the boundaries of the trained operant class. As a case in point, Antonitis (1951), placed rats in a chamber with a horizontal slot along one wall. When the nose poking response was reinforced at the center of the slot, nose poking at other positions near the center were observed, indicating the spread of effects of reinforcement to variations in the nose poking response class.

In applied research, a major outcome of script-fading procedures is response generalization. Krantz and McClannahan (1993) taught youth with autism to initiate conversations to their peers by using written scripts. Once the scripts were faded, unscripted responses, responses that varied from the trained responses, were emitted routinely. This outcome demonstrated the occurrence of different responses in the presence of the same stimulus conditions, the presence of a conversational partner.

Maintenance of responding, that is, responding that persists after all or a portion of the treatment has been removed, can be another valuable form of learning (Cooper, Heron, & Heward, 1987). Koegel and Rincover (1977) emphasized the distinction between the failure of a response to generalize from a training environment to a non-training environment and the failure of a learned response to maintain over time. In this study three children with autism learned imitative responses and to follow one-step instructions in a training environment with the experimenter. Two of the children performed these responses in the absence of programmed reinforcement outside of the training environment with another adult, demonstrating generalized responding. Nevertheless, with continuous measurement, the generalized response extinguished over time. The authors argued that this was a problem of maintenance. Another child never demonstrated the learned imitative responses in the non-training environment, so the authors argued that this was a problem of generalization.

One often overlooked, but powerfully important form of learning in applied research is learning to learn. For example, it is sometimes critical to demonstrate that subjects learn the next target response in less time and with fewer reinforced trials, particularly by people with developmental disabilities. Baer, Peterson, and Sherman (1967) demonstrated that children with developmental disabilities required fewer trials to reach criterion when learning to imitate new gross motor models as training progressed.

Conclusions

Inferences regarding stimulus generalization are more safely made when consideration is given to the experimental design and the discriminability of the reinforcement contingencies associated with training and testing stimuli. As reviewed in this paper, using non-reinforced probes interspersed among reinforced training trials during baseline and treatment provides important information needed to draw inferences regarding stimulus generalization. This type of procedure also may enhance the discriminability of the physical properties of the testing stimuli while diminishing the discrimination of their reinforcement contingencies, thereby reducing problems associated with extinction when testing for generalization from training. It is also noteworthy that interspersing generalization probes may interfere with the beginning stages of skill acquisition. Nevertheless, the procedural variations described in this paper may provide examples of ways to address this issue. The use of reinforcement in trial-unique procedures also reduces problems associated with extinction when testing for generalization from training. Finally, extinction may not always interfere with the use of generalization probes, as was evident in studies demonstrating other sources of stimulus control.

As described above, there are many valuable inferences one may draw from data about learning apart from stimulus generalization. These include inferences regarding response generalization, maintenance, and learning to learn. In the final analysis, all are valuable inferences from treatment data even when they are not all the product of stimulus generalization. We may find it helpful to refine our language to more accurately describe the types of measurement and experimental design procedures we are using and the resulting inferences we may draw from them, and to not refer to them all as stimulus generalization.

