Available online at www.sciencedirect.com
SCIENCE DIRECT
Behavioural Processes 69 (2005) 281–293
BEHAVIOURAL PROCESSES
www.elsevier.com/locate/behavproc

Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols

Paul M. Smeets [a,*], Dermot Barnes-Holmes [b]

[a] *Department of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands*
[b] *National University, Maynooth, Ireland*

Received 6 June 2004; received in revised form 19 August 2004; accepted 6 December 2004

Abstract

The present study set out to assess if the different probabilities reported in the literature of obtaining equivalence after baseline training with MTO and OTM protocols could be attributed to individual differences and, if so, whether equivalence formation could be facilitated by using familiar stimuli as nodes. In Experiment 1, 16 preschool children were trained on four sets of 2-choice match-to-sample tasks, eight with a OTM protocol (A–B, A–C, A–D, A–E) and eight with a MTO protocol (B–A, C–A, D–A, E–A). For four OTM and four MTO children only abstract stimuli were used. The other four OTM children and four MTO children received the same training but with familiar stimuli as nodes. All children received tests for equivalence (first) and symmetry (second). In Experiment 2, eight children who served in Experiment 1 participated again, four who had passed the equivalence test, and four who had failed that test. All children received the same baseline training as in Experiment 1 but with the opposite type of nodes (abstract instead of familiar, and vice versa) and training protocol (MTO instead of OTM, and vice versa). The results showed that (a) the children's performances (training and testing) were not affected by the training protocol; (b) equivalence formation occurred more readily when being trained with all abstract stimuli than when familiar stimuli served as nodes; and (c) most children who passed or did not pass the equivalence test in Experiment 1 repeated their performance in Experiment 2, irrespective of the conditions that were used.

Keywords: Stimulus equivalence classes; Equivalence reversal; Training protocols; Stimulus familiarity; Preschool children

1. Introduction

Studies on stimulus equivalence typically start with a baseline training of multiple symbolic match-to-sample tasks with a set of common stimuli (nodes). The nodes can be used as samples (e.g., A–B, A–C), as comparisons (B–A, C–A), or as comparisons and as samples (B–A, A–C). In the literature, these three training structures are frequently referred to as the one-to-many protocol (OTM), the many-to-one protocol (MTO), and the linear series protocol (LS), respectively. Although OTM is by far the most frequently used protocol in

* Corresponding author. Tel.: +31 71 5273679; fax: +31 71 5273619.
E-mail address: smeets@fsw.leidenuniv.nl (P.M. Smeets).

0376-6357/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.beproc.2004.12.009

human research (Saunders and Green, 1999), all three protocols have been considered equally effective in producing novel match-to-sample performances indicative of stimulus equivalence (B–C, C–B).

A relatively small number of comparative studies, however, reported that the chances of generating equivalence differ among protocols (Arntzen and Holth, 1997, 2000; Barnes, 1994; Fields et al., 1999; Hove, 2003; Saunders et al., 1988, 1999; Spradlin and Saunders, 1986). Most of these studies and related reviews (Saunders and Green, 1999; Saunders et al., 1993), reported that (a) consistent with what has been reported in animal research (e.g., Hall et al., 1993; Urcuioli, 1996; Urcuioli et al., 1995), MTO is a more effective training protocol than OTM; and (b) LS is the least effective protocol. These discrepancies are remarkable because the original Sidman analysis of stimulus equivalence did not suggest that the equivalence outcomes should vary as a function of training structure, order, or direction (Sidman and Tailby, 1982; Saunders and Green, 1999).

Upon closer examination, however, the superiority of MTO over OTM appears far less impressive than has been suggested. Consider the following analyses for which we divided the aforementioned eight studies into two groups. The first group consists of five studies, in chronological order: Spradlin and Saunders (1986), Saunders et al. (1988, 1999), Fields et al. (1999), and Hove (2003). In each of these studies, equivalence and symmetry were assessed and all five of them have been repeatedly cited as evidence for the superiority of MTO over OTM. This conclusion, however, can be contested because no consideration was given to whether or not the participants passed the equivalence test before or after being exposed to the symmetry test. Yet, the efficacy of the training protocol can only be validly assessed if equivalence is tested before symmetry (e.g., train A–B, A–C; test C–B then B–A, C–A). If equivalence is measured together with or after symmetry, the C–B performances can be based on any demonstrated relations, that is, the trained relations (A–B, A–C: OTM), the tested symmetry relations (B–A, C–A: MTO), or a combination of trained and tested relations (B–A [tested], A–C [trained]: LS). Even though the test responses are not reinforced, the relations established during the symmetry probes may affect the formation of equivalence relations (Smeets et al., 1997).

