Published in final edited form as:

Psychol Rec. 2008 January 1; 58(1): 15–36.

REVERSAL LEARNING SET AND FUNCTIONAL EQUIVALENCE IN CHILDREN WITH AND WITHOUT AUTISM

Karen M. Lionello-DeNolf,

University of Massachusetts Medical School-Shriver Center, Worcester

William J. McIlvane,

University of Massachusetts Medical School-Shriver Center, Worcester

Daniela S. Canovas,

Universidade Federal de São Carlos, Sao Carlos, Brazil

Deisy G. de Souza, and

Universidade Federal de São Carlos, Sao Carlos, Brazil

Romariz S. Barros

Universidade Federal do Pará, Belem, Brazil

Abstract

To evaluate whether children with and without autism could exhibit (a) functional equivalence in the course of yoked repeated-reversal training and (b) reversal learning set, 6 children, in each of two experiments, were exposed to simple discrimination contingencies with three sets of stimuli. The discriminative functions of the set members were yoked and repeatedly reversed. In Experiment 1, all the children (of preschool age) showed gains in the efficiency of reversal learning across reversal problems and behavior that suggested formation of functional equivalence. In Experiment 2, 3 nonverbal children with autism exhibited strong evidence of reversal learning set and 2 showed evidence of functional equivalence. The data suggest a possible relationship between efficiency of reversal learning and functional equivalence test outcomes. Procedural variables may prove important in assessing the potential of young or nonverbal children to classify stimuli on the basis of shared discriminative functions.

The capability of behaving organisms to respond to physically dissimilar stimuli as if they are the same has been of considerable interest to researchers who study both human and nonhuman beings from various theoretical and methodological perspectives (e.g., human and nonhuman cognitive psychology, Pavlovian conditioning, and behavior analysis). In recent years, one focus of interest has been the phenomenon of stimulus equivalence as defined by Sidman and Tailby (1982). According to their set theory definition, stimuli are said to be equivalent if the individual's behavior toward them exhibits the relational properties of reflexivity, symmetry, and transitivity—typically as exhibited in emergent stimulus-stimulus relations in conditional matching-to-sample procedures. The emergence of new relations without explicit discrimination training, combined with the formal relationship to mathematical equivalence,

Address correspondence to Karen Lionello-DeNolf, University of Massachusetts Medical School, Shriver Center, 200 Trapeld Rd, Waltham, Massachusetts 02452. Karen.Lionello-DeNolf@umassmed.edu.

Publisher's Disclaimer: Copyright of Psychological Record is the property of Psychological Record and its content may not be copied or emailed to multiple sites or posted to a listserve without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

constitutes the principal argument that stimulus equivalence is the foundation for true symbolic matching to sample.

Vaughan (1988), building on an earlier suggestion by Lea (1984), used another set-theoretic definition that led to a different method for establishing and verifying functional stimulus equivalence. Vaughan used the concept of the *partition*: division of a given set into two smaller subsets. Each member of each subset is defined mathematically as equivalent to the other member or members of the subset containing that member, if the union of the two subsets is equal to the universe and if the intersection between the two subsets corresponds to a blank set. In Vaughan's implementation, a set of 40 stimuli was divided arbitrarily into two nonintersecting subsets through the establishment of stimuli of one subset as positive (20 S+) and those of the other as negative (20 S-) in a successive (i.e., go, no-go) discrimination procedure with pigeons. When the pigeons mastered the 20 S+ versus 20 S- problems, the discriminative functions of the subsets were abruptly reversed. Originally positive stimuli became negative, and vice versa. When this discrimination reversal problem was mastered, the contingencies were re-reversed, thus reinstating the original discrimination contingencies. Such reversals were programmed many more times over the 2-year course of the study.

Used in this manner, the repeated-reversal procedure allows one to ascertain (1) whether subjects become more efficient in mastering reversed contingencies as training progresses (often termed "reversal learning set") and (2) whether spontaneous discrimination function reversals are observed, that is, if experiencing the consequences of the reversed contingencies with some members of a subset leads to spontaneous discriminative function reversals with other members without explicit discrimination training. The second of these outcomes—demonstrated and emphasized in Vaughan's study—indicates the formation of functional stimulus classes, one way of demonstrating functional equivalence.

Since Vaughan's pioneering study, the contingency-reversal procedures have been replicated with verbally able adolescents and adults (McIlvane, Dube, Kledaras, Iennaco, & Stoddard, 1990; Sidman, Wynne, Maguire, & Barnes, 1989) and with sea lions having extensive conditional discrimination training histories (Kastak, Schusterman, & Kastak, 2001). Findings have often been positive. Very little is known, however, about the conditions under which functional stimulus classes established in this manner can be demonstrated in the absence of verbal ability or an extensive history of intraexperimental discrimination training.

Also yet to be established is the relationship, if any, between reversal learning set and functional equivalence. It seems possible that establishing the behavioral prerequisites for reversal learning set might promote positive outcomes on tests for functional equivalence. Indeed, the spontaneous reversals observed in contingency reversal experiments could be viewed as a logical endpoint of a continuum of reversal learning set effects: maximally efficient behavioral allocation within the concurrent schedules programmed via the reversal learning procedure.

Our group has an ongoing interest in assessing the potential of populations with limited-to-nonexistent verbal abilities to exhibit relational stimulus control consistent with equivalence relations. For example, we recently reported a study that demonstrated stimulus equivalence according to the criteria outlined by Sidman and Tailby (1982) in persons with minimal verbal ability (Carr, Wilkinson, Blackman, & McIlvane, 2000). To our knowledge, no one has yet studied the potential of humans who lack verbal abilities *and* extensive discrimination training experience to form functional stimulus classes. Notably, the positive outcomes reported by Vaughan (1988) and Kastak and colleagues (2001) were obtained only after animals had undergone years of directly relevant preparatory discrimination training. To what extent is it possible to obtain positive contingency-reversal outcomes with nonverbal persons within a time frame more typical of research with humans? This question is of both academic and

practical interest. As yet, we have very little information about the capability of nonverbal persons to acquire and potentially expand a repertoire of true symbolic relations. Moreover, we do not presently know whether individuals from this population can exhibit learning set.

The research reported here was conducted as part of a larger program to assess the relational learning potential of populations who are not generally thought to function symbolically. The program includes not only nonhuman primates (e.g., Galvão et al., 2005) but also preverbal children (Gil & Oliveira, under review) and nonverbal children with autism spectrum disorders. The second study reported here addressed the last of these populations. We report also a companion study, conducted with normally capable preschool children, that was run in parallel to verify the integrity of the procedures and to serve as contrast data for our work with the nonverbal children. Although our typical-child study is reported first in the present sequence, it was in fact launched somewhat after our study of children with autism spectrum disorders. The procedures of Experiment 1 were modified in part by suggestions from early results from Experiment 2.