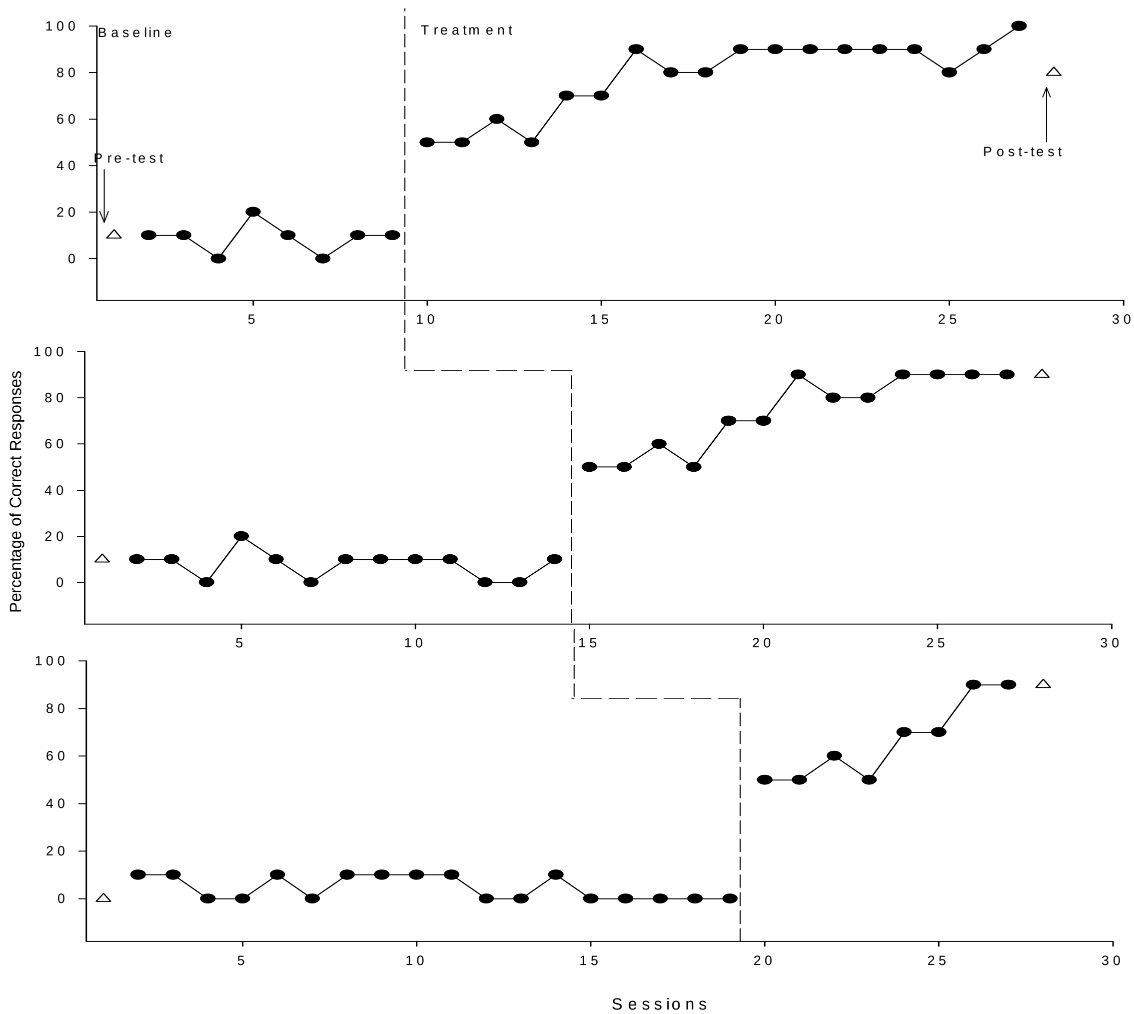

Figure 1A. Hypothetical example of pre-post test measures within a multiple- baseline design.