In the Spradlin and Saunders (1986) study (Experiments 3 and 4), five adolescents with mental retardation were trained on four sets of 2-choice match-to-sample tasks. Two participants received a OTM protocol (A–B, A–C, A–D, A–E) and three a corresponding MTO protocol (B–A, C–A, D–A, E–A). None of the OTM participants evidenced equivalence before or after the presentation of the symmetry test. All three MTO participants demonstrated equivalence before symmetry was tested. This discrepancy (0/2 versus 3/3), however interesting, does not yield statistical significance and, as acknowledged by the authors, could be related to individual differences.

In the Saunders et al. (1988) study, six persons with mental retardation were trained on four sets of 2-choice match-to sample tasks, three with a OTM protocol (A–B, A–C, A–D, A–E) and three with a MTO protocol. Then they received a protracted series of equivalence tests alternated with symmetry tests and training sessions (Phase 1). Only one participant (MTO) evidenced equivalence test before symmetry was tested. Thereafter, one OTM participant and two more MTO participants showed equivalence. At that point, the participants were trained on a new series of match-to-sample tasks (Phase 2). Participants who had received OTM training in Phase 1, now received MTO training (A–E, F–E, G–E, H–E), while those who in Phase 1 had received MTO training now received OTM training (E–A, E–F G–E, E–H). The same participants who had evidenced equivalence in Phase 1 also showed equivalence in Phase 2. Given the authors' acknowledgement that these findings could be attributed to individual differences, it is unclear why this study has been repeatedly cited as evidence for the superiority of the MTO protocol (Arntzen and Holth, 2000; Fields et al., 1999; Hove, 2003; Saunders et al., 1999; Saunders and Green, 1999).

In the following study by Saunders et al. (1999), 11 normally developing preschool children received training on four sets of 2-choice match-to-sample tasks. Six children received a OTM protocol (A–B, A–C, A–D, A–E) and five a MTO protocol. Then they received a series of test sessions in which one part of the equivalence test was tested before and the other part of that test was presented after the symmetry test. Children who failed any or both tests were tested again (retest). Three MTO children evidenced equivalence (and symmetry). So did three OTM children, one of whom failed the symmetry test. During the retest, two more MTO children responded accurately on the equivalence trials while one OTM child, who initially had passed the equivalence test, now failed that test. At the end, therefore, equivalence was seen in only 2/6 OTM children and in all five MTO children. Although this difference yields statistical significance (Fisher test, $p = 0.045$)[1], this effect did not emerge until the retest, that is, after exposure to the symmetry test.

Fields et al. (1999) trained four groups of adults on multiple sets of 2-choice match-to-sample tasks. Two groups received training on four sets, one group with a OTM protocol (A–B, A–C, A–D, A–E) and one group with a MTO protocol. The other two groups were trained on six sets, one with a OTM protocol (A–B, A–C, A–D, A–E, A–F, A–G) and one with a MTO protocol. After demonstrating criterion performance, the participants received test blocks (no feedback) of equivalence trials mixed with symmetry and baseline trials. During the first block, emergence of seven-member classes was seen in 45% of the MTO participants and in 15% of the OTM participants. The percentages of MTO and OTM participants who showed five-member classes were about the same, that is, 55% and 60%, respectively. During the following blocks, for any given class size, the percentage of participants who showed the emergence of classes was only slightly higher after MTO training than after OTM training. Thus, except for a modest initial effect seen among participants who might have been trained to the limits of their capacities, the equivalence performances obtained with either training protocol were very similar.

Similar results, albeit with much smaller classes, were reported by Hove (2003). Twenty college students were trained on two sets of three-choice match-to-sample tasks, 10 with a OTM protocol (B–A, B–C) and 10 with a MTO protocol (A–B, C–B). Then they received a test block of equivalence trials mixed with symmetry and baseline trials and the performances on the first and second half of the symmetry and equivalence test were compared. Although the performances of the MTO participants were superior to those of their OTM counterparts during both halves of the equivalence test, the between-group difference during the second half was no longer statistically significant. Overall, the results of these studies permit the conclusion that the discrepancies between OTM and MTO (a) are relatively small or can be qualified as quickly evaporating initial effects; (b) cannot be unequivocally related to the training protocol; and (c) may be accounted for by individual differences.