Experiment 1

There is a substantial literature on discrimination reversal learning with typically and atypically developing children (e.g., Blank, 1966; Heal, Ross, & Sanders, 1966; House, 1964; Kaufman & Gardner, 1969; Plenderleith, 1956). Much of it focused on the ability to learn a single discrimination reversal after extradimensional shift learning or extended experience with nonreversed simple discrimination problems. Notable by their absence are studies of reversal learning set in children (i.e., does learning become more efficient in the course of repeated reversal training?). Such data seemed more directly relevant to yoked-contingency reversal procedures than those available in the current literature. Thus, we deemed it necessary to characterize the performance of typically developing children to establish a data basis for contrasts with the performance of nonverbal children on similar procedures. To that end, children were first trained on three simple discrimination problems using three different stimulus sets. The reinforcement contingencies were then yoked and reversed repeatedly to determine whether the children would exhibit reversal learning set and functional equivalence.

Method

Participants—The participants were 6 typically developing preschool children; their age, gender, and language-age equivalent are listed in Table 1. Language-age equivalence scores were obtained with the Peabody Picture Vocabulary Test-R (Dunn & Dunn, 1981), translated into Portuguese from English and compared against English-speaking norms. All children attended a local day-care facility in São Carlos, Brazil. Experimental sessions took place at the day-care facility, were typically conducted 5 days per week, and lasted approximately 10 min (not including scheduled post session play time, described below). Generally, only one session per day was conducted, except for participant Lia, who occasionally requested two sessions.

Setting, Apparatus, and Stimuli—Sessions took place in a quiet room furnished with a table and chairs. An Apple Macintosh Performa 6360 computer, running MTS version 11.6 software (Dube, 1991; Dube & Hiris, 1996), was used to control all experimental events and to collect data. Stimuli were six non representative, three-dimensional color forms presented on a computer touch screen (see Figure 1). The stimuli were divided into three sets of two (Sets A, B, and C).

Procedure

Pretraining: Participants underwent 3–10 of such sessions with physically dissimilar stimuli that were not used subsequently. Initial pretraining sessions consisted of 15 trials, and over successive sessions, the number of trials was increased to 20, 25, and 30.

During the first five trials, only the S+ was presented. When the child touched it, reinforcing consequences followed: stimuli disappeared from the screen, a musical jingle played, and a display of different-colored stars flashed on the screen for 2 s. Simultaneously, six multicolored plastic tokens were delivered into a clear container in the participant's view. Token delivery was accompanied by praise from the experimenter. On subsequent trials, the S– was introduced. If the participant selected the S– stimulus, the stimuli disappeared from the screen and no music, stars, tokens, or praise was delivered. Rather, a 2 s time-out period with a blank screen commenced. Criterion was 100% correct after the introduction of S–. Once this initial discrimination was learned, participants were trained similarly with a second set of stimuli (also not used later in the study). At the end of pretraining and all subsequent sessions, tokens could be exchanged for playtime with a toy of the participant's choosing. Playtime lasted 10 to 30 min, depending on the experimental phase and the participant's performance.

Initially, the child was required to touch the screen with his or her finger to select a stimulus displayed there. Thereafter, she or he was instructed in the use of the computer mouse and the required response was the familiar “point-click” sequence associated with typical mouse operation. The only instructions involved operation of the mouse and token trade-in. The experimenter demonstrated use of the mouse and said, “Hold the button with your finger and push it when the arrow is on the picture.” These instructions were given on the first few trials when only the S+ was presented. Referring to token trade-in, the experimenter said, “See, you can get many stars and tokens when you point to the correct picture. Pay attention and get a lot of stars and tokens. At the end of this game, you can play with a toy; if your cup is full of tokens, you can play with two toys.”

Initial discrimination and reversal training: Each session consisted of 30 trials. Each trial was separated by a 0.5 s intertrial interval. The method used to teach discriminations varied across participants. Four participants (Bia, Eva, Ana, Rai) received initial training on a simple simultaneous discrimination problem with stimulus Set A using a delayed-cue procedure. At the beginning of the first delayed-cue training trial, the S+ and S– (i.e., A1 and A2, respectively) were presented in two of the four corners of the computer screen. After a delay of 1 s, the S– was removed to prompt the child's selection of the remaining S+ stimulus. When he or she did so, reinforcing consequences followed.

After every correct training-trial selection, the interval between next-trial onset and S– removal was increased progressively by 0.5 s, thus giving the child a progressively longer period to make the S+ versus S– discrimination before prompt delivery. If the child selected the S+, reinforcing consequences followed. If the child selected the S–, stimuli disappeared from the screen, a 2 s time-out period commenced, and the delay between trial onset and prompt delivery was decreased by 0.5 s. The delayed cue procedure remained in effect for the entire condition, even if all responses occurred before the S– was removed.

Two participants (Bela and Lia) were not exposed to the delayed-cue procedure. Since the other four participants learned the discrimination quickly, the delayed cue procedure became unnecessarily long. The criterion was 100% correct responding after the cue. However, because the cue was 1 s, often the children were physically unable to respond quickly enough and the response occurred just after the S– removal. We came to believe that five initial trials with just S+ presentation would have the same effect as the delayed cue procedure (i.e., minimizing errors during training), and so we discontinued the procedure with the final two participants.

For these, the first five trials of the Set A discrimination presented the S+ alone. Thereafter, simple differential reinforcement contingencies were in effect. In this procedure, stimuli appeared at the beginning of the trial and remained on the screen until a response was made to one of them. Training continued until children achieved an accuracy criterion of no more than two errors per 30-trial session. Next, they were exposed to a discrimination reversal with Set A, and then to a re-reversal (i.e., return to the baseline) with the same accuracy criteria.

In the next step of training, Set A training was discontinued. Set B stimuli were introduced, and discrimination and reversal training proceeded as described, except that two participants (Ana and Rai) received the simple differential reinforcement procedure instead of the delayed-cue procedure. These procedures were subsequently repeated with Set C. Set B trials were discontinued, and training commenced with Set C using the delayed-cue procedure with Bia and Eva and the simple differential reinforcement method for the remaining children.

Trial-type intermixing: When all three discriminations had been demonstrated in separate sessions, trial types from all three sets were programmed within the same sessions. The procedure was simple differential reinforcement. In the initial intermixing sessions, trials were presented in blocks for each stimulus set: in the first session, there were 10 consecutive trials with each stimulus; in the second session, there were blocks of 5 trials; and in the third session there were blocks of 3 trials. In the fourth session, all trial types were intermixed in an unsystematic order. Such sessions were programmed until the child reached combined accuracy criteria of (a) no more than two errors and (b) no more than one error per stimulus set.

Repeated reversal training and testing: Contingency reversals began in the session following that of intermixing criteria achievement. Table 2 shows the stimulus pairs presented during training. In Reversal 1, stimulus Sets A and B were presented and stimuli that were formerly the S+ were now S- (A1 and B1), whereas stimuli that were formerly the S- were now S+ (A2 and B2). Consequences for correct and incorrect choices were the same as those in training. Repeated sessions of Reversal 1 were conducted until participants achieved an accuracy of no more than one error per session. Then stimuli from Set C (S+: C2; S-: C1) were added, and sessions were conducted until the combined accuracy criteria mentioned above were achieved.