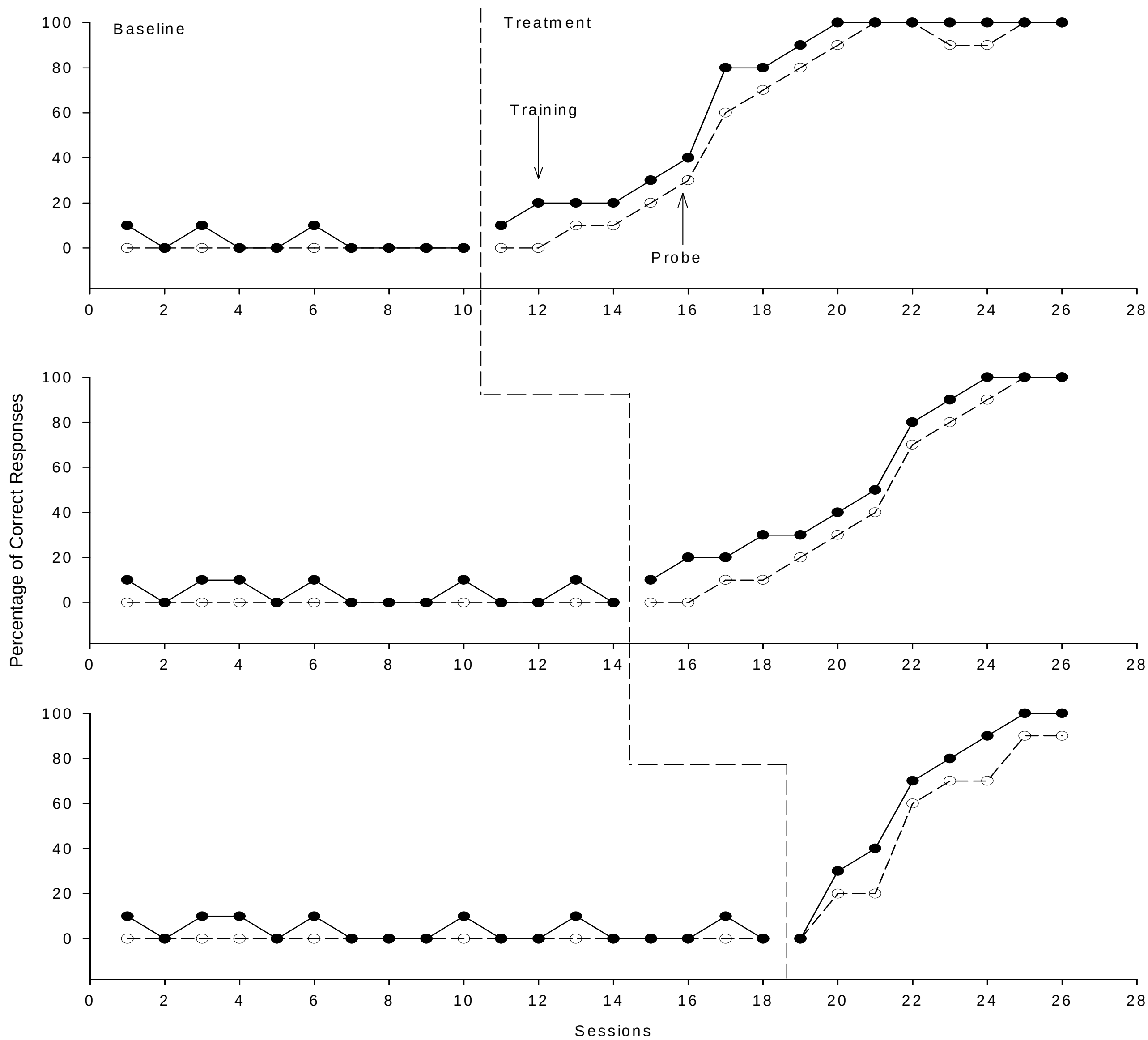

Figure 2A. Hypothetical example of interspersed generalization probes throughout baseline and training conditions.