Similar inconclusive findings were obtained in the second group of studies, in chronological order: Barnes (1994), and Arntzen and Holth (1997, 2000). These studies used tests that consisted only of equivalence trials (i.e., no symmetry trials and no baseline trials). Assuming that the trained baseline relations remained in tact during testing, the equivalence performances can be directly related to the training protocol. In Barnes' (1994) study, 20 adults received training on two sets of two-choice match-to-sample tasks, 10 with a MTO protocol (A–B, C–B) and 10 with a OTM protocol, followed by one block of equivalence test trials (C–A). Nine of the MTO participants and only four of the OTM participants passed the C–A test (Fisher test, $p = 0.027$).[1]

Arntzen and Holth (1997) trained adults on two sets of three-choice match-to-sample tasks, 10 with a MTO protocol (A–B, C–B) and 10 with a OTM protocol, and compared the individual performances during the first and second halves of a single block of equivalence trials (C–A). All OTM participants evidenced equivalence during both test halves. Of the MTO participants, five evidenced equivalence during the first half (Fisher test, $p < 0.05$) and seven during the second half (not significant). Although this finding has been reported as a demonstration of OTM superiority over MTO (e.g., Arntzen and Holth, 1997, 2000; Hove, 2003; Saunders and Green, 1999), the differential effect was very short lived.

In the Arntzen and Holth (2000) study, 23 adults participated. Each participant was trained on multiple sets of match-to-sample tasks, one set with MTO and one set with OTM (Experiment 1), or two sets with MTO and two sets with OTM (Experiments 2 and 3). The numbers of designated classes (2 or 3) and class members (3 or 4) varied over experiments. Across all four experiments, OTM and MTO produced accurate responding on 48/53 (91%) and 44/53 (83%) of the tested equivalence relations, respectively. Collectively, therefore, these studies did not provide convincing evidence for the superior efficacy of one training protocol over another. The different outcomes of these three

[1] This statistical test was carried out by the authors of the current article.

studies could have resulted from unbalanced assignment of participants.

The present study was modeled after the Saunders et al. (1999) study because of theoretical considerations (see Discussion section) predicting that MTO superiority is more likely to emerge with young normal children (and persons with mental retardation) than with adults. Would the current study support the above analysis and show that OTM and MTO produce equivalence with equal probability? If not, could the efficacy of either training protocol be improved by using familiar and easy to name stimuli as nodes (Arntzen, 2004; Arntzen and Holth, 1998)? In Experiment 1, we assessed the differential effects of MTO and OTM training protocols with abstract and with familiar stimuli as nodes. Experiment 2 was designed to assess if the children, who passed or failed the equivalence test in Experiment 1, would show similar equivalence performances after being trained with the other protocol. If so, this finding would support the notion that the differential test results, if any, which were obtained with either protocol probably resulted from individual differences.

2. Experiment 1

This experiment was an exact replication of Experiment 1 of the Saunders et al. (1999) study except that: (a) a table-top setting was used; (b) abstract and familiar stimuli were used; and (c) equivalence was tested before symmetry.

2.1. Method

2.1.1. Children

Eight boys and eight girls participated. Their ages ranged from 5 years and 0 months to 6 years and 0 months (mean: 5 years, 4 months). The gender and age of each child are listed, together with the training conditions and training results, in Table 1. The children were recruited through school contacts and participated with their parents' approval on a voluntary basis. None of the children had participated in experimental studies before. The population was divided into four groups of four children each. Each group received a different baseline training (see below).

2.1.2. Sessions, setting, and adults

Sessions were conducted individually in a quiet room of the school building, once a day, and lasted 8–20 min (mean: 12). Across children, the numbers of sessions varied from 7 to 18 (mean: 12).

Two adult females served as experimenters, Experimenters 1 and 2. Each experimenter carried out the procedures with two children (50%) of each condition (see below). The experimenter and child were seated at

Table 1
Children, training conditions, experimenters, and numbers of training blocks in Phases 1 and 2 of Experiment 1