Reversals 2–4 (or 5 for 2 participants, see the footnote to Table 2) were conducted in the same manner, except that the stimulus sets differed, rotating among the possible combinations. During Reversal 2, for example, stimulus Sets B and C were presented during initial reversal training (S+: B1 and C1; S-: B2 and C2); after the participant met the accuracy criterion of no more than one error in a session. Set A was added (S+: A1; S-: A2). In Reversals 5–7 (or 8 for 2 children), all three stimulus sets were presented from the outset of training and testing.

These procedures thus allowed for an assessment of reversal learning set, which was indicated by progressively faster acquisition as the number of discriminative function changes increased and as reversal problems were mastered. It also allowed for an assessment of functional stimulus classes. Strong evidence for functional equivalence would be errorless reversal performance on one or more sets after the participant experienced the reversed contingencies with the other or others. For example, such classes would be indicated in the initial reversal procedures (Reversals 1–4 or 5) if the participant made no incorrect selections on trials that involved the third stimulus set. The strongest possible evidence would be no errors after Trial 1 on the first session of a reversal when discriminative functions of all three stimulus sets were reversed simultaneously (i.e., 96.6% accuracy in a 30-trial session during Reversals 5 [or 6] through 8).

Data were analyzed by means of within-subjects analysis of variance, with the significance level set at .05 for all comparisons. When appropriate, post hoc comparisons were made with the use of the Scheffé test.

Results and Discussion

Participants completed pretraining in an average of 6.8 sessions (range: 3–10).

Training and Intermixing—The children learned the original discriminations in averages of 2.8 sessions (range: 1–4), 1.3 sessions (range: 1–2), and 1.5 sessions (range: 1–2) for Sets A, B, and C stimuli, respectively. The initial reversal phase was learned in fewer sessions on average: 2.3 (range: 1–5), 1.3 (range: 1–2), and 1.3 (range: 1–2) for the Sets A, B, and C stimuli, respectively. In the second initial reversal (return to original contingencies), children met criterion on average in 1.5 (range: 1–3), 1.3 (range: 1–2), and 1.5 (range: 1–3) for the Sets A, B, and C stimuli, respectively. During the intermixing phase (i.e., when all three sets were combined into a single session), participants reached criterion in an average of 5.1 sessions (range: 4–7, with 4 being the minimum given the use of the programmed block-reduction procedure).

Repeated Reversals—The number of sessions in each reversal and overall accuracy on the first session of each reversal are shown in Table 3 for individual participants. For the first four reversals (or first five reversals for Bia and Eva), accuracy reflects performance with two stimulus sets only. All participants took fewer sessions (mean: 2.8; range: 2–4 sessions) to learn their final reversal than to learn their first reversal (mean: 6.8; range: 4–10 sessions), indicating the formation of reversal learning set, $F(5, 25) = 5.49, p < .01$. This interpretation is supported additionally by considering that the first-reversal-to-last-reversal comparison entails comparing a two-stimulus-set problem with a three-stimulus-set problem. That is, the children learned their final three-set problem significantly more rapidly than they learned their initial two-set problem.

Reversal learning set can also be examined on the basis of whether the reversals involved two or three pairs of stimuli. When only two pairs were reversed, accuracy on the first reversal was very high (83.3%; range: 60%–96.6%) and only 1 participant (Bela) achieved an accuracy score below 80%. On the final reversal phase with only two stimulus pairs, accuracy was 91.6% (range: 83.3%–96.6%). There were no statistically significant differences between accuracy on first and last reversal involving two stimulus pairs, $F(2, 5) = 5.62, p > .06$.

Starting with reversal 5 (or reversal 6 for Bia and Eva), all three stimulus sets were presented in an intermixed fashion from the start of the reversal phase. In the first reversal session, accuracy was 86.7% (range: 76.6%–93.3%), showing a slight decrement from the previous reversal (which had only two stimulus pairs). Accuracy on the final reversal was 91.6% (range: 83.3%–96.6%) and did not differ statistically from accuracy on the first session, $F(2, 5) = 4.84, p > .07$. Notably, there tended to be savings across both the two-set problems and the three-set problems. That these savings escaped statistical significance may be attributed to the small numbers of each type of problem. The obtained across-type data of statistical significance (i.e., acquisition comparing Reversal 2 with 7 or 8) thus appears to be the more important finding.

The strongest evidence for functional equivalence is indicated by accuracy after Trial 1 in a reversal. In other words, will the children respond on trials 2–30 as they had been trained, or will they respond in accordance with the newly reversed contingencies (i.e., one error in a session, or 96.6% correct)? Table 3 indicates four instances of such accuracy with two stimulus sets (Reversal 3 for Bia and Reversals 1, 2, and 4 for Rai) and three instances with three stimulus sets (Reversal 7 for Bia and Reversals 7 and 8 for Lia).

Functional equivalence also can be evaluated for sessions in which two stimulus sets were initially reversed (and trained, if necessary) before the third set was added. One way to do so is to examine Trial 1 accuracy for the third stimulus set when it was added to the two already reversed sets in Reversals 1–4 (or 1–5 for Bia and Eva). Class formation here is indicated by little or no disruption in accuracy for the already-reversed sets (Sets 1 and 2 combined) and no errors on Trial 1 with the third set. As Table 4 indicates, seven data points (across 4 of the 6 participants) indicate perfect accuracy with Set 3—the strongest evidence for functional stimulus class formation. In addition, Reversal 3 for Bia shows an accuracy of 90%. This participant responded correctly on Trial 1 with Set 3 but then made a single error on this discrimination later in the session (the fourth time that trial type was encountered). Although participants made Trial 1 errors with Set 3 on the remaining reversals, the table also shows high overall session accuracy (only two instances below 80%).

Finally, class formation can be assessed by examining data over the first five trials for each pair during the first session in which the third pair was added (see Table 5). Again, we expect little or no disruption on the already reversed sets (90%, or 9 of 10 trials, correct with Sets 1 and 2) and one or fewer errors with Set 3 (80%, or 4 of 5 trials, correct). These data too show strong evidence of functional equivalence: there were 12 instances across 5 of the participants (indicated in bold in Table 5) in which no more than one error was made with the added pair and no more than one error was made with the other two pairs combined. In addition, there are five instances in which no more than one error was made with the added pair and two errors were made with the other pairs combined.

The high accuracy levels overall during the reversals are supplemental supporting evidence for functional stimulus classes, a point that will be developed more fully in the discussion of the results of the next experiment.

Experiment 2

In this experiment, we were interested in assessing the ability of nonverbal children with autism to exhibit reversal learning set and functional equivalence. As noted earlier, there is virtually no information available, to our knowledge, about either ability in this population. Regarding more capable persons with neurodevelopmental disabilities, the available data suggest that they tend to perform more poorly on reversal tasks than typically developing control subjects (Heal et al., 1966; Plenderleith, 1956; see Kaufman & Prehm, 1966). In one study (Prior & Chen, 1977), however, learning set ability was compared among matched verbal children with autism, Down syndrome, and typical development, and the first group showed statistical evidence of superior learning.

As noted earlier, this experiment was already under way when Experiment 1 commenced. The training and reversal tests were procedurally similar, except that all of the reversal procedures were similar to those of the later stages of Experiment 1.

