ACTA DE INVESTIGACIÓN PSICOLÓGICA, 2011, 1 (1), 55-76

A Developmental Analysis of Children's Equivalence-class Formation and Disruption

Carol Pilgrim, Rebecca Click & Mark Galizio[1]
University of North Carolina Wilmington

Abstract

Developmental differences in children's conditional discrimination learning, equivalence-class formation, and equivalence-class disruption were investigated in two experiments. In Experiment 1, children between 2 and 9 years of age demonstrated age-related differences across a series of preliminary training steps, such that time to acquisition was more variable for younger than for older children on an initial identity matching and category matching task. However, upon completion of the preliminary training, there were no age-related differences in time to acquisition of the two arbitrary conditional discriminations that would serve as the basis for equivalence-class formation, nor were there differences in time to demonstrate stable equivalence classes (Experiment 2). Also in Experiment 2, children between 2 and 14 years of age were exposed to a potential challenge to the demonstrated equivalence classes; the reinforcement contingency for the AC conditional discrimination was reversed (i.e., given A1, A2 or A3, reinforcers were produced by selecting C2, C3, or C1 respectively). While there was little change in performance on reflexivity or BA symmetry tests following the challenge, age-related differences were obtained for CA symmetry and combined tests for equivalence. The older children were more likely to demonstrate an orderly change in equivalence-class membership consistent with the reversal training, while the younger children showed either little change or substantial disruption in their equivalence patterns. These data are considered in relation to more traditional investigations of children's category formation, as well as their implications for the study of equivalence-class formation and flexibility.

Key words: Equivalence-class formation, Equivalence-class flexibility, Contingency reversal, Category formation, Children, Mouse click.

Un Análisis del Desarrollo de la Formación de Equivalencia de Clases y de su Disrupción por Niños

Resumen

En dos experimentos se investigaron las diferencias en el desarrollo infantil en el aprendizaje de discriminaciones condicionales, la formación de equivalencia de clases y la disrupción de la equivalencia de clases. En el Experimento 1 se demostraron diferencias relacionadas con la edad en niños entre 2 y 9 años, a través de una serie de pasos preliminares de entrenamiento, de tal manera que el tiempo para la adquisición de la igualación de identidad y en una tarea de categorías fue más variable para los niños más jóvenes que para los niños mayores. Sin embargo, después de completar el entrenamiento preliminar, no hubo diferencias relacionadas con la edad en el tiempo de adquisición de las dos discriminaciones condicionales arbitrarias que servirían como base para la formación de equivalencia de clases, así como tampoco hubo

---

[1] The authors express their most sincere thanks to the many students of the Stimulus Equivalence lab who were essential to the conduct of this work. Address correspondence to Carol Pilgrim, College of Arts and Sciences, 109 Bear Hall, UNC Wilmington, Wilmington, North Carolina 20493 (E-mail: pilgrimc@uncw.edu).

diferencias en el tiempo para mostrar equivalencia de clases (Experimento 2). En el Experimento 2 también se expuso a los niños entre 2 y 14 años a un desafío potencial de las clases de equivalencia formadas, la contingencia de reforzamiento para la discriminación condicional AC se revirtió (i.e., dado A1, A2 o A3, la selección de C2, C3 o C1 produjo reforzamiento, respectivamente). Mientras que la ejecución en la reflexividad o la simetría BA en pruebas de simetría cambió poco después del desafío, se obtuvieron diferencias relacionadas con la edad para la simetría CA y en pruebas combinadas de equivalencia. Los niños mayores mostraron con mayor facilidad un cambio ordenado en la membresía a una clase equivalente consistente en entrenamiento en reversión, mientras que los niños más jóvenes mostraron poco cambio o una disrupción substancial en sus patrones de equivalencia. Estos datos se consideraron en relación con investigaciones más tradicionales sobre la formación de categorías en niños, así como sus implicaciones para el estudio de formación y de flexibilidad de equivalencia de clases.

Palabras clave: Formación de clases de equivalencia, Flexibilidad de las clases de equivalencia, Reversión de la contingencia, Formación de categorías, Niños, Clic del mouse.

Developmental differences in children's concepts and categories have frequently been observed in the study of perceptually based categories (e.g., Hayes & Taplin, 1992; Markman, 1989; cf., Osborne & Calhoun, 1998), as well as with respect to the influence of beliefs about the category (e.g., Keil, 1992). In contrast, much less is known about developmental differences involving categories for which members share no perceptual features or correlated attributes (i.e., functional or arbitrary categories). An increasingly important experimental approach to the study of such categories is exemplified in behavior-analytic work on stimulus equivalence (e.g., Sidman, 1994; 2000; Sidman & Tailby, 1982). Over the past several years, this approach has been widely applied to the study of complex cognitive functions in normally developing and developmentally delayed populations (e.g., Galizio, Stewart, & Pilgrim, 2001, 2004; Lipkens, Hayes, & Hayes, 1993; Pilgrim, Jackson, & Galizio, 2000; Wilkinson, Dube, & McIlvane, 1996, 1998; Wilkinson & McIlvane, 2001).

Standard procedures used to study stimulus equivalence begin with arbitrary match-to-sample (MTS) training, where physically dissimilar stimuli are used to establish at least two interrelated conditional discriminations. On each trial, children are presented with one of at least two possible sample stimuli (e.g., A1 or A2), and at least two comparison stimuli (e.g., B1 and B2). Selection of the correct comparison stimulus produces reinforcers, and the comparison designated as correct on any given trial is conditional on the specific sample presented (e.g., comparison B1 would be designated as correct given A1 as a sample, while comparison B2 would be correct given A2). A second conditional discrimination is trained in a similar manner using new comparison stimuli (e.g., comparison C1 would be designated as correct given A1 as a sample, while comparison C2 would be correct given A2).

What has captured the attention of researchers most about such procedures is that after learning these baseline discriminations, both children and adults have reliably shown the emergence of untrained stimulus relations when presented with novel trial types based on mathematical set theory; symmetry, transitivity, and reflexivity (Sidman & Tailby, 1982). Using the training examples given above, a symmetry test could involve the presentation of say, B1 as sample stimulus, and A1 and A2 as comparisons. Choice of A1 on this trial would reflect stimulus symmetry in that the trained functions of sample and comparison stimuli are reversible. A transitivity test could involve presentation of B1 as a sample stimulus, and C1 and C2 as comparisons. Choice of C1 on this trial would reflect stimulus transitivity in that the sample and comparison stimuli have never been directly related on training trials. Emergent symmetry would also be required on such a trial in that the sample stimulus has never previously functioned in that role; such trials are frequently referred to as "combined tests". Finally, a reflexivity trial could involve presentation of A1 as sample stimulus, and A1 and A2 as comparisons. Choice of A1 on this trial would reflect stimulus reflexivity in that untrained relations are demonstrated between each stimulus and itself. Thus, after directly training the four relations described above (A1B1, A2B2, A1C1, A2C2), stimuli become related to each other in ways that were never reinforced, and an additional 14 stimulus-control relations emerge (i.e., B1A1, B2A2, C1A1, C2A2, B1C1, B2C2, C1B1, C2B2, A1A1, A2A2, B1B1, B2B2, C1C1, C2C2). The stimuli that become related to each other in this manner are termed equivalence classes (i.e., A1, B1, and C1 as one class; A2, B2, and C2 as another), in that all elements are functionally substitutable within a given context (Sidman, 1994).