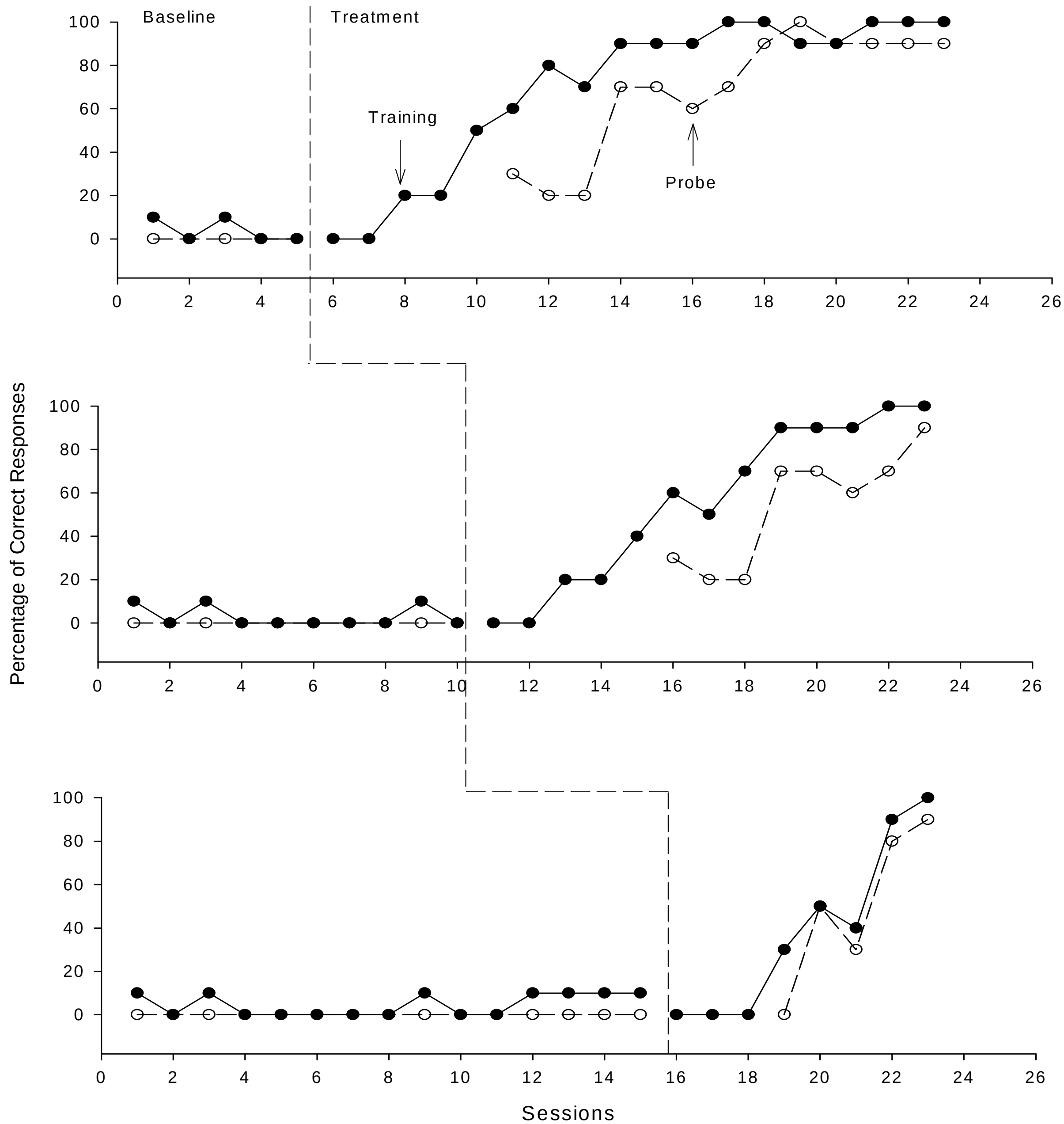

Figure 3A. Hypothetical example of generalization probes interspersed during baseline and then delayed in their introduction during training within a multiple- baseline design.