Method

Participants—Six students at a behaviorally oriented New England school for children with neurodevelopmental disabilities participated in the experiment. Chronological age, clinical diagnosis, and language age equivalence (as assessed with the Peabody Picture Vocabulary Test-III, Dunn & Dunn, 1997) are listed in Table 6. All were nonverbal, and all had preparatory discrimination learning histories that have been extensively described elsewhere (Lionello-DeNolf, Barros, & McIlvane, in press). In brief, this history involved training the students to open and close apparatus doors to obtain food items presented in compartments. Training involved nonverbal prompts (e.g., slowly nudging the door closed). Once compartment training was completed, the participants were taught a simple discrimination that involved a delayed S

+ procedure (see preliminary training below). This initial training involved the Set 1 stimuli reported in this experiment for each participant.

Setting, Apparatus, and Stimuli—All sessions took place in an automated teaching laboratory. The laboratory room was split into two enclosures by the apparatus. On one side, researchers operated controls that programmed experimental events; on the other side, participants interacted with the teaching equipment. Details of the apparatus have been described extensively elsewhere (Lionello-DeNolf & McIlvane, 2003), and only aspects relevant to the current study will be described here (see Figure 2).

The participant was seated before a tripartite stimulus panel (the teaching area). Directly in front of the seated participant were two stimulus display compartments with clear, sliding doors that could be locked or unlocked by the researcher at any time. Centered beneath those compartments was a third display compartment that was not used in this experiment. The compartments could be lit with white light and contained movable platforms so that stimuli could be rotated into and out of view.

There were four additional, doorless compartments, two on each side of the third display compartment. The exterior two compartments were used to deliver tokens, which could be inserted into a slot, positioned 8 cm to the left and 5 cm above the left compartment, to obtain food reinforcers. The interior two compartments were used to deliver food reinforcers.

In the programming (i.e., researchers') area, two researchers conducted the session. Experimental events were controlled with a Macintosh G4 computer running LabView (National Instruments) software. One researcher entered commands into the controlling computer and recorded the participants' responses. (Unlike Experiment 1, this experiment involved a computer that was unable to automatically record responses). The other researcher loaded and unloaded stimuli into the upper display compartments and delivered reinforcers into the food wells. The participant could not view any of these programming operations, because the teaching area was fully enclosed by laminated wall units. The experimenters could observe the participants' behavior at all times via a four-camera closed-circuit television-and-videotaping system. Videotape records were used as an additional check on the experimenter-recorded accuracy scores.

The stimuli for each participant consisted of different-colored food items or different-colored poker chips. Participants were given a food preference assessment to determine a highly preferred food. Foods for the assessments were based on recommendations from the participants' teachers. To be considered an appropriate edible stimulus for this experiment, the food item had to come in a variety of colors, such as M&Ms or Skittles. For some participants, food preferences changed on a daily basis or were inappropriate choices for the experiment (e.g., Cheetos cheese puffs being available in only one color). In such cases, different-colored tokens were used instead. Table 7 lists the stimuli used for each participant.

For some participants, stimuli to be discriminated were switched from edible stimuli to tokens after training had commenced. For example, this switch occurred during preliminary training for Betty, because she began to refuse to consume the previously preferred Skittles. Two other participants (Nancy and Sal) were switched later in training when they began to throw the edible stimuli on the floor. When a switch was made to tokens, these stimuli could be exchanged for a variety of preferred food items.

Procedure

Preliminary training: Experimental sessions lasted 15 to 30 minutes and were conducted two to three times per week. In the participants' first session, they underwent the preference

assessments. In the following session, the children were given the Peabody Picture Vocabulary Test-III. During the subsequent several sessions, the participants were accustomed to the automated laboratory and taught how to operate the compartment doors. Finally, the participants were given preliminary training designed to teach them to refrain from responding to food items displayed in the compartments unless one or more of them were illuminated with white light. This training involved use of a delayed S+ procedure (McIlvane, Kledaras, Callahan, & Dube, 2002) and was designed to discourage indiscriminate responding to displayed food items.

Initial discrimination and reversal training: When delayed S+ training was completed, participants were taught simple discriminations involving three sets of food items. Procedures were similar to those used in Experiment 1, except that the initial discrimination training was accomplished with a delayed cue procedure for all three stimulus sets for all participants. Each discrimination training session consisted of 30 trials in which the left-right position of the S+ was balanced and varied unsystematically. As in Experiment 1, discrimination training with each of the three stimulus sets was accomplished separately.

At the beginning of a delayed cue trial, the S+ and S− stimuli were rotated into the participant's view, both lights were turned on, and the door to the S+ compartment was unlocked. The door to the S− was always locked. Participants could hear (but not see) the operation of the locking mechanism; however, the locks were immediately adjacent to one another in a central location, and it was thus difficult to discriminate which lock was operating on a given trial, particularly as it was masked by the sound of the rotation of compartments. During early delayed cue trials, the light in the S− compartment was then turned off after a brief delay (500 ms), a value selected to be less than the participants' typical response latency. The remaining compartment light cued the participant to open its door to retrieve the S+ food item or token. When food items were to be discriminated, the participant could then consume the S+ item immediately. When token stimuli were to be discriminated, the participant could deposit the token in the slot, and an edible reinforcer was then delivered into the lit food well. The intertrial interval (10–30 s) commenced when the participant closed the compartment door. Selections of the dark compartment ended the trial without food or token deliveries.

Over trials, the time between trial onset and cue delivery was gradually increased (in 250 ms increments) to 2,000 ms. At that point, use of the cue was discontinued (i.e., the lights in both the S+ and S− compartments remained lit throughout the trial). A correct response was touching and opening the door containing the S+, and an incorrect response was touching the door containing the S−. Participants remained in initial discrimination training until they reached or exceeded a 90% accuracy for two consecutive sessions in which the delayed cue procedure was not used.

Once the aforementioned criteria were met, the reinforcement contingencies for the discrimination were reversed. The stimulus previously designated S+ was now S−, and the stimulus previously designated S− was now S+. The delayed cue procedure was not used in this phase. Participants remained in this initial reversal phase until the 90% accuracy criterion was met. Then the reinforcement contingencies were reversed again (i.e., the original reinforcement contingencies were again in effect, an A-B-A design).

Trial-type intermixing: In this phase, discrimination training was conducted with all three stimulus sets in the same session. As in Experiment 1, sets were initially presented in blocks of 10 trials. The number of trials per block was then reduced to five and three. Finally trials were presented in an intermixed fashion. The order of stimulus presentation varied with each session according to three prearranged sequences. Trial-type intermixing training continued

until participants met a combined accuracy criterion of (a) $\geq 90\%$ in a 30-trial session and (b) no more than one error per stimulus set.

Repeated reversals: Reversal procedures were similar to those of Experiment 1. For each reversal, (a) the order of stimulus presentation varied according to three prearranged sequences of trials, with intermixing of all three trial types, and (b) contingencies remained in place until participants made no more than three errors in a single session and no more than one error within a stimulus set. Participants were exposed to a varying number of reversals, depending principally on their availability for study.