Equivalence classes allow for the study of many interesting features of class or category formation in that unfamiliar, physically dissimilar stimuli with which a child has no experience come to function similarly and interchangeably in novel ways. Such classes capture the sort of efficiency that is often held to be a defining feature of categories, if not their primary function, and provide a basis for what is often described as inductive inference, also argued to be criterial for categories (e.g., Markman, 1989).

In addition, the stimulus-equivalence paradigm provides for important methodological rigor in the study of category formation in that the experiences giving rise to these classes and the extent of their exposure can be controlled.

Although demonstrations of equivalence-class formation have proven difficult with non-human animal populations (e.g., Dugdale & Lowe, 1990; Lipkens, Kop, & Matthijs, 1988; Sidman, Rauzin, Lazar, Cunnigham, Tailby & Carrigan, 1982; but note also Kastak, Schusterman & Kastak,, 2001; Schusterman & Kastak, 1993), after acquiring the prerequisite baseline conditional discriminations, equivalence classes have been reliably shown in typically developing children (e.g., Barnes, Smeets, & Leader, 1996; Devany, Hayes, & Nelson, 1986; Michael & Bernstein, 1991; Pilgrim, Chambers, & Galizio, 1995; Sidman & Tailby, 1982) and even with developmentally delayed populations (e.g., Carr, Wilkinson, Blackman, & McIlvane, 2000; Sidman, 1994). Although frequently studied in children, developmental analyses of equivalence-class formation are few.

One complicating factor in developmental comparisons of children's equivalence performances lies in the difficulty often associated with acquisition of the baseline conditional discriminations (e.g., Augustson & Dougher, 1991; Gollin, 1966; Gollin & Savoy, 1968; Lipkens et al., 1993; Pilgrim et al., 2000). For example, Lipkens et al. describe acquisition failure with a 12 month-old subject when training procedures involved differential reinforcement only; the same child showed rapid learning at 16 months when discriminations were taught in a verbal context involving animal names and noises. Other procedures successful with young children have involved modeling, instructions, naming, or other unspecified training aids, thus making developmental comparisons across, or even within, studies difficult.

Further questions about children's equivalence classes involve their stability once formed. Indeed, in addition to the study of category formation, the equivalence paradigm allows for investigation of how categories change. Traditional views of categories hold organization at any given point to be the outcome of two conflicting tendencies; 1) the tendency to modify categories to reflect new experiences, and 2) the tendency to resist change, due to the effort required and the loss of continuity with previous systems (e.g., Markman,1989; Piaget & Inhelder, 1969). Pilgrim et al. (1995) studied equivalence-class flexibility as a way of approaching category change in children. Children aged 5-7 years learned A1B1, A2B2, A1C1, and A2C2 discriminations, and demonstrated the emergence of two 3-member equivalence classes (i.e., A1B1C1 and A2B2C2). These classes were then challenged by training a reversal of the AC relations. During the class-challenge condition, choosing comparison stimulus C2 (instead of C1) was reinforced when A1 served as sample, while choosing comparison C1 (instead of C2) was reinforced when A2 served as sample. This reversal training might have been expected to bring about a change in equivalence-class organization (i.e., A1B1C2 and A2B2C1), reflected by altered performances on CA symmetry trials (e.g., choosing A2 given a C1 sample) and on BC and CB transitivity/equivalence trials (e.g., choosing C1 given B2 as a sample). However, despite mastering the AB and reversed AC relations, all of the children showed disrupted probe performances that were consistent with neither the originally established classes nor the classes that might have been expected to follow from the reversed baselines. These findings were in marked contrast to the effects of baseline reversals in adults who, for any given probe type, have shown patterns consistent with either the original or the new training relations (Dube, McIlvane, Mackay, & Stoddard, 1987; Garotti, de Souza, de Rose, Molina, Renata, & Gil, 2000; Pilgrim & Galizio, 1990, 1995; Saunders, Saunders, Kirby, & Spradlin, 1988; Spradlin, Cotter, & Baxley, 1973; Spradlin, Saunders, & Saunders, 1992; Wirth & Chase, 2002).

An interesting aspect of the Pilgrim et al. (1995) data was that the youngest children showed the most disrupted probe patterns, and the oldest child showed the most adult-like profile, suggesting a developmental trend in class flexibility. The possibility of such a trend is further supported by data from Micheal and Bernstein (1991), whose young participants (4 and 5 year olds) also showed disrupted probe performances following a conditional discrimination reversal, and by data from Spradlin et al. (1992), who found more adult-like patterns (as described above) in an 8 and a 12 year-old under similar conditions. However, a follow-up study (Saunders, Drake, & Spradlin, 1999) with 3-5 year olds reported that while one subject showed disrupted probe performances, three others showed probe patterns that were predominantly consistent with the reversed baseline relations – a pattern previously observed only in adults, as noted above. Many methodological differences other than subject age distinguish these studies, including experimenter instructions, training sequence, type of stimuli, stability criteria, duration of exposure to the original training relations prior to reversal, etc. To date, the effects of equivalence-class challenges in children of a range of ages have not been explored by any single laboratory or by any single set of experimental methods, underscoring the need for systematic developmental analyses using standardized procedures. The present study was designed to study the acquisition of conditional discriminations in young children (Experiment 1), the emergence of equivalence classes (Experiment 2), and the flexibility of those classes (Experiment 2) in children from a broader range of ages.

Experiment 1

Method

Participants

All children attending two preschools and two elementary schools were invited to participate via letters and consent forms sent home to parents. Participants were the 97 children whose parents completed permission forms. All children were students in classrooms for typically developing children. Ages ranged from 2 years 1 month to 8 years 9 months; 56 of the participants who began the study were female and 41 were male.

Apparatus

Experimental stimuli were black-and-white line drawings approximately 1.5 to 2 cm square presented on a white screen background on either a MacIntosh Performa or Power PC computer (30 cm diagonal screens), according to specialized MTS programming (Dube, 1991). The sample stimulus always appeared in the center of the screen, and comparison stimuli could appear in any of the corners. Manipulating a mouse moved a cursor on the screen. (Children who had no previous experience with computers were taught to point and click with the mouse using commercial software.) When the cursor was situated on or near a stimulus, clicking on the mouse registered a response. Following a response designated as correct, a brief fanfare sounded during which colored stars transversed the computer screen. Following a response designated as incorrect, a buzzer sound was produced and the screen immediately went blank. All stimulus presentations and data collection occurred automatically.

Procedure

General Procedure. Sessions were conducted five days a week, or as often as possible given scheduling conflicts and absences. Each session lasted approximately 15 min and was programmed to include either 24 (when tasks involved one or two comparisons) or 36 (when tasks involved three comparisons) MTS trials. A trial began with the presentation of a sample stimulus in the center of the screen. A response to the sample resulted in the presentation of two or three comparison stimuli in the corners of the screen. A response to one of the comparison stimuli produced the appropriate consequences, and the next trial began following a 1.5s inter-trial interval.