Children	Gender	Age	Training	Experimenter	Training blocks	
					Phase 1	Phase 2
1	F	5; 3	OTM-Abs	1	2	10
2	F	5; 3	OTM-Abs	2	2	31
3	M	5; 6	OTM-Abs	1	1	12
4	F	6; 0	OTM-Abs	2	1	8
5	M	5; 2	OTM-Fam	1	1	8
6	F	5; 2	OTM-Fam	2	2	14
7	M	5; 4	OTM-Fam	1	1	11
8	F	5; 10	OTM-Fam	2	1	9
9	M	5; 1	MTO-Abs	1	1	25
10	F	5; 4	MTO-Abs	2	1	7
11	M	5; 6	MTO-Abs	2	1	21
12	M	5; 9	MTO-Abs	1	1	10
13	M	5; 0	MTO-Fam	2	1	8
14	F	5; 0	MTO-Fam	1	1	20
15	M	5; 6	MTO-Fam	2	2	10
16	F	5; 8	MTO-Fam	1	1	12

the same table facing one another. The experimenters had received extensive training on the correct execution of the training and testing procedures with special emphasis on the prevention of cues (e.g., facial expression, eye darting, timing of feedback) that could influence the children's responses. During training, the experimenter looked at the child's face when giving instructions and delivering programmed consequences. During the remainder of these trials (i.e., when presenting stimulus materials and while the child responded), the experimenter gazed at a fixed location on the table. Precautions were taken to prevent the children from observing the experimenter's recordings on the data sheets.

Two other adults served as reliability observers, one at a time. The observer was in the same room but situated such (i.e., behind and slightly to the left or right from the child) that she could clearly observe the children's responses but not the experimenter's data sheet.

2.1.3. Tasks, stimuli, and materials

Identity and symbolic match-to-sample tasks were used. The stimuli (3 cm × 3 cm) consisted of four geometric forms, 10 abstract configurations (same as in Saunders et al., 1999), and two familiar color-form compounds (yellow colored happy face and red colored heart). The geometric forms were only used for pretraining. The other stimuli (see Fig. 1) were used for baseline training and equivalence testing. For convenient reference, the stimuli are identified by alphanumeric codes (e.g., A1, C2). The numerals refer to the classes and the letters to the members of these classes. The children never saw these codes.

The stimuli were presented on laminated white cards (21.0 cm × 14.8 cm). Each card showed two comparisons (8.5 cm apart) and a sample centered 4.5 cm below. Additional materials included a cup with 50 beads and a standing glass tube showing a mark. Filling the tube to the mark required 50 beads.

2.1.4. Trials, contingencies, and feedback

Each trial started with the experimenter presenting a stimulus card. The responses were recorded correct, incorrect, or invalid. Invalid responses, which rarely occurred, were recorded when a child pointed without looking at the card or pointed to both comparisons. During training, correct responses were followed by positive feedback ("Good. Take a bead") and incorrect responses by negative feedback ("Wrong. No bead").

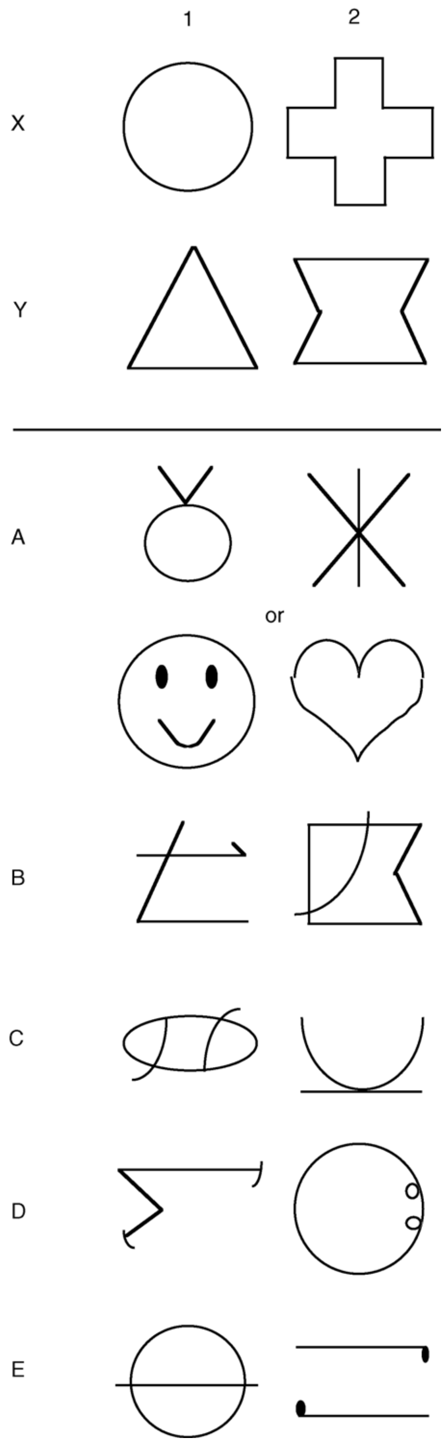

Fig. 1. Stimuli that were used during pretraining (X, Y) and experimental stimuli (A–E).