Results and Discussion

Agreement on interobserver reliability on experimenter-recorded data, calculated from videotape records of 40% of sessions, was virtually 100%.

Pretraining, Initial Discrimination, Reversal Training, and Trail-Type Intermixing

—Participants completed preliminary training in an average of 16.7 sessions (range: 6–30 sessions). Table 8 shows the number of sessions to attain criterion performances on the initial training, reversal, and return to initial contingencies for each participant when training was conducted separately with each stimulus set. Also shown is the number of sessions to criterion during trial-type intermixing and terminal accuracy (on the final session) for each of the training phases.

Initial training with each stimulus set took 14.8, 5.2, and 3.5 sessions, on average, for Sets 1, 2, and 3, respectively. The difference in sessions to criterion was statistically significant, $F(2, 5) = 4.61$, $p < .04$, and post hoc contrasts indicate that it took longer to learn Set 1 than Sets 2 and 3, $p < .04$ and $.08$, respectively. Table 8 shows also that most participants learned the reversal phase in fewer sessions than the Initial discrimination: in 13 of 18 reversal opportunities, the reversal was learned more quickly than the initial discrimination. On average, it took 11.7, 2.7, and 2.5 sessions to learn the reversal for Sets 1, 2, and 3, respectively. The difference in reversal sessions to criterion was statistically significant, $F(2, 5) = 4.09$, $p < .05$, and post hoc contrasts indicate it took longer to learn the Set 1 reversal than the Set 3 reversal, $p < .04$. Finally, when researchers returned to the original contingencies there were no significant differences in the number of sessions needed to re-achieve the criterion accuracy: 6, 3.7, and 3 sessions for Sets 1, 2, and 3, respectively, $F(2, 5) = 0.13$, $p > .88$. Overall, these data seem noteworthy in that our nonverbal cohort showed good evidence of “learning to learn” across problems when they were presented in separate sessions.

When trial types were intermixed, there was substantial disruption of previously mastered discriminations and substantial variability across participants. On average, the number of sessions required to meet the intermixing criteria were 19.2, but the range was large (7–43 sessions). Three participants met criterion in 10 or fewer sessions, while the remainder needed many more. These data seem particularly noteworthy in that participants had previously performed each of these discriminations separately at very high levels of accuracy. For these children, unlike the children of Experiment 1, intermixing of problems thus posed a substantial acquisition challenge.

Repeated Reversals—Table 9 shows, for each reversal and for each participant, the first session accuracy score and the number of sessions required to reach criterion. Of the overall results, there appear to be three cases in which reversal learning set was clearly demonstrated (Andy, Joe, and Sal), two cases in which there was no suggestion of reversal learning set (Betty and Nancy), two cases in which functional equivalence was clearly demonstrated (Joe and Sal), and two anomalous but nevertheless interesting findings (Andy and Kerry).

Reversal learning set is indicated by fewer sessions to criterion with each successive reversal. On average, participants needed 10.7 sessions (range: 6–14) to learn the first reversal and 11.7 sessions (range: 1–47) to learn the last, and this difference was not significant, $F(1, 5) = 0.01$, $p > .05$. However, an examination of the individual participant data show that for four of the six participants, the final reversal was learned in fewer sessions than the first. In two of these cases, criterion was met in a single session. Moreover, in one case, Kerry, the final reversal took 47 sessions (see below for a further discussion of this participant). When his data are excluded from the analysis, the final reversal was completed in an average of 4 sessions (range: 1–12). When compared with sessions to criterion with the first reversal, the difference just escaped statistical significance, $F(1, 4) = 6.9$, $p > .06$.

Individual participant data give a clearer picture of reversal learning set formation. For three participants, Andy, Joe, and Sal, reversal learning set is suggested by a decreasing trend in the number of sessions needed to achieve criterion performance (see Table 9). Data from Betty and Nancy do not indicate reversal learning set: the number of sessions required to learn each reversal did not decrease with repeated reversal experience.

Functional equivalence can be conservatively defined as no errors after the first trial in an initial reversal session (96.6% in a session). Table 9 indicates three instances across two participants (Joe and Sal) where such equivalence was achieved. If the criteria to demonstrate functional equivalence are relaxed slightly to no more than three errors in an initial reversal session (90% correct), there are eight instances: Andy (R5), Joe (Rs 12, 13, and 14), Kerry (R2) and Sal (Rs 4, 5, and 6).

An interesting anomaly was performance by Kerry, who completed only three reversals. Overall accuracy on the first reversal was 60% correct. For the second reversal, accuracy was 90% or better for all three stimulus sets, data suggesting both learning set and contingency classes. In the first session of the third reversal, Kerry scored 83.3%, substantially above the chance levels of 50%, and he scored 100% on one of the three sets. Thereafter, however, his performance deteriorated, with reasonably high scores intermixed with low scores. Various stimulus control shaping methods were used in an unsuccessful attempt to assist him in meeting the Reversal 3 criterion. Kerry's participation was discontinued after 47 sessions, because it appeared that these discriminations had become permanently faulty, perhaps a result of the development of stable error patterns (cf. Terrace, 1963; Stoddard & Sidman, 1967).

Also of notable interest was Andy's performance. He clearly showed reversal learning set over the course of the seven reversals that he experienced, and his first session accuracy score improved progressively across reversals. Intriguingly, he achieved a criterion score for contingency classes in his fifth reversal but did not maintain that accuracy level in the subsequent reversal. It is unfortunate that his participation ended after the first session of the seventh reversal (due to a transfer to a different school). Had he remained available for continued study, perhaps reliable contingency classes would have emerged later, as they did for Joe.

A final related point of interest in our data is a possible relationship between the development of learning set and the development of functional equivalence. Both of the children who showed reliable functional stimulus classes also showed the learning set outcome, and, as just noted, Andy's learning set development was accompanied by one test score that was consistent with functional equivalence. Perhaps these relationships are not coincidental. On the one hand and as noted earlier, contingency classification is a logical endpoint of a process of progressively more efficient learning. On the other hand, learning set itself has never been explained at the level of a basic behavioral process or processes. (See Catania, 2007, for a discussion of stimulus

equivalence in relationship to such processes.) Perhaps there exists a more intimate relationship between these two behavioral phenomena than has been suspected heretofore.

General Discussion

Six typically developing children and six children with autism were trained on a series of separate simple discrimination and reversal tasks. Then all discriminations were trained within the same session, and reversals were again given. Accuracy on the tasks was analyzed in terms of development of learning set and functional equivalence class formation.

Regarding learning set, the results systematically replicate prior studies with typically and atypically developing children and extend them to reversal learning set. The rapidity and efficiency of reversal-set learning in the typically developing children was somewhat surprising in that (a) the cohort was younger than has been studied in typical learning set research, (b) a two-dimensional rather than three-dimensional stimulus set was used, and (c) the response mode (mouse “pointing and clicking”) involved displaced responding. We think it unlikely that our findings with our typically developing children would have been predictable from the developmental psychology of several decades ago, when learning set research was last in vogue (Stevenson, 1973). It may be that children developing in the information age—even very young children—come to studies with behavioral repertoires that prepare them especially well for interacting with procedures such as those that we employed.