For any given experimental condition, stimulus presentations were arranged such that each sample stimulus appeared an equal number of times in an irregular sequence, and no one sample was presented on more than three consecutive trials. Comparison stimuli appeared an equal number of times in each corner position, each was correct an equal number of times in an irregular sequence, and the correct comparison stimulus was not in the same position for more than three trials. The mastery criterion for each training phase of the study required two consecutive sessions with correct responses on 90% or more of the trials.

At the end of each session, the participants received edible reinforcers of their choosing (e.g., fruit bits, candy), the opportunity to play a commercially available computer game unrelated to the experiment (e.g., Reader Rabbit, Storybook Weaver, Thinking Science), and a sticker. To encourage continued participation, stickers were accumulated and exchanged for age-appropriate prizes (e.g., yo-yos, CDs).

Initial training sequence. Given the difficulties noted above in establishing arbitrary conditional discriminations with young children, a standardized three-phase training sequence was used with all participants. In Phase 1, all stimuli were pictures of familiar objects. A response was reinforced if the selected comparison stimulus was physically identical to the sample (e.g., given a heart as the sample stimulus, choosing the heart, but not the fish or the pencil, was reinforced). Phase 2 also involved identity matching, but with unfamiliar, abstract stimuli. In Phase 3, the sample and reinforced comparison stimulus were not physically identical, but rather were members of a common class or category such as animals, vehicles, or body parts (e.g., given a cow as a sample, selecting a pig, but not a truck or a hand, was reinforced).

If a participant failed to meet mastery criterion for a particular phase within 10 sessions and if there was no trend toward acquisition, the training sequence was systematically altered to facilitate learning by simplifying the task. First, the three-choice conditional discrimination task was reduced to a two-comparison and then if necessary, a one-comparison task. When mastery criteria were met on any task, the next-most complex task was reinstated.

Arbitrary conditional discrimination training. In this training phase, all stimuli were abstract black-and-white line drawings. The sample stimulus on each trial was either A1, A2, or A3, and the comparison stimuli were B1, B2, and B3. (These alphanumeric labels were not available to the children.) When A1 was the sample, the reinforced choice was B1. When A2 or A3 served as sample, the reinforced choices were B2 and B3, respectively. When mastery criteria were met, a second conditional discrimination was introduced and trained in the same manner. Sample stimuli again consisted of A1, A2, or A3, and the comparison stimuli were C1, C2, and C3 (see Figure 1). Choosing C1, C2, or C3 was reinforced in the presence of A1, A2, or A3, respectively. In the final phase of Experiment 1, 12 AB and 12 AC trials were randomly intermixed. The number of sessions to master the AB, the AC, and the mixed discriminations were the dependent measures.

Results

The upper left panel of Figure 1 shows the total number of sessions to master the three-phase initial training sequence (e.g., identity matching with familiar stimuli, identity matching with abstract stimuli, and category matching) for each child, plotted as a function of age. Each data point represents performance of an individual child. There were children at every age who met the mastery criteria in the minimum number of sessions (6); however, there was much greater variability in the number of sessions required to meet mastery among the younger children. This resulted in a strong negative correlation, $r(65) = .665$, $p < .0001$, between age and the number of sessions required to master the initial training sequence. To determine the source of this developmental effect, Figure 1 also presents data for each of the three phases independently. The top right panel presents data from the initial phase, which involved identity matching with familiar stimuli. Most children mastered this task with little difficulty, but again, there was a significant negative correlation between age and sessions to mastery $r(95) = .46$, $p < .0001$. Much of the variance was accounted for by the performances of the very youngest children (2-4 years of age). The lower left panel presents data for the identity-matching task involving abstract stimuli. What is striking here is the rapid mastery of this task by most children. It would appear that the match-to-sample performances acquired in the initial phase generalized to the novel, abstract stimuli in most cases. While some younger children made this transition less smoothly, there was not a strong relation between age and mastery $r(80) = .27$, $p = .015$. However, the developmental trend reappeared when the category task was introduced $r(66) = .47$, $p < .0001$.

Figure 2 shows the number of sessions required to master the initial AB (top panel) and a second, AC (bottom panel), arbitrary conditional discrimination for each child. Of particular interest for both measures is the absence of a relation between age and sessions to mastery (for AB mastery, $r(63) = 0.15$, $p = .24$; for AC mastery, $r(51) = 0.08$, $p = .59$). Most children acquired the conditional discriminations within a few sessions of the minimum number of sessions required by the mastery criteria. There were several outliers for each discrimination-training phase, but these were found across a range of ages. T-tests revealed no difference in the number of sessions to mastery for the AB (mean = 6.09 sessions) and AC (mean = 6.67 sessions) conditional discriminations $t(48) = .65$, $p > .05$.