Invalid responses were followed by a correction procedure (e.g., "You can point to only *one* picture"). During testing, the procedures were the same except that correct and incorrect responses were followed by no programmed consequences other than the presentation of the next trial.

2.1.5. *Experimental sequence*

The program consisted three phases: pretraining, baseline training, and testing equivalence and symmetry. The pretraining produced identity-matching performances. These relatively easy tasks served to familiarize the children with the procedures and to prepare them for the more difficult symbolic matching tasks that were used in the following two phases. In the baseline phase, four sets of two-choice symbolic matching tasks were trained: A–B, A–C, A–D, A–E (OTM) or B–A, C–A, D–A, E–A (MTO). Eight children received the OTM protocol and eight the MTO protocol. For four OTM children (1–4) and four MTO children (9–12), only abstract stimuli were used. These groups are referred to as OTM-Abstract and MTO-Abstract. The other four OTM children (5–8) and four MTO children (13–16) received the same training but with familiar stimuli as nodes (happy face and heart). These groups are referred to as OTM-Familiar and MTO-Familiar (see Table 1). Finally, all children received tests for equivalence (first) and symmetry (second).

2.1.5.1. Phase 1: pretraining. Two sets of identity matching tasks were trained, X–X and Y–Y. Blocks of 16 trials were used, 8 X–X trials quasi-randomly mixed with 8 Y–Y trials. Immediately before starting the first trial, the experimenter said, "I am going to show you three pictures. First look at this picture (experimenter pointed to sample). Now look at these pictures (experimenter pointed to each comparison). One is right, one is wrong. Can you show me the right picture?" All responses were followed by feedback. Children who responded correctly on at least 15 trials of a block (94%) proceeded to Phase 2.

2.1.5.2. Phase 2: baseline training. Four sets of symbolic match-to-sample tasks were trained. The training occurred in a stepwise fashion. The A–B or B–A relations were trained in Step 1. Blocks of 16 trials were used. All children were instructed during the first two trials. When dealing with OTM-Abstract children, the experimenter said during the first trial, "This (A1) is an apple, this (B1) a nose, and that (B2) a flag. Point to nose", and during the second trial, "This (A2) is a star, this (B1) a nose, and that (B2) a flag. Point to flag." When dealing with OTM-Familiar children, the experimenter said, "This (A1) is a smiley face, this (B1) a nose, and that (B2) a flag. Point to nose" (Trial 1), or "This (A2) is a heart, this (B1) a nose, and that (B2) a flag. Point to flag" (Trial 2). The MTO children were given the same instructions as their OTM counterparts but with the stimulus names reversed in each sentence. These instructions were not repeated in any following trials. Children, who responded correctly on at least 15 trials of a block, proceeded to the following step. If at any given training trial of this step or of any following components of the program the number of accumulated beads reached the mark on the tube, the experimenter interrupted the session, allowed the child to exchange the beads for a colored picture, and advanced to the next trial.

From Step 2 on, new match-to-sample tasks were mixed with previously trained tasks (review trials). In Step 2, the A–C or C–A tasks were trained. Each block consisted of 22 trials, 16 A–C or C–A trials quasi-randomly mixed with 6 A–B or B–A trials. The procedures were the same as in Step 1 but without instructions. Children who responded correctly on 20 (91%) or more trials of a block while making no more than one error on a review trial, advanced to the following step. Children who did not show criterion performance on the A–C or C–A trials while demonstrating criterion performance on the review trials, received Step 2 again. Children who failed to demonstrate criterion performance on the A–C or C–A trials and on the review trials, returned to Step 1 and, after demonstrating criterion performance, received Step 2 again. The procedures for Steps 3 and 4 were the same. In Step 3, each block consisted of 16 A–D or D–A trials quasi-randomly mixed with 3 A–B or B–A trials and 3 A–C or C–A trials. In Step 4, each block consisted of 16 A–E or E–A trials, mixed with 2 A–B or B–A, 2 A–C or C–A, and 2 A–D or D–A trials.