Regarding our nonverbal children with autism, there was also substantial evidence of “learning to learn,” although the results were different and more variable than those that we obtained from the typically developing children. On average, the efficiency of both groups of children’s discrimination learning tended to improve during the course of their initial training. One difference between the two groups of children, however, was the effect of intermixing discriminations with the three stimulus sets. Whereas the greatest number of sessions required to achieve criterion in a typically developing child was seven, that number reflected the fewest number of sessions needed by the children with autism. Another difference was that all of the typically developing children ultimately came to exhibit reasonably efficient reversal learning by the end of their participation, whereas only three of the children with autism did so. Finally, no typically developing child exhibited the protracted failure that was observed with Kerry.

We cannot conclude on the basis of the data in hand that the observed group differences were due to the autism or mental retardation, or both, of one group. The typically developing children did not constitute a matched control group. These children were verbal and exhibited no evidence of behavioral retardation. A true developmental control group would have been composed of toddler-aged or younger children such that one could compare directly the groups on variables such as language-age equivalent score. Regrettably, we are not aware of any published data available on learning-set development in typically developing children with developmental levels comparable to those of our nonverbal children. Recent research suggests, however, that such research could be accomplished (Gil & Oliveira, under review), and such studies are clearly needed to make a more adequate comparison with our second experiment.

Regarding our other major interest, functional equivalence class formation, these studies were successful along two important dimensions. First, Experiment 1 contributes the first published data on this topic in young, typically developing children. Although not all functional equivalence test outcomes were perfectly consistent, many were. The generally high scores overall on class tests with these children raise an additional question: whether such test procedures inherently set up a perhaps unrealistic standard for fully convincing data (i.e., perfect test performance). Indeed, any performance short of perfection may be attributed to very rapid discrimination learning, particularly when participants also exhibit learning set. It would be helpful if we could develop modifications of, or alternatives to, the repeated reversal

procedure for the assessment of functional class development. For example, the initial training and class test procedures used with the typically developing children were a response to certain results with some of the children with autism in which test data were suggestive of class formation but inconclusive. We reasoned that fully completing reversal training with the first two class members might make it more likely to achieve a perfect outcome with the third, but that procedure was only partly successful in accomplishing our aims. Another approach could be use of a group design and test methods modeled after the consistent-inconsistent contingency methods developed by Urcuioli and colleagues (Lionello-DeNolf & Urcuioli, 2002; Urcuioli & Zentall, 1993; Urcuioli, Zentall, Jackson-Smith, & Steirn, 1989).

Our second success in this work was providing convincing data of functional equivalence class formation in nonverbal children with autism, albeit limited to a few individuals. Indeed, it is hard to account for the performances of Joe and Sal without conceding that functional equivalence was evident, and the initial data from Kerry were also highly suggestive. These data, of course, are not consistent with the position that equivalence classes depend on overt or covert naming (Horne & Lowe, 1996). They are consistent, however, with other theoretical analyses, particularly relational frame theory (Hayes, 1991), which allows analogy between the development of operant abstraction via multiple exemplar training and the development of equivalence relations. Data from Joe seem to fit particularly this type of account. Sal, by contrast, showed rapid development of functional stimulus classes: findings of the type that seem more in line with tenets of stimulus control topography coherence theory (McIlvane & Dube, 2003).

The limited success that we achieved with the nonverbal cohort—demonstrating learning set and functional equivalence in a subset of the children with autism—seems all the more impressive in that it was achieved via the repeated contingency reversal method using discriminative stimuli that were highly similar in appearance. On its face, one could scarcely design a procedure that was a poorer match for our nonverbal population, which is characterized by unusually restricted attending (Dickson, Deutsch, Wang, & Dube, 2006) and unusually perseverative and stereotypical responding. Indeed, discrimination reversal tasks have been used traditionally to assess perseverative responding and behavioral inflexibility in clinical populations such as this. For this population especially, it seems reasonable to predict that training and test methodology that minimizes or avoids reversal methods might yield superior evidence of functional class formation or reduce the level of intra- and interparticipant variability that we observed in this first experiment.

We believe it important that in future work with nonverbal children, the various aspects of the training and test procedures be rendered as highly discriminable as possible (cf., McIlvane & Dube, 2003). For example, it would be desirable to retain three-dimensionality of discriminative stimuli for this population, but it may not be necessary to use food or token reinforcers as the stimuli to be discriminated, thus avoiding potential problems of reinforcer preference. Another possible improvement in the procedures would be use of outcome-specific consequences to define the contingency classes (cf., Dube, McIlvane, Mackay, & Stoddard 1987; Dube, McIlvane, Maguire, Mackay, & Stoddard, 1989). That procedure was shown to be helpful by Kastak and colleagues (2001) in their work with sea lions; their subjects were able to demonstrate functional equivalence only when outcome-specific reinforcement was used.

To conclude, we think that our findings encourage further research to assess the variables that may influence the development of learning set and functional equivalence in children who exhibit both typical and atypical development. For both populations, an important goal will be development of methodology to render such studies less time- and labor-intensive. Few laboratories have the resources to devote dozens or hundreds of training sessions to the

problems, perhaps one reason why the extant data sets are so meager with both human and nonhuman populations. We do not believe that phenomena such as learning set and functional stimulus classes develop only over extended time frames, particularly when the discriminations required seem straightforward, as in situations that we studied in this work. While painstaking procedures may be necessary at first to define critical variables that determine success or failure of training and test procedures, an important subsequent goal is procedural development and refinement such that processes of interest may be controlled and studied in an optimally efficient manner.

Acknowledgments

This research was supported by FAPESP/PRONEX Grant 2003/09928-4 and FAPESP Grant 06/55848-0 (Experiment 1), by CAPES BEX0640/03-4 and CNPq 201155/2004 (Experiment 2), and by NICHD grants HD 39816 and HD 04147 (Experiment 2).

We thank Lidia Postalli for her assistance in Experiment 1 and Camila Domeniconi, Sarah Luthern, Jeff Kilpatrick, and Kerrilyn Lacerte for their assistance in data collection for Experiment 2. Manuscript preparation was supported by HD 04666.