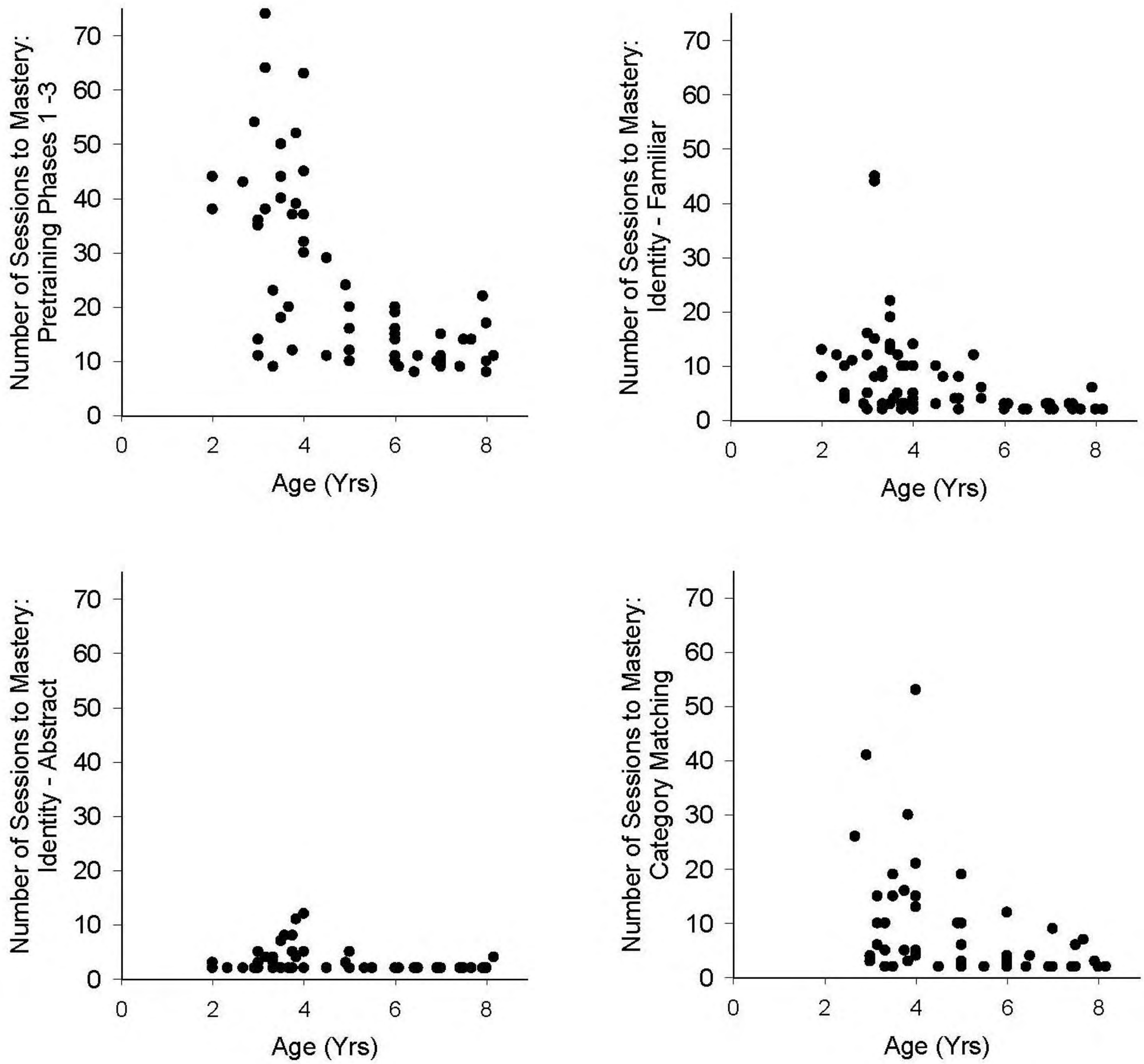

Figure 1. Number of sessions required to meet mastery criterion for preliminary training phases 1 – 3, collectively (top, left panel) and individually (top, right panel for identity matching with familiar pictures; bottom, left panel for identity matching with abstract line drawings; bottom, right panel for category matching) as a function of age. Each data point represents the number of sessions required by an individual child.

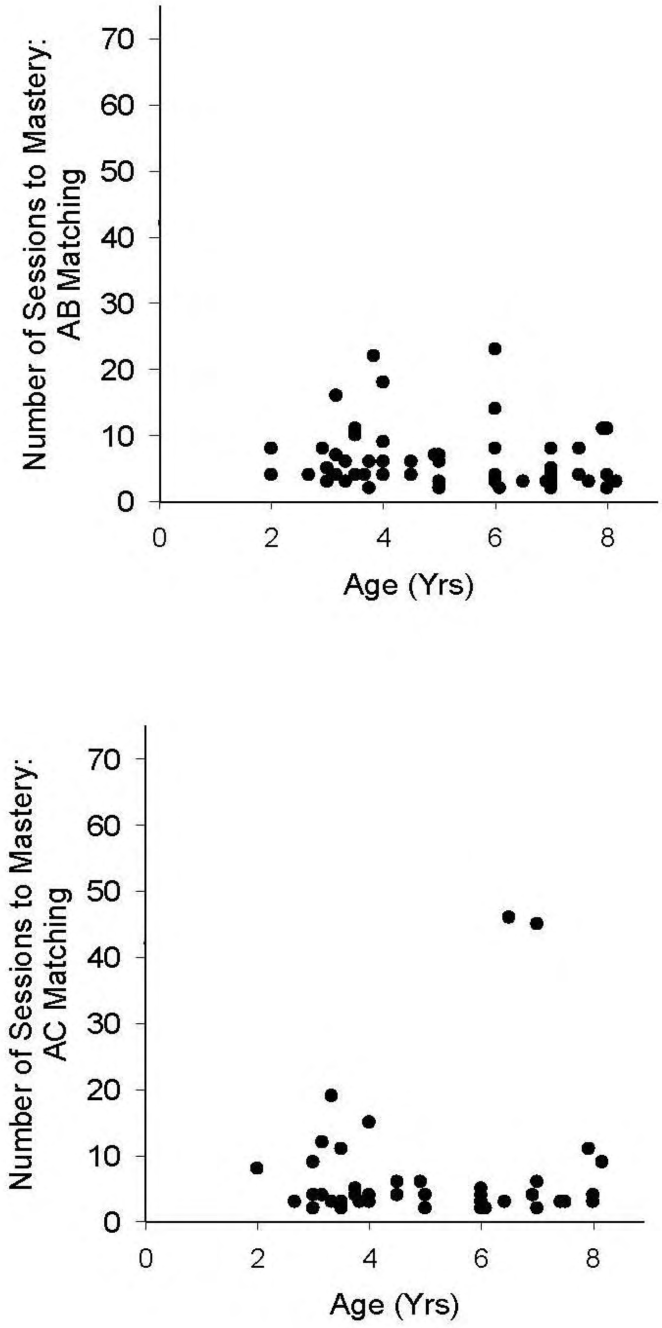

Figure 2. Number of sessions required to meet mastery criterion for AB (top panel) and AC (bottom panel) conditional discrimination training as a function of age. Each data point represents the number of sessions required by an individual child.

Discussion

The data of Experiment 1 are interesting to consider in light of the frequently reported finding that arbitrary conditional discrimination acquisition can be problematic for young children (e.g., Augustson & Dougher, 1991; Gollin, 1966; Gollin & Savoy, 1968; Lipkens et al., 1993; Pilgrim et al., 2000). Here, developmental differences were clear for mastery of the initial training sequence, but not for the arbitrary relations. Older children tended to show rapid mastery of all three phases of the initial training sequence. In contrast, there was far greater variability in the number of sessions required for the younger children to master the identity-matching task with familiar stimuli in Phase 1. These results replicate findings from previous developmental studies of matching to sample (e.g., Kraynak & Raskin, 1971), and expand the age-range compared. Of interest was that even children who had difficulty with the Phase 1 problems, showed rapid mastery in Phase 2, identity matching of abstract shapes. Thus, it appeared that Phase 1 training resulted in generalized matching to sample across all ages. However, when this perceptually based matching task was shifted to the Phase 3 category matching, in which correct selections could not be based on physical identity, the developmental differences reappeared; young children were again much more variable with respect to the number of sessions to mastery. However, after the categorical match-to-sample task was mastered, arbitrary matching (AB and AC) was acquired at a comparable rate across ages. It might be argued that the age differences in Phase 3 were related to the fact that accurate matching could no longer be based on perceptual similarities alone. While shared physical features may have been important in the pre-experimental acquisition of these categories, accurate matching in Phase 3 would also seem to require relations among features that were not perceptually based. Control by non-physical relations in Phase 3 seemed to facilitate acquisition of the purely arbitrary AB and AC relations. Thus, the training sequence appeared to shape, in successive steps, control by increasingly arbitrary stimulus relations, and might be usefully considered in the context of learning set (e.g., Harlow, 1949) and other generalized learning phenomena (e.g., higher-order operants like generalized imitation; Catania, 1998; Baer, Peterson, & Sherman, 1967; Pilgrim & Galizio, 1996).