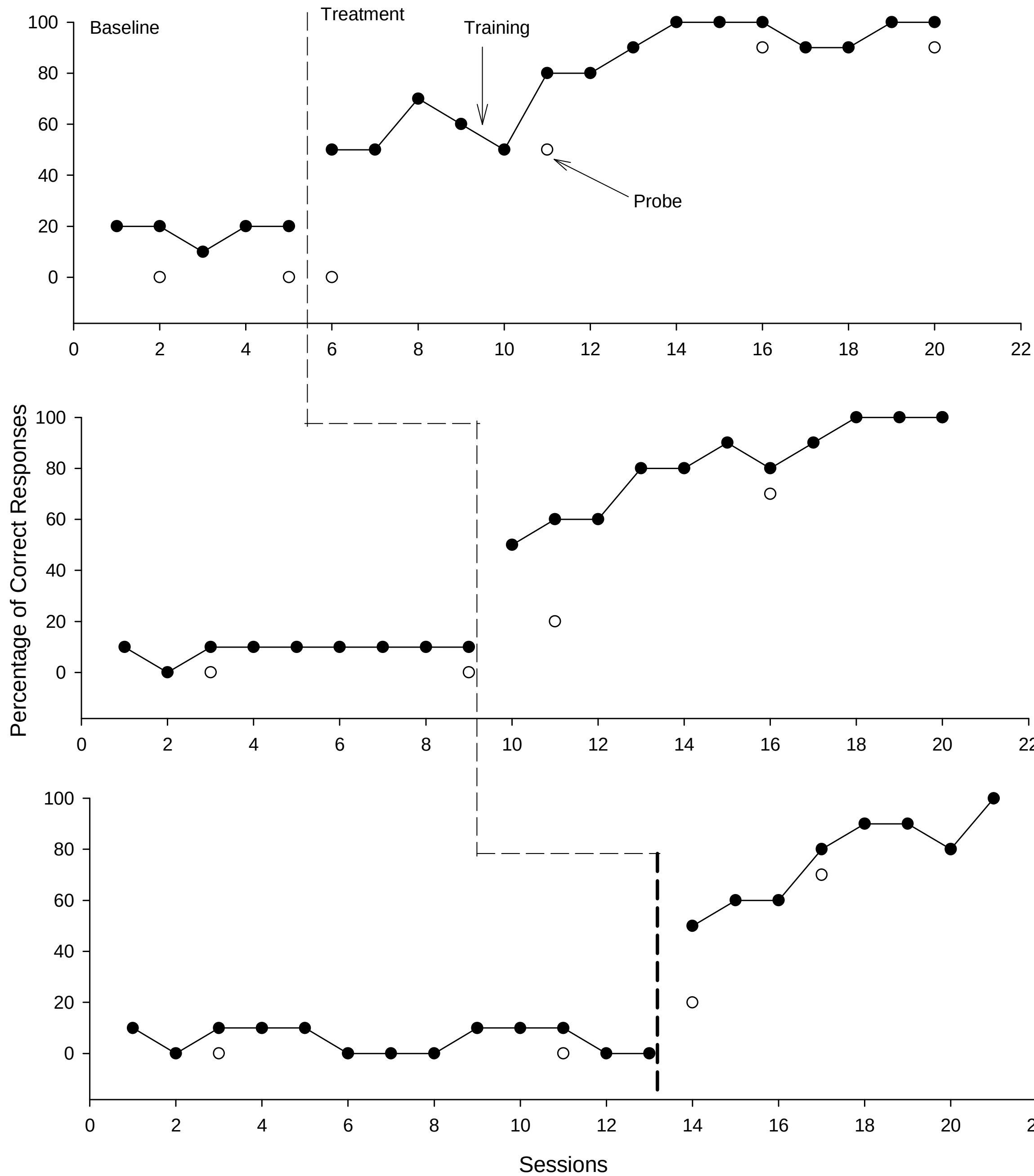

Figure 4A. . Hypothetical example of infrequent interspersed presentation of generalization probes within a multiple-baseline design.