References

- Blank M. The effects of training and verbalization on reversal and extra-dimensional learning. *Journal of Experimental Child Psychology* 1966;4:50–57. [PubMed: 5971017]
- Carr D, Wilkinson KM, Blackman D, McIlvane WJ. Equivalence classes in individuals with minimal verbal repertoires. *Journal of the Experimental Analysis of Behavior* 2000;74:101–114. [PubMed: 10966098]
- Catania, AC. *Learning*. 4th ed.. Cornwall-on-Hudson, NY: Sloan Publishing; 2007. The Cambridge Center-Sloan Century Series in Behavior Analysis.
- Dickson CA, Deutsch CK, Wang SS, Dube WV. Matching-to-sample assessment of stimulus overselectivity in students with intellectual disabilities. *American Journal on Mental Retardation* 2006;111:447–453. [PubMed: 17029502]
- Dube WV. Computer software for stimulus control research with Macintosh computers. *Experimental Analysis of Human Behavior Bulletin* 1991;9:28–30.
- Dube, WV.; Hiris, EJ. MTS v 11.6 [Software for research and education]. Waltham, MA: E. K. Shriver Center for Mental Retardation; 1996.
- Dube WV, McIlvane WJ, Mackay HA, Stoddard LT. Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior* 1987;47:159–175. [PubMed: 3572297]
- Dube WV, McIlvane WJ, Maguire RA, Mackay HA, Stoddard LT. Stimulus class formation and stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior* 1989;51:65–76. [PubMed: 2921589]
- Dunn, LM.; Dunn, LM. *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service; 1981.
- Dunn, LM.; Dunn, LM. *Peabody Picture Vocabulary Test-III*. Circle Pines, MN: American Guidance Service; 1997.
- Galvão OF, Barros RS, Brandão S, Lavratti CM, Santos JR, Brino ALF, Dube WV, McIlvane WJ. Extent and limits of the matching concept in *Cebus apella*: A matter of experimental control? *The Psychological Record* 2005;55:219–232.
- Gil MSCA, Oliveira TP. Identity matching-to-sample in the study of acquisition of conditional discriminations by infants. 2007 Manuscript submitted for publication.
- Hayes, SC. A relational control theory of stimulus equivalence. In: Hayes, LJ.; Chase, PN., editors. *Dialogues on verbal behavior*. Reno, NV: Context Press; 1991.
- Heal LW, Ross LE, Sanders B. Reversal and partial reversal in mental defectives and normal children of a comparable mental age. *American Journal of Mental Deficiency* 1966;71:411–416.

- Horne PJ, Lowe CF. On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior* 1996;65:185–241. [PubMed: 16812780]
- House B. The effect of distinctive responses on discrimination reversals in retardates. *American Journal of Mental Deficiency* 1964;69:79–85. [PubMed: 14173763]
- Kastak CR, Schusterman RJ, Kastak D. Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior* 2001;76:131–158. [PubMed: 11599636]
- Kaufman ME, Gardner WI. Transfer of training of learning sets in mental defectives: I. Discrimination reversal. *American Journal of Mental Deficiency* 1969;73:801–803. [PubMed: 5780067]
- Kaufman ME, Prehm HJ. A review of research on learning sets and transfer of training in mental defectives. *International Review of Research in Mental Retardation* 1966;2:123–149.
- Lea, SEG. In what sense do pigeons learn concepts?. In: Roitblat, HL.; Bever, TG.; Terrace, HS., editors. *Animal cognition*. Hillsdale, NJ: Erlbaum; 1984. p. 263–276.
- Lionello-DeNolf KM, Barros RS, McIlvane WJ. A novel method for teaching the first instances of simple discrimination to nonverbal autistic children in a laboratory environment. *The Psychological Record*. (in press).
- Lionello-DeNolf KM, McIlvane WJ. Rebirth of the Shriver automated teaching laboratory. *Experimental Analysis of Human Behavior Bulletin* 2003;21:12–17.
- Lionello-DeNolf KM, Urcuioli PJ. Stimulus control topographies and test symmetry in pigeons. *Journal of the Experimental Analysis of Behavior* 2002;78:476–495.
- McIlvane WJ, Dube WV. Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst* 2003;26:195–213.
- McIlvane WJ, Dube WV, Kledaras JB, Iennaco FM, Stoddard LT. Teaching relational discrimination to individuals with mental retardation: Some problems and possible solutions. *American Journal on Mental Retardation* 1990;95:283–296. [PubMed: 2261161]
- McIlvane WJ, Kledaras JB, Callahan TC, Dube WV. High-probability stimulus control topographies with delayed S+ onset in a simultaneous discrimination procedure. *Journal of the Experimental Analysis of Behavior* 2002;77:189–198. [PubMed: 11936251]
- Plenderleith M. Discrimination learning and discrimination reversal learning in normal and feeble-minded children. *The Journal of Genetic Psychology* 1956;88:107–112. [PubMed: 13319675]
- Prior MR, Chen CS. Learning set acquisition in autistic children. *Journal of Abnormal Psychology* 1975;84:701–708. [PubMed: 1194533]
- Sidman M, Tailby W. Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior* 1982;37:5–22. [PubMed: 7057129]
- Sidman M, Wynne CK, Maguire RW, Barnes T. Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior* 1989;52:261–274. [PubMed: 16812597]
- Stevenson, HW. *Children's learning*. New York: Appleton-Century-Crofts; 1973.
- Stoddard LT, Sidman M. The effects of errors on children's performance on a circle-ellipse discrimination. *Journal of the Experimental Analysis of Behavior* 1967;10:261–270. [PubMed: 6056799]
- Terrace HS. Discrimination learning with and without "errors". *Journal of the Experimental Analysis of Behavior* 1963;6:1–27. [PubMed: 13980667]
- Urcuioli PJ, Zentall TR. A test of comparison-stimulus substitutability following one-to-many matching by pigeons. *The Psychological Record* 1993;43:745–759.
- Urcuioli PJ, Zentall TR, Jackson-smith P, Steirn JN. Evidence for common coding in many-to-one matching: retention, intertribal interference, and transfer. *Journal of Experimental Psychology: Animal Behavior Processes* 1989;15:264–273.
- Vaughan W Jr. Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes* 1988;14:36–42.

Figure 1.

The stimuli used in Experiment 1. S+ and S- indicate the function of each stimulus in the first discrimination (prior to reversal training).

Figure 2. Portions of the automated teaching lab used in Experiment 2. The top panel shows the stimulus display in the teaching area, and the bottom panel shows the same view from the experimenter area.

Table 1

Age, Gender, and Language-Age Equivalent Score for Each Participant in Experiment 1

Child's Name	Age (years-months)	Gender	Language Age (years-months)
Bia	2-11	Female	—
Bela	3-01	Female	<1-11
Eva	3-11	Female	2-00
Lia	4-01	Female	2-11
Ana	4-02	Female	2-03
Rai	4-03	Male	3-04

Note. A language-age equivalent score was not available for Bia. Language-age scores for Ana and Rai were obtained prior to the study, at ages 3 years 5 months and 3 years 6 months, respectively.

Table 2

Order of Reversal Sessions and Stimulus Pairs Presented in Experiment 1

Phase	Stimulus Pairs and Trios (Training)		
R1	A	B	–
R2	A	–	C
R3	–	B	C
R4	A	B	–
R5	A	B	C
R6	A	B	C
R7	A	B	C
R8	A	B	C

Note. For Bia, R4 Involved initial presentations of A and C for a phase identical to that of R2. R5 involved presentations as depicted in R4 in the table. For Eva, R5 involved initial presentations of B and C for a phase identical to R3. All remaining phases are as depicted in the table. For reversals in which two stimulus pairs were initially presented, the third pair was presented once criterion was achieved with the other two pairs.