A point to note is that not all children completed the experiment, and the total number of children included in the analysis for each phase drops from the pre-training steps to the AB and AC training. In an experiment of this sort that requires extended testing, attrition due to a number of factors is common. In the present study, attrition was due exclusively to children leaving the preschool or after-school program that served as the study site. Despite these losses, each age-level was well represented at each phase in the present study. An important feature of our training sequence is that it permitted even very young children to acquire the arbitrary matching task in a number of sessions comparable to that required by the older children. Control over exposure to these prerequisites made possible a developmental analysis of the emergence and flexibility of equivalence performances following acquisition of the baseline conditional discriminations.

Experiment 2 examined these variables and, based on previous data from this laboratory (Pilgrim & Galizio, 1990; 1995; Pilgrim, Chambers, & Galizio, 1995), included older children in order to capture the full range of developmental differences. Following their AB and AC baseline acquisition, the children in Experiment 2 were tested for the emergence of equivalence classes (i.e., A1B1C1, A2B2C2, and A3B3C3), exposed to an AC class challenge (i.e., A1C2, A2C3, and A3C1), and then re-tested for the emergence of modified classes (i.e., A1B1C2, A2B2C3, and A3B3C1). Evidence for such modification would come from altered patterns on CA-symmetry and transitivity/equivalence probe trials, but not on BA-symmetry or reflexivity trials. Any evidence of altered patterns on these latter trial-types would be indicative of a more general disruption (Pilgrim et al., 1995).

Experiment 2

Method

Participants

Twenty-two children from Experiment 1 were available for sufficient time to complete the remaining phases of the study. Ages ranged from 2 years, 9 months to 8 years, 7 months; 14 were female and eight were male. An additional 10 older children (ages 9 years to 13 years 3 months) were tested at a local elementary school and a local middle school in Experiment 2. Five were female and five were male.

Apparatus

The apparatus was the same as for Experiment 1.

Procedure

The general procedures were the same as in Experiment 1 with exceptions critical to each phase described below.

Equivalence Testing. The new participants completed exactly the same sequence of training steps described for Experiment 1 up through the mixed AB/AC training phase. The children from Experiment 1 moved from mastery of the mixed training phase immediately to Experiment 2. The first new phase of Experiment 2 continued the mixed discrimination training with reduced reinforcer density, such that 75% and then 50% of the trials included programmed reinforcers, and mastery was required with each reduction. These steps were designed to prepare the children for equivalence-test trials on which no reinforcers were available. No instructions were given regarding no-reinforcement trials. If questions were raised, the child was reassured that the equipment was functioning properly and encouraged to continue.

After demonstrating mastery of the mixed AB and AC conditional discriminations with reinforcers available on 50% of the trials, each subsequent session included one of the following probe-trial types; either symmetry, reflexivity, or combined tests for transitivity and symmetry (hereafter, combined tests). Probe trials, for which reinforcers were never available, were unsystematically intermixed with AB and AC trials, for which reinforcers were intermittently available, such the overall reinforcer density for the session was maintained at approximately 50%. The symmetry and combined test sessions included six probe trials (B1:A1A2A3; B2:A1A2A3; B3:A1A2A3; C1:A1A2A3; C2:A1A2A3; and C3:A1A2A3 for symmetry and B1:C1C2C3; B2:C1C2C3: B3:C1C2C3; C1:B1B2B3; C2:B1B2B3; C3:B1B2B3 for combined tests, where the first stimulus indicates the sample and the next three, the comparisons) randomly intermixed with 18 AB and AC trials. No two probe trials were presented in succession. The reflexivity sessions included nine probe trials (A1:A1A2A3; A2:A1A2A3; A3:A1A2A3; B1:B1B2B3; B2:B1B2B3; B3:B1B2B3; C1:C1C2C3; C2:C1C2C3; C3:C1C2C3) intermixed with 27 AB and AC trials. A cycle of three sessions included symmetry, reflexivity, and combined tests, in that order. The principal dependent measure was the percentage of trials in each session on which the selected comparison was consistent with the equivalence classes that would be expected to follow from the training contingencies (e.g., on a symmetry trial, comparison A1 would be selected given a B1 sample). This testing cycle was repeated until performance on all trial types met a six-session stability criterion where the difference between the mean percentage on the first three and the second three sessions did not exceed the grand mean by more than 10%.

Equivalence-class Challenge. Upon completion of equivalence testing, a class challenge was arranged by altering the reinforcement contingencies for the AC baseline conditional discrimination. Each session included 36 AC trials. When stimulus A1 was presented as sample, selection of comparison C2, rather than C1, was reinforced. Similarly, in the presence of sample A2, choosing C3 was reinforced, and in the presence of sample A3, choosing C1 was reinforced. When AC performances met mastery criteria, 18 AB and 18 AC trials were intermixed. Reinforcement contingencies for AB performances were unchanged from the original training. Reinforcement density was reduced in successive steps to 75% and then to 50%. When mastery criteria were met for each of these steps, equivalence probe trials were re-introduced. Session composition during equivalence testing was exactly the same as during the original equivalence-testing phase.

Return to Baseline. When stability criteria were met and availability allowed, the original reinforcement contingencies for the AC conditional discrimination were re-instated (i.e., in the presence of sample A1, A2, or A3, reinforcers followed choice of comparison stimulus C1, C2, or C3, respectively). The same training steps used in the class-challenge phase were followed (i.e., AC trials only, intermixed AB and AC trials, reinforcer reductions, and equivalence testing).

Results

The ten children new to the study rapidly mastered the AB and AC conditional discriminations, and all of the children progressed quickly through the mixed discrimination training with reduced reinforcer density. All children, regardless of age, also showed strong and stable evidence of equivalence as defined by high percentages of class-consistent responding on reflexivity, symmetry, and combined tests (symmetry and transitivity). The upper panel of Figure 3 examines equivalence performances more closely by presenting the number of sessions required to meet the stability criteria. Some children showed evidence of delayed emergence in their equivalence performances (see Sidman, 1994), but there were children at all ages who met stability criteria in the minimum number of sessions (18). As Figure 3 shows, there was more variability in the number of sessions required by children 8 years old and younger, but the overall correlation between age and sessions to stability was not significant $r(30) = 0.059$, $p = .75$.

The lower panel of Figure 3 presents the number of sessions required to meet mastery criteria when the AC contingencies were altered. Most children acquired the altered conditional discrimination rapidly, and the variability accounted for by age was not significant $r(23) = 0.37$, $p = .08$. Similarly, all children quickly mastered the mixed AB/AC training and the reduced reinforcement phases.

Figure 4 presents performances for each child on probe trials testing for each of the properties of equivalence following the AC class challenge. The data are expressed as percent change from the stable pre-class challenge performances (that is, the difference between the means for the final 6 sessions of the pre- and post-class challenge probe conditions). There was little evidence of change in either reflexivity or BA-symmetry performances for children at any age (see top panels of Figure 4; for the relation between reflexivity performances and age, $r(23) = 0.03$, $p = .88$; for BA symmetry patterns, $r(23) = 0.20$, $p = .39$). In contrast, performances on CA-symmetry (see bottom, left panel) and transitivity/equivalence trials (see bottom, right panel) showed much more change, at least among the older children. In both cases, there was a strong positive correlation between age and percent change from baseline (for CA symmetry, $r(23) = 0.53$, $p = .014$; for transitivity/equivalence, $r(23) = 0.58$, $p = .001$). In general, children 10 and above approached 100% change on both measures; for example, on CA-symmetry trials with C1 as sample, comparison A3 was chosen almost exclusively, while A1 was chosen given C2 as sample, and A2 was chosen given C3 as sample. Similar changes occurred on the BC and CB transitivity/equivalence trials which are presented together in Figure 4. In short, the older children tended to demonstrate new classes following the AC challenge, as evidenced by emergent probe patterns that were consistent with the new contingency.