Steps 5 and 6 involved proportionate mixed training. In Step 5, 16 trials were used: four A–B trials quasi-randomly mixed with four A–C, four A–D, and four A–E trials (OTM), or four B–A trials, quasi-randomly mixed with four C–A, four D–A, and four E–A trials (MTO). Children who responded correctly on at least 15 trials advanced to Step 6. Children who did not meet this criterion received this step again unless they made multiple errors on the same set (e.g., A–C). In that case, training of that set (Step 2) was repeated until criterion performance was demonstrated, at which point the child returned to Step 5. Step 6 was the same as Step 5 except that, for each set of tasks, the reinforcement density was reduced from 100 to 75%. Each block consisted of 12 reinforced and four nonreinforced trials. The nonreinforced trials were always interspersed between reinforced trials. Starting the first trial, the experimenter informed each child that, “Now I would like to see whether you can play the game without me telling you that you are right or wrong. Therefore, I will sometimes say nothing. Do your best.” Following the completion of each block, the children received additional four beads irrespective of their performance on the nonreinforced trials. Children who responded correctly on at least 14 trials advanced, without any introduction, to Phase 3. Those who made less than 14 correct responses, returned to Step 5 (mixed training, 100% reinforcement) before Step 6 was presented again.

2.1.5.3. Phase 3: testing equivalence and symmetry. The procedures were the same as in Step 6 (12 reinforced baseline trials mixed with four nonreinforced test trials), except that the four test trials measured equivalence (B–C, C–B; B–D, D–B; B–E, E–B; C–D, D–C; C–E, E–C; D–E, E–D) or symmetry (OTM: B–A, C–A, D–A, E–A; MTO: A–B, A–C, A–D, A–E).

Equivalence was tested first. All 24 trials were presented once. The testing took place in six consecutive blocks (spread over several sessions), four trials per block, two from Class 1 and two from Class 2. Then symmetry was tested the same way. Testing the eight relations required two blocks. The children proceeded from one block to another provided that they responded correctly on at least 11/12 reinforced baseline trials. If they failed that criterion, they returned to Step 5 of Phase 2 (mixed baseline training) until criterion was reached, at which point, the test block was presented again. For example, if during the second block of the equivalence test a child responded correctly on only 9/12 baseline trials, that child returned to the mixed baseline training before the second equivalence test block was presented again. Thus, failures to demonstrate equivalence or symmetry could not be related to deterioration of baseline skills. Equivalence and symmetry were assumed if a child responded correctly on at least 20/24 equivalence trials (83%) and on at least 7/8 symmetry trials (87%). Children whose testing performances met both these criteria were not tested again. Children who failed the equivalence or symmetry test received that test again (retest).

2.1.6. Interobserver reliability

The experiment consisted of 6598 training trials and 844 test trials. Twelve thousand training trials (18%) and 308 test trials (37%) were monitored. The experimenters and observers agreed on all but two training trials and one test trial.

2.2. Results

One MTO child failed to learn the B–A relations and was replaced by another child. The results of the pre- and baseline training (Phases 1 and 2) are shown in Table 1. All children learned the identity matching tasks in one or two blocks and required 7–31 blocks to complete the baseline training. All conditions required about the same mean numbers of training blocks, OTM-Abstract: 15.3, MTO-Abstract: 15.5, OTM-Familiar: 10.5, and MTO-Familiar: 12.5. Although the children learned the baseline tasks somewhat faster during the Familiar (mean = 11.5 blocks) than during the Abstract conditions (mean = 15.5 blocks), a *t*-test revealed that this difference was not statistically significant (9.7, 1.15, $p = 0.28$). When the OTM protocol was used, Experimenter 1 needed fewer blocks for training children to complete the baseline phase (mean = 10.3) than Experimenter 2 (mean = 15.5). When the MTO protocol was used, Experimenter 1 needed more blocks (mean = 16.8) than Experimenter 2 (mean = 11.5). Although these findings suggest an interaction effect, an ANOVA indicated that it was not statistically significant ($F[1,12] = 2.13$, $p = 0.17$).