Table 3
Number of Sessions and First Session Accuracy (%) for Each Reversal Phase in Experiment 1

	Bela	Bia	Eva	Ana	Lia	Rai
R1	6 (60)	7 (80)	9 (86.6)	10 (90)	5 (86.6)	4 (96.6)
R2	7 (90)	3 (93.3)	4 (93.3)	4 (90)	4 (93.3)	4 (96.6)
R3	5 (76.6)	9 (100)	5 (93.3)	4 (93.3)	6 (93.3)	5 (90)
R4	7 (83.3)	7 (86.6)	7 (90)	5 (93.3)	6 (93.3)	5 (96.6)
R5	5 (76.6)	5 (93.3)	6 (90)	7 (90)	3 (80)	3 (93.3)
R6	4 (83.3)	3 (90)	3 (90)	6 (90)	2 (93.3)	2 (90)
R7		2 (96.6)		3 (90)	2 (96.6)	2 (86.6)
R8					2 (96.6)	3 (93.3)

Note. Data in **bold** type indicate sessions during which three stimulus pairs were reversed. Data in regular type indicate sessions during which only two pairs were reversed.

Table 4
Accuracy During Reversal Sessions 1–5 When the Third Stimulus Pair Was Added to the Session in Experiment 1

	Bela		Bia		Eva		Ana		Lia		Rai	
	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set
R1	100	80	100	90	100	90	95	90	95	80	95	100
R2	90	100	100	100	100	90	100	100	95	80	95	70
R3	100	80	80	90	100	60	90	100	95	80	100	90
R4	90	90	95	100	55	90	100	90	100	90	100	80
R5			100	100	90	90						

Note. 1st and 2nd set refers to average performance on trials with the two pairs of stimuli that were initially reversed in the phase, on the first session when the third set was added. 3rd set refers to the final stimulus pair to be reversed, after criterion had been achieved with the other pairs. Data in **bold** indicate accuracy meeting criterion for functional equivalence: no errors made on Trial 1 with Set 3.

Table 5

Accuracy Over the First Five Trials During Reversal Sessions 1–5 When the Third Stimulus Pair Was Added to the Session in Experiment 1

Bela		Bia		Eva		Ana		Lia		Rai	
1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set	1st and 2nd Set	3rd Set
R1	100	60	100	80	100	90	80	90	60	90	100
R2	80	100	100	100	80	100	100	90	60	90	60
R3	100	60	80	80	100	80	100	90	60	100	80
R4	80	80	90	100	50	100	80	100	80	100	60
R5			100	100	80		80				

Note. 1st and 2nd set refers to average performance on trials with the two pairs of stimuli that were initially reversed in the phase, on the first session when the third set was added. 3rd set refers to the final stimulus pair to be reversed, after criterion had been achieved with the other pairs. Data in bold indicate 1 or fewer errors with the added pair combined with high accuracy (1 or fewer errors) with the other two pairs.

Table 6

Age, Gender, Diagnosis, and Language-Age Equivalent Score for Each Participant in Experiment 2

Child's Name	Age (years-months)	Gender	Diagnosis	Language Age (years-months)
Andy	18-07	Male	Autism	1-09
Betty	12-03	Female	Autism/PDD	<1-09
Joe	11-02	Male	Autism	1-09
Kerry	13-06	Male	Autism	2-08
Nancy	10-09	Female	Autism	1-09
Sal	6-08	Male	Autism	1-09

Note. PDD = pervasive developmental disorder.

Table 7

Stimulus Type and Color for Each Participant in Experiment 2

	Stimulus Type	Color
Andy	Set 1: chocolate-covered pretzels	Set 1: brown and white
	Set 2: Doritos (tortilla chips)	Set 2: green and orange
	Set 3: Jelly Bellies (jelly beans)	Set 3: blue and yellow
Betty	Skittles (candy)/poker chips	Set 1: green and red
		Set 2: purple and white
		Set 3: pink and black
Joe	Poker chips	Set 1: blue and orange
		Set 2: brown and yellow
		Set 3: red and green
Kerry	M&Ms (candy)	Set 1: orange and brown
		Set 2: red and green
		Set 3: purple and yellow
Nancy	Mint Skittles and Sweets Tarts (Candy) poker chips	Set 1: white and green
		Set 2: purple and yellow
		Set 3: red and orange
Sal	Skittles/poker chips	Set 1: red and green
		Set 2: purple and orange
		Set 3: black and yellow

Note. The color listed first was the stimulus designated as the S+, and the color listed second was the stimulus designated as the S– during initial training.

Table 8

Number of Sessions and Terminal Accuracy (%) for Initial Discrimination, Reversal, and Trial-Type Intermixture Training in Experiment 2

Participant	Stimuli	Initial	Reversal	Return to Initial	Intermixed
Andy	Set 1	17 (100)	5 (100)	7 (100)	
	Set 2	7 (96.7)	4 (96.6)	2 (93.6)	
	Set 3	3 (100)	2 (100)	1 (90.2)	
	All				28 (96.7)
Betty	Set 1	26 (96.2)	12 (100)	*	
	Set 2	9 (96.7)	2 (96.7)	11 (95)	
	Set 3	2 (100)	2 (96.7)	1 (96.7)	
	All				19 (96.7)
Joe	Set 1	3 (100)	32 (100)	10 (95.8)	
	Set 2	4 (100)	3 (100)	2 (93.3)	
	Set 3	10 (100)	6 (100)	5 (96)	
	All				10 (100)
Kerry	Set 1	11 (100)	2 (93.3)	3 (100)	
	Set 2	2 (96.7)	3 (96.7)	2 (93.3)	
	Set 3	1 (93.3)	2 (93.3)	1 (96.7)	
	All				7 (100)
Nancy	Set 1	29 (95.5)	14 (90)	*	
	Set 2	4 (100)	2 (93.3)	2 (100)	
	Set 3	3 (100)	2 (93.3)	3 (100)	
	All				8 (96.7)
Sal	Set 1	3 (100)	5 (96.7)	4 (100)	
	Set 2	5 (100)	2 (100)	3 (90)	
	Set 3	2 (100)	5 (100)	7 (90)	
	All				43 (96.7)

* Return to initial phase was inadvertently omitted.

Table 9
Number of Sessions and First Session Accuracy (%) for Each Reversal Phase in Experiment 2

	Andy	Betty	Joe	Kerry	Nancy	Sal
R1	12 (56.7)	9 (83.3)	12 (38.7)	14 (50)	11 (13.3)	6 (56.7)
R2	8 (63.3)	7 (58.3)	14 (27.3)	1 (90)	9 (66.7)	4 (80)
R3	4 (70)	11 (75.3)	21 (11)	47 (83.3)	3 (13.3)	2 (83.3)
R4	3 (56.7)	5 (60)	5 (63.3)		4 (60.00)	1 (93.3)
R5	2 (93.3)	8 (76.7)	3 (83.3)		5 (73.3)	1 (96.7)
R6	3 (76.7)	10 (39)	4 (73.3)		6 (63.3)	1 (100)
R7	1 (76.7)	6 (67)	2 (86.7)		12 (23.3)	
R8		8 (59.3)	5 (70)			
R9		5 (70)	5 (76.7)			
R10			3 (60)			
R11			2 (70)			
R12			1 (90)			
R13			1 (95.7)			
R14			1 (93.3)			