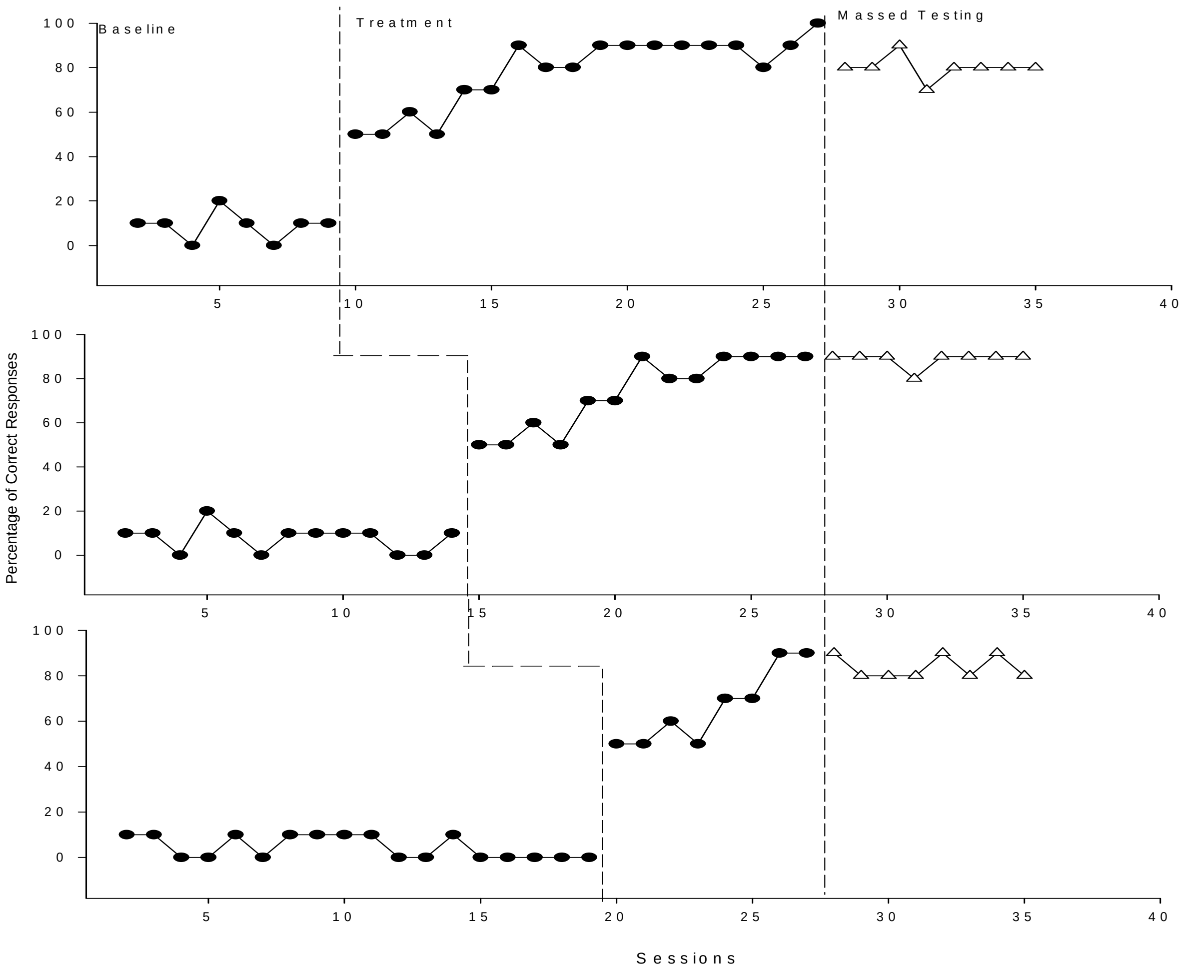

Figure 5A. Hypothetical example of massed generalization testing following baseline and training conditions.

Bibliography

Antonitis, J. J. (1951). Response variability in the white rat during conditioning, extinction, and reconditioning. *Journal of the Experimental Psychology, 42,* 273-281.

Baer, D.M. & Sherman, J. A. (1964). Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology, 1,* 37-49.

Baer, D.M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior, 10,* 405- 416.