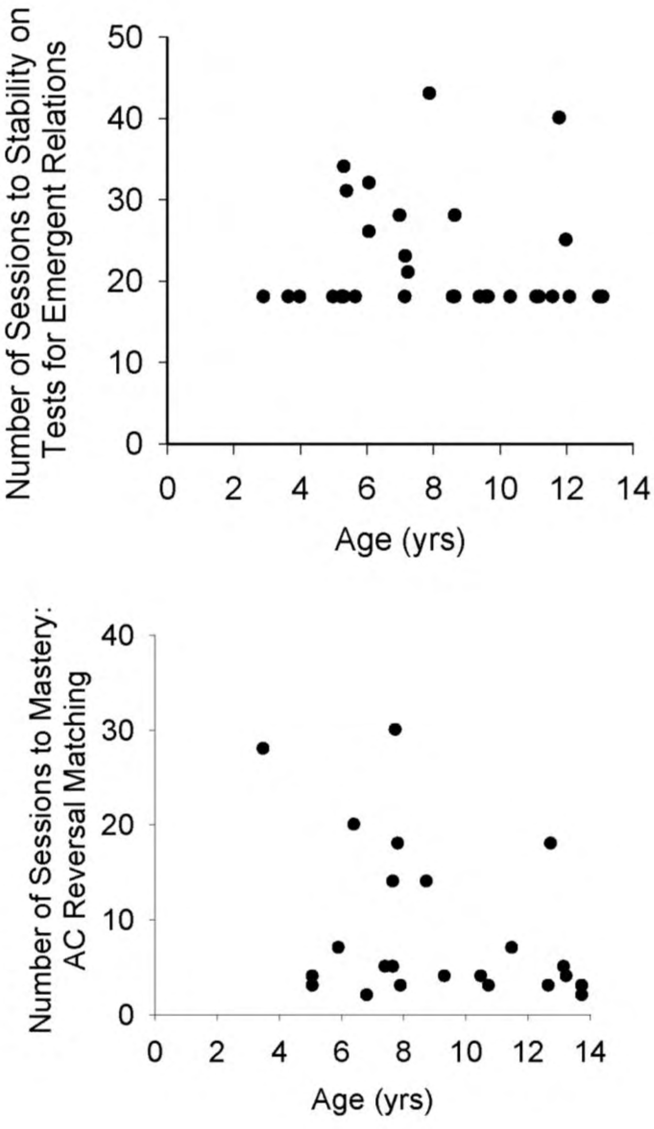

Figure 3. Number of sessions required to meet stability criterion for the emergent relations indicative of equivalence-class formation as a function of age. Each data point represents the number of sessions required by an individual child.

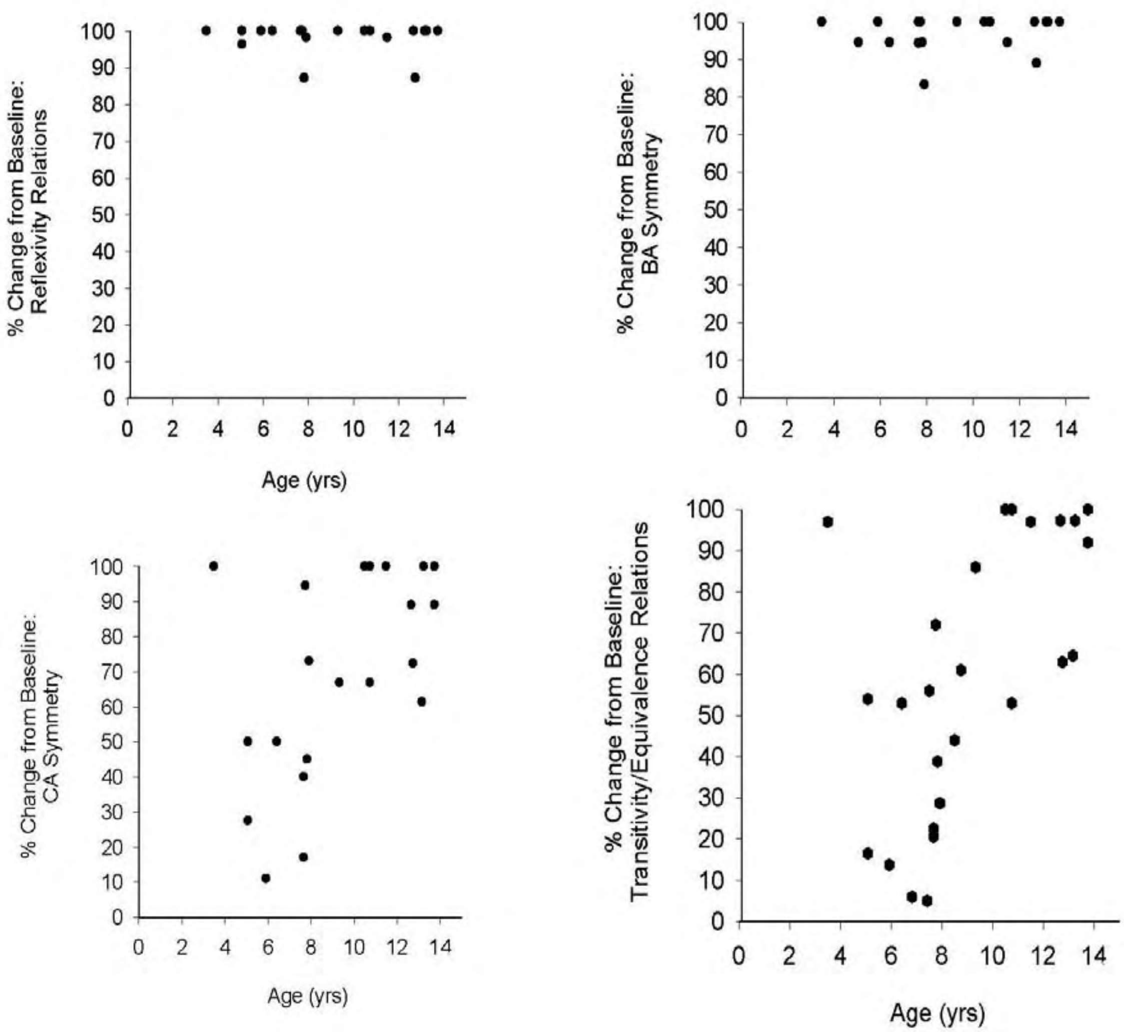

Figure 4. Percent change in equivalence-probe performances from the final 6 sessions of stable pre-class-challenge levels to the final 6 sessions of stable post-class-challenge levels as a function of age. Data are shown for percent change in reflexivity performances (top, left panel), BA symmetry performances (top, right panel), CA symmetry performances (bottom, left panel) and performances on combined tests for equivalence (bottom, right panel). Each data point represents the percent change for an individual child.