Table 2 shows the numbers of correct test responses of each child. Three analyses were carried out. First, we used the same criteria as Saunders et al. (1999): 20/24 equivalence trials correct (83%) and 7/8 symmetry trials correct (87%). Based on these criteria, all children but one (4) passed the symmetry test. Six children passed the equivalence test during the Abstract condition, five before (two OTM children and three MTO children), and one (OTM) after the symmetry

Table 2
Numbers of correct responses during equivalence and symmetry tests in Experiment 1

Children	Test		Retest		Children	Test		Retest	
	Equivalence	Symmetry	Equivalence	Symmetry		Equivalence	Symmetry	Equivalence	Symmetry
OTM-Abstract					OTM-Familiar				
1	19	8	23		5	22	8		
2	23	8			6	16	7	10	
3	20	8			7	19	8	24	
4	9	4	13	4	8	17	8	22	
MTO-Abstract					MTO-Familiar				
9	20	8			13	19	8	24	
10	20	8			14	14	8	15	
11	12	7	17		15	24	8		
12	23	8			16	11	7	11	

test. During the Familiar condition, five children passed the equivalence test, two before (one OTM child and one MTO child) and three after the symmetry test (one OTM children and one MTO child). Thus, there were no obvious differences between conditions (OTM versus MTO, Abstract versus Familiar). Although the abstract conditions seemed to produce equivalence more readily before symmetry than the familiar conditions, this difference (5 versus 2) was not statistically significant (Fisher test, $p = 0.13$) There were also no experimenter effects. Six children evidenced equivalence in the presence of Experimenter 1 (four before and two after symmetry) and five in the presence of Experimenter 2 (three before and two after symmetry).

Second, we analyzed the data according to a more stringent criterion for assessing equivalence. Although the criterion of 83% or more correct responses is consistent with Sidman's (1987) recommendations for assessing two 3-term equivalence classes, this criterion may not be stringent enough for classes of more than three members. Consider Child 9 who passed the equivalence test on the basis of 20 correct responses. He responded correctly on all trials of three relations (B ↔ C, C ↔ E, D ↔ E), but on chance level (two or three trials correct) on each of the other three relations (B ↔ D, B ↔ E, C ↔ D). This performance is not essentially different from that of Child 11 who failed the equivalence retest with a score of 17 correct responses. This child also responded correctly on all trials of three relations (B ↔ E, C ↔ E, D ↔ E), and at chance level (one or three trials correct) on each of the other three relations (B ↔ C, B ↔ D, C ↔ D). Given that neither one of these two children demonstrated evidence for the emergence of two 5-member classes, the 83% percent correct criterion for measuring equivalence classes of more than three members clearly is inadequate. We therefore analyzed the equivalence test data again but with the following, more stringent, criterion: at least 22/24 trials correct (92%) with no more than one error on any specific stimulus relation (e.g., B ↔ C). The retest data were not taken into account because only some children had been retested. Using this criterion, only three children evidenced 5-term equivalence classes, Child 2 (OTM-Abstract), Child 12 (MTO-Abstract), and Child 15 (MTO-Familiar). Again, no differential effects from training structure (OTM versus MTO) or stimulus familiarity (Abstract versus Familiar) were found.

Finally, we compared the mean number of correct responses of each group on the equivalence test. Although such an across-subjects approach is inconsistent with the single-subject approach in human equivalence research, one could argue that the mean number of correct responses gives some indication about the efficacy of each condition. Again, the results did not point in any direction. The mean numbers of correct responses in OTM-Abstract, OTM-Familiar, MTO-Abstract, and MTO-Familiar were almost the same: 17.8 (range: 9–23), 18.5 (range: 16–22), 18.8 (range: 12–23), and 17.0 (range: 11–24), respectively. The results of the retests were very similar: OTM-Abstract (mean: 17.5, range: 13–22), OTM-Familiar (mean: 18.7, range: 10–24), MTO-Abstract (mean: 17.0, no range), and MTO-Familiar (mean: 16.7, range: 11–24). In essence, there was no statistically supported evidence that the training protocols (MTO, OTM) or nature of the nodal stimuli (Abstract or Familiar) affected equivalence formation. The only apparent differences were those between individual children.

3. Experiment 2

This experiment was designed to identify the individual differences that contributed to the equivalence performances in Experiment 1. Two possibilities were considered. First, the training protocols (and nature of the nodal stimuli) were irrelevant—the results simply identified children with different aptitudes for deriving class-consistent match-to-sample tasks. If correct, most children probably would show similar performances (i.e., pass or fail the equivalence test) after being trained on a new set of baseline tasks with a different protocol (and with different stimuli as nodes). Second, the training protocols were relevant but had different implications across individual children. Some children benefited more from a MTO protocol, others more from a OTM protocol. Thus, children who evidenced equivalence in Experiment 1 after MTO training might not have done so after OTM training. Likewise, children who failed the equivalence test after MTO training might have passed that test after OTM training. If correct, children who, in Experiment 1, passed the equivalence test with one protocol (e.g., MTO) probably would not pass that test after receiving novel baseline training with the other protocol (OTM). Alternatively, children who failed the equivalence test in Experiment 1 after being trained with one protocol (e.g., MTO) might pass that test after being trained with a different, possibly more effective, protocol (OTM).