Brim, D. , Townsend, D.B., DeQuinzio, J.A., & Poulson, C.L. (in press). Analysis of social referencing skills in children with autism. *Research in Autism Spectrum Disorders.*

Brown, A. K., Brown, J. L., & Poulson, C. L. (1995). Generalization of children's identity matching-to-sample performances to novel stimuli. *The Psychological Record, 45,* 29-43.

Burgess, R. L., Burgess, J. M., & Esveldt, K. C. (1970). An analysis of generalized imitation. *Journal of Applied Behavior Analysis, 3*, 39-46.

Campos, J. J., & Sternberg, C. (1981). Perception, appraisal, and emotion: The onset of social referencing. In M. E. Lang & L. R. Sherrod (Eds.), *Infant Social Cognition: Empirical and Theoretical Considerations* (pp. 273-314). Hillsdale, NJ: Erlbaum.

Catania, A. C. (1998). *Learning.* 4th Ed. New Jersey: Prentice-Hall Inc.

Cooper, J.O., Heron, T. E., & Heward, W. L. (1987). *Applied Behavior Analysis.* New Jersey: Prentice-Hall Inc.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring stimuli. *Journal of Autism and*

Developmental Disorders, 28, 479-485.

Eikeseth, S., & Nesset, R. (2003). Behavioral treatment of children with phonological disorder: The efficacy of vocal imitation and sufficient response-exemplar training. *Journal of Applied Behavior Analysis, 36*, 235-337.

Feinman, S. (1992). In the broad valley: an integrative look at social referencing. In S.Feinman (Ed.), *Social Referencing and the Social Construction of Reality in Infancy* (pp.3-13). New York: Plenum.

Fields, L. (1981). Early and late introduction of probes and stimulus control acquisition in fading. *Journal of the Experimental Analysis of Behavior, 36*, 363-370.

Fields, L., & Reeve, K. F. (2000). Synthesizing Equivalence Classes and Natural Categories from Perceptual and Relational Classes, in J. C. Leslie and D. Blackman (Eds.) *Experimental and Applied Analysis of Human Behavior,* pp. 59-83. Reno, NV: Context Press.

Garcia, E., Baer, D. M., & Firestone, I. (1971). The development of generalized imitation within topographically determined boundaries. *Journal of Applied Behavior Analysis, 4,* 101-112.

Gewirtz, J. L. & Pelaez-Nogueras, M. (1992). Social Referencing as a Learned Process, in S. Feinman (Ed.) *Social Referencing and the Social Construction of Reality*, pp.151-173. New York, NY: Plenum Press.

Guttman, N., & Kalish, H. J. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51*, 79-88.

Haring, T. G., Breen, C. G., & Laitinen, R. E. (1989). Stimulus class formation and concept learning: Establishment of within- and between-set generalization and transitive relationships via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *52*, 13-25.

Hearst, E., Besley, S., & Farthington, G. W. (1970). Inhibition and the stimulus control of operant behavior. *Journal of the Experimental Analysis of Behavior, 14,* 373-409.

Honig, W. K., & Urcuioli, P. J. (1981). The legacy of Guttman and Kalish (1956): 25 years of research on stimulus generalization. *Journal of the Experimental Analysis of Behavior, 36, 405-445.*

Jenkins, H.M., & Harrison, R. H. (1960). Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology, 59*, 246-253.

Jerome, J., Frantino, E. P., & Sturmey, P. (2007). The effects of errorless learning and backward chaining on the acquisition of Internet skills in adults with developmental disabilities. *Journal of Applied Behavior Analysis, 40,* 185-189.

Koegel, R. L. & Rincover, A. (1977). Research on the difference between generalization and maintenance in extra-therapy responding. *Journal of Applied Behavior Analysis, 10,* 1-12.

Krantz, P. J. & McClannahan, L. E. (1993). Teaching children with autism to initiate to peers: Effects of a script-fading procedure. *Journal of Applied Behavior Analysis, 26*, 121-132.

Leslie, J. C. (1996). *Principles of Behavioral Analysis.* Australia: Harwood Academic Publishers.