Different patterns were observed among the younger children. Several of the youngest children showed very little impact of the AC-class challenge, as evidenced by a lack of change from their original probe patterns. Many of the younger children showed inconsistent responding on probe trials, with intermediate change patterns somewhere between 25% and 75% change. While class reorganization was generally confined to children 10 and older, it should be noted that the youngest child tested under these conditions showed virtually 100% change from her original equivalence patterns.

Discussion

The strong equivalence performances demonstrated in the original testing conditions by the children in Experiment 2 are of note for several reasons. First, there have been few studies of equivalence in very young children, particularly in the absence of extensive verbal prompting. In fact, the 2 years, 9 month-old child here appears to be one the youngest in the published literature to have shown equivalence in a situation in which verbal interventions and inadvertent experimenter cueing were ruled out. The absence of a developmental difference in the emergence of equivalence is striking, particularly given that class or category formation did not involve physical similarities between class members and that the performances indicative of equivalence emerged without direct reinforcement. Of course, not all of the children who began the study in Experiment 1 completed the training, but all who did showed virtually perfect class-consistent probe performances. The consistency with which equivalence emerged following the baseline training suggests that equivalence-class formation is a robust phenomenon not influenced by the sorts of major individual differences represented in this sample and that it is not dependant on particular educational experiences, sophisticated verbal abilities, or formal reasoning skills (see also, Carr et al., 2000). Indeed, it has even been argued that equivalence represents a basic behavioral process (Sidman, 1994; 2000). The present data are consistent with the possibility of equivalence as either a basic process or at least one with relatively simple pre-requisites that could be mastered very early in life (e.g., Hayes, Barnes-Holmes, & Roche, 2001; Horne & Lowe, 1996).

In contrast to the original equivalence-probe patterns, strong developmental trends were apparent on probe trials following the class challenge. The older children tended to change their patterns of responding on just those probe trials (i.e., CA symmetry and combined trials) where change would be predicted, based on the class challenge. Thus, the older children showed evidence of equivalence classes different from those of the original testing phase. The younger children showed little evidence of new class formation on probe trials, despite their near-perfect performances on the baseline trials that were directly involved in the contingency manipulation. These findings replicate the early developmental trend noted previously for young children (Michael & Bernstein, 1991; Pilgrim et al., 1995), and extend it to older participants. In the Pilgrim et al. study, only the eldest child (7 years) showed evidence of control by the contingency manipulation and even then, probe patterns were not completely consistent with class re-organization. Data from Spradlin et al. (1992) with older participants (8 and 12 years) are also consistent with this developmental trend, but more recent findings by Saunders et al. (1999) reflect greater evidence of class re-organization in three of four pre-school children than was observed here. There were many differences between their training procedures and those used in the present study. For example, the baseline conditional discrimination that was reversed was established without reinforcement, and thus may have been less resistant to change (Nevin, 1992). An important advantage of the present study was the common set of training and testing procedures used across ages. While the studies noted above were suggestive, the present study confirms the presence of a developmental trend in response to class challenge.

The nature of this developmental trend warrants closer analysis. Interpretation of the patterns at either extreme is relatively straightforward (see the CA symmetry and combined test panels of Figure 7); either the original equivalence classes were maintained, despite the class challenge (some younger children), or new classes emerged following the challenge (most older children). However, the majority of the children showed intermediate patterns. In these cases, the class-challenge did produce an effect, but rather than produce class re-organization, disrupted probe performances were the result. On any given session, AB and AC conditional discriminations were completely consistent with the class-challenge contingencies, but choices on CA-symmetry trials and combined tests for equivalence varied from trial to trial. Some responses were consistent with the classes that would follow from the challenge contingencies and others were consistent with the original equivalence classes.

The basis for these mixed patterns and their age-related occurrence remains to be determined, but several alternatives may be considered. One possibility is that the two sets of training contingencies (i.e., the original training and the class-challenge training) established two sets of equivalence classes that compete for control of responding on any given probe trial following the class challenge. For older children, the baseline relations currently in effect seem more likely to select the set of equivalence classes that would be most appropriate, while the younger children seemed less sensitive to contextual control by the baselines in effect. By this account, the baseline relations serve two functions for the older children; they provide the basis for the emergence of new equivalence classes (as they would for younger children) and serve as contextual determinants of the classes to be demonstrated on any given trial. This would suggest that the problem for the younger children is one related to contextual control over equivalence classes (e.g., Bush, Sidman, & de Rose, 1989, Wulfert & Hayes, 1988) or stimulus-control topographies (i.e., Dube & McIlvane, 1996).

Another alternative is that the class challenge provided the basis for equivalence class collapse, or class union (Sidman, 1994). Following a history in which stimulus C1, for example, was directly related to A1 (and related through equivalence to B1), the new training directly relates C1 to A2 (and therefore to B2). C1 is then held in common by these potentially separate sets of stimuli, and thus could function as a node, allowing two smaller classes to merge into one larger class. If such a merger occurred, disrupted performances might be expected; in essence, any given trial would require a choice between comparison stimuli that were all related to the sample. By this account, older children are more sensitive than younger children to the bases for class partition (i.e., the baseline relations currently in effect). The class-merger view might suggest that providing young children with experiences designed to enhance discrimination between classes would result in more orderly performances on post-reversal probe trials (Sidman, 2000).

General Discussion

In summary, the present experiments found developmental differences in some aspects of children's conditional discriminations, while these were notably absent for other measures. For example, younger children had much more difficulty in mastering the pre-training tasks (Experiment 1), but once they had, mastery of arbitrary conditional discriminations (Experiment 1) and emergence of equivalence-class performances (Experiment 2) were rapid and unrelated to age. Following the class-challenge, however, development trends were again apparent (Experiment 2) in that older children showed more orderly patterns on any given probe type, as has been noted in studies with adults (see Pilgrim & Galizio, 1996, and Spradlin et al., 1992, for reviews). That developmental differences were observed in the pre-training and class-challenge conditions but not during the arbitrary or symbolic match-to-sample training and testing suggests that these age differences were not the simple product of selecting a task that was inappropriate for younger children. Indeed, the arbitrary training and testing might appear in some ways to be more demanding than the pre-training in that equivalence-class members are not related by perceptual similarity or correlated attributes. Thus it would appear that even young children generate the emergent performances that might be said to indicate the efficiency and inductive inference characteristic of categories when provided with the appropriate pre-requisites. These findings may parallel those from studies of natural language categories in which the developmental progression (e.g., from thematic to taxonomic categorization) has been eliminated by experiential variables (e.g., Markman, 1989; Osborne & Calhoun, 1998).

In many ways, the class-challenge data are a more interesting reflection of age differences in that older children showed class performances that quickly incorporated new learning experiences while younger children did not. One possibility suggested by the previous discussion is that this developmental difference might also be impacted by providing the appropriate training experiences to the younger participants, although these remain to be identified. It seems important that the equivalence methodologies employed here produce findings similar to those obtained with more naturalistic procedures often employed in research with young children (e.g., Markman, 1989). Further research will be needed to determine how closely the phenomena studied here parallel those observed with natural language categories (e.g., Galizio et al., 2001). Until that time, the stimulus-equivalence paradigm may raise issues of ecological validity in the present context (but see Wilkinson & McIlvane, 2001). However, equivalence methodologies have the advantage of allowing study of the basic processes involved in class formation in a manner that is relatively unconstrained by the influence of extra-experimental factors. Because equivalence class formation can be studied without the need for verbal instructions or context-setting stories it provides an interesting alternative to the sorts of procedures often used to study artificial categories with children (e.g., Markman, 1989). Equivalence methodologies may permit the sort of direct experimental control over relevant variables that is necessary to test questions about the factors responsible for developmental differences in the emergence and stability of stimulus classes or categories.

References

Augustson, K. G., & Dougher, M. J. (1991). Teaching conditional discrimination to young children: Some methodological successes and failures. *Experimental Analysis of Human Behavior Bulletin, 9*, 21-24.

Barnes, D., Smeets, P. M., & Leader, G. (1996). Procedures for generating emergent matching performances: Implications for stimulus equivalences. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus Class Formation in Humans and Animals* (pp. 153-171). Holland, The Netherlands: Elsevier.

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.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior, 51*, 29-45.

Carr, D., Wilkinson, K. M., Blackman, D., & McIlvane, W. J. (2000). Equivalence classes in individuals with minimal verbal repertoires. *Journal of the Experimental Analysis of Behavior, 74*, 101-114.

Catania, A. C. (1998). *Learning, 4th Edition*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*, 243-257.

Dube, W. V. (1991). Computer software for stimulus control research with Macintosh computers. *Experimental Analysis of Human Behavior Bulletin, 9*, 28-30.

Dube, W. V., & McIlvane, W. J. (1996). Implications of a stimulus control topography analysis for emergent behavior and stimulus classes. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus Class Formation in Humans and Animals* (pp. 197-218). Amsterdam, The Netherlands: Elsevier.

Dube, W. V., McIlvane, W. J., Mackay, H. A., & Stoddard, L. T. (1987). Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior, 47*, 159-175.

Dugdale, N. A., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman, & H. Lejeune (Eds.), *Behaviour Analysis in Theory and Practice: Contributions and Controversies* (pp. 115-138). Hove, England: Erlbaum.

Galizio, M., Stewart, K., & Pilgrim, C. (2001). Clustering in artificial categories: An equivalence analysis. *Psychonomic Bulletin & Review, 8*, 609-614.

Galizio, M., Stewart, K., & Pilgrim, C. (2004). Typicality effects in contingency-shaped generalized equivalence classes. *Journal of the Experimental Analysis of Behavior, 82*, 253-273.

Garotti, M., DeSouza, D. G., & De Rose, J. C., Molina, J. C., Renata, C., & Gil, M. S. A. (2000). Reorganization of equivalence classes after reversal of baseline relations. *Psychological Record, 50*, 35-48.

Gollin, E. S. (1966). Solution of conditional discrimination problems by young children. *Journal of Comparative and Physiological Psychology, 62*, 454-456.

Gollin, E. S., & Savoy, P. (1968). Fading procedures and conditional discrimination in children. *Journal of the Experimental Analysis of Behavior, 11*, 443-451.

Harlow, H. (1949). The formation of learning sets. *Psychological Review, 56*, 51-65.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. Reno, NV: Context Press.

Hayes, B. K, & Taplin, J. E. (1992). Developmental changes in categorization processes: Knowledge and similarity-based modes of categorization. *Journal of Experimental Child Psychology, 54*, 188-212.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 185-241.

Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior, 76*, 131-158.

Keil, F. C. (1992). *Concepts, kinds, and cognitive development*. Cambridge, Mass.: The MIT Press.

Kraynak, A. R., & Raskin, L. M. (1971). The influence of age and stimulus dimensionality on form perception by preschool children. *Developmental Psychology, 4*, 389-393.

Lipkens, G., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of derived stimulus relations in an infant. *Journal of Experimental Child Psychology, 56*, 201-239.

Lipkens, G., Kop, P. F. M., & Matthijs, W. (1988). A test of symmetry and transitivity in the conditional discrimination performances of pigeons. *Journal of the Experimental Analysis of Behavior, 49*, 395-409.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, Mass.: The MIT Press.

McIlvane, W. J., & Dube, W. V. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst, 26*, 195-213.

Michael, R. L., & Bernstein, D. J. (1991). Transient effects of acquisition history on generalization in a matching-to-sample task. *Journal of the Experimental Analysis of Behavior, 56*, 155-166.

Nevin, T. A. (1992). An integrative model for the study of behavioral momentum. *Journal of the Experimental Analysis of Behavior, 57*, 301-316.

Osborne, J. G., & Calhoun, D. O. (1998). Themes, taxons, and trial types in children's matching to sample: methodological considerations. *Journal of Experimental Child Psychology, 68*, 35-50.

Piaget, J., & Inhelder, B. (1969). *The Psychology of the Child*. New York, NY: Basic Books.

Pilgrim, C., Chambers, L., & Galizio. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior, 63*, 239-254.

Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior, 54*, 213-224.

Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63*, 225-238.

Pilgrim, C., & Galizio, M. (1996). Stimulus equivalence: A class of correlations, or a correlation of classes? In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus Class Formation in Humans and Animals* (pp. 173-195). Amsterdam, The Netherlands: Elsevier.

Pilgrim, C., Jackson, J., & Galizio, M. (2000). Acquisition of arbitrary conditional discriminations by young normally developing children. *Journal of the Experimental Analysis of Behavior, 73*, 177-193.

Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior, 71*, 195-214.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior, 50*, 145-162,

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (Zalophus californianus) is capable of forming equivalence classes. *The Psychological Record, 43*, 823-839.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA.: Authors Cooperative, Inc.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127-146.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior, 37*, 23-44.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5-22.

Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency, 77*, 556-566.

Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes, & L. J. Hayes (Eds.), *Understanding Verbal Relations* (pp. 29-42). Reno, NV: Context Press.

Wilkinson, K. M., Dube, W. V., & McIlvane, W. J. (1996). A crossdisciplinary perspective on studies of rapid word mapping in psycholinguistics and behavior analysis. *Developmental Review, 16*, 125-148.

Wilkinson, K. M., Dube, W. V., & McIlvane, W. J. (1998). Fast mapping and exclusion (emergent matching) in developmental language, behavior analysis, and animal cognition research. *The Psychological Record, 48,* 407-422.

Wilkinson, K. M., & McIlvane, W. J. (2001). Methods for studying symbolic behavior and category formation: Contributions of stimulus equivalence research. *Developmental Review, 21*, 355-374.

Wirth, O., & Chase, P. N. (2002). Stability of functional equivalence and stimulus equivalence: Effects of baseline reversals. *Journal of the Experimental Analysis of Behavior, 77*, 29-47.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125-144.