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin, 86,* 1236-1254.

MacDuff, G. S., Krantz, P. J., & McClannahan, L. E. (1993). Teaching children with autism to use photographic activity schedules: Maintenance and generalization of complex response chains. *Journal of Applied Behavior Analysis, 26*, 89-97.

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*. Austin, TX: PRO-ED.

Najjar, N.K. (2007). *Analysis of three components of affective behavior in children with autism.* Unpublished doctoral dissertation, City University of New York, New York.

Paone, D. (2006). *An examination of stimulus generalization in generalized imitation.* Unpublished doctoral dissertation, City University of New York, New York.

Peterson, R. F. (1968). Some experiments on the organization of a class of imitative behaviors. *Journal of Applied Behavior Analysis*, *1*, 225-235.

Peterson, R. F., & Whitehurst, G. J. (1971). A variable influencing the performance of generalized imitative behaviors. *Journal of Applied Behavior Analysis, 4*, 1-9.

Poulson, C. L. (2009). Behavioral theory and language acquisition. In A. Fitzer & P. Sturmey (Eds.), *Language and Autism: Applied Behavior Analysis, Evidence, and Practice*. Austin, TX: Pro-ed.

Poulson, C. L., Kyparissos, N., Andreatos, M., Kymissis, E., & Parnes, M. (2002). Generalized imitation within three response classes in typically developing infants. *Journal of Experimental Child Psychology, 81*, 341-357.

Reeve, S. A., Reeve, K. F., Townsend, D. B., & Poulson, C. L. (2007). Establishing a generalized repertoire of helping behavior in children with autism. *Journal of Applied Behavior Analysis, 40*, 123-136.

Saunders, K.J., & Spradlin, J.E. (1989). Conditional discrimination in mentally retarded adults: The effects of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior, 52*, 1-12.

Schlinger, H. D. (1995). *A Behavior Analytic View of Child Development.* New York: Plenum Press.

Schuster, R. H. (1969). A functional analysis of conditioned reinforcement. In D.P. Hendry (Ed.),*Conditioned Reinforcement* (pp. 192-234). Homewood, IL: Dorsey Press.

Shroeder, G. L., & Baer, D. M. (1972). Effects of concurrent and serial training on generalized vocal imitation in retarded children. *Developmental Psychology, 6*, 293-301.

Sigman, M. D., Kasari, C., Kwon, J. & Yirmiya. (1992). Responses to the negative emotions of others by autistic, mentally retarded, and normal children. *Child Development, 63,* 796-807.

Steinman, W. M. (1970). Generalized imitation and the discrimination hypothesis. *Journal of Experimental Child Psychology, 10,* 79-99.

Stokes, T. & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349-376.

Terrace, H. S. (1963). Discrimination learning with and without errors. *Journal of the Experimental Analysis of Behavior, 6,* 1-27.

Vorndran, C. M. & Lerman, D. C. (2006). Establishing and maintaining treatment effects with less intrusive consequences via a pairing procedure. *Journal of Applied Behavior Analysis, 39,* 35-48.

Warreyn, P., Roeyers, H., & De Groote, I. (2005). Early social communicative behaviors of preschoolers with autism spectrum disorder during interaction with their mothers. *Autism, 9*, 342-361.

Williams, G., Carnerero, J. J., & Perez-Gonzalez, L. A. (2006). Generalization of tacting actions in children with autism. *Journal of Applied Behavior Analysis, 39,* 233-237.

Williams, D. C., Johnston, M. D., & Saunders, K. J. (2006). Intertrial sources of stimulus control and delayed matching-to-sample performance in humans. *Journal of the Experimental Analysis of Behavior, 86*, 253-267.

Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning and Behavior, 16*, 436-444.

Young, J. M., Krantz, P. J., McClannahan, L.E., & Poulson, C. L. (1994). Generalized imitation and response class formation in children with autism. *Journal of Applied Behavior Analysis, 27*, 685-697.