In Experiment 2, children who participated in Experiment 1 received the same baseline training again but with the other training protocol and with different A stimuli as nodes. Children, who in Experiment 1 had received OTM training, now received MTO training, and vice versa. Likewise, children who, in Experiment 1, had been trained with abstract A stimuli, were now trained with familiar A stimuli, and vice versa. Furthermore, the trained relations between the A and B stimuli were class-inconsistent with those between the A stimuli and the C, D, and E stimuli. This training should lead to the emergence of two 5-term equivalence classes different from those in Experiment 1. For children who passed the equivalence test in Experiment 1, the implication was that, in order to pass the equivalence test again, they had to reverse three of the six tested relations ($B \leftrightarrow C$, $B \leftrightarrow D$, $B \leftrightarrow E$) while maintaining the other relations ($C \leftrightarrow D$, $C \leftrightarrow E$, $D \leftrightarrow E$). For children who failed the equivalence test in Experiment 1, Experiment 2 provided them with a second opportunity to show equivalence, now perhaps under more favorable training conditions.

3.1. Method

Eight children from Experiment 1 participated again, two from each condition. The children were divided into two groups of four children each. Group 1 consisted of Children 2, 5, 12, and 15. These children had shown equivalence before symmetry in Experiment 1, three (2, 12, 15) according to the 92% criterion and one (5) according to the 83% criterion (Child 5 also met the 92% criterion but, unlike the other three children, had made two errors on the same [$B \leftrightarrow D$] relation). Group 2 consisted of Children 4, 6, 11, and 16. These children had failed the equivalence test (before and after symmetry) in Experiment 1. All children received the same baseline training as in Experiment 1 but with the following modifications.

First, the B, C, D, and E stimuli were conditionally related to different A stimuli. Abstract A stimuli were used with children who, in Experiment 1, had been trained with familiar A stimuli (5, 6, 15, 16), while familiar A stimuli were used with children who, in Experiment 1, were trained with abstract A stimuli (2, 4, 11, 12). Second, the A stimuli were conditionally related to B stimuli of another class (e.g., A1–B2, A2–B1) than the C, D, and E stimuli (e.g., A1–C1, A2–C2; A1–D1, A2–D2; A1–E1, A2–E2). This training was carried out with a MTO protocol for children who, in Experiment 1, had been trained with a OTM protocol (2, 4, 5, 6), and with a OTM protocol for children who, in Experiment 1, were trained with a MTO protocol (11, 12, 15, 16). The test procedures were identical to those used in Experiment 1.

The experiment started 2 months after the completion of the data collection of Experiment 1 and was conducted by a new experimenter (male) in collaboration with two reliability observers. Because the children were expected to recognize some of the tasks, notably those with unchanged stimulus configurations

Table 3
Number of required training blocks and correct test responses in Experiment 2

Children	Conditions	Training blocks	Test		Retest	
			Equivalence	Symmetry	Equivalence	Symmetry
Group 1						
2	MTO-Fam	16	11	8	12	
5	MTO-Abs	12	23	7		
12	OTM-Fam	11	19	8	24	
15	OTM-Abs	16	22	8		
Group 2						
4	MTO-Fam	12	17	7	13	
6	MTO-Abs	15	10	8	15	
11	OTM-Fam	10	16	7	17	
16	OTM-Abs	14	12	8	13	

(e.g., B–C), the experimenter was informed that these children had participated in a similar experiment before but given no information on the children's conditions or performances.

Reliability checks were made on 760 training trials (24%) and on 178 test trials (41%). The experimenter and observers agreed on all but two training trials and two test trials.

3.2. Results

The training and test results are listed in Table 3. All children completed the baseline training. The mean numbers of required trial blocks for the OTM-Abstract, OTM-Familiar, MTO-Abstract, and MTO-Familiar conditions were 15.0, 10.5, 13.5 and 14.0, respectively. Group 1 required a mean of 13.8 trials, Group 2 a mean of 12.8 trials.

All children passed the symmetry test. Three children of Group 1 passed the equivalence test, two before (Child 15/OTM-Abstract, Child 5/MTO-Abstract) and one after symmetry (Child 12/OTM-Familiar). The fourth child of this group (Child 2/MTO-Familiar) reversed all equivalence relations, that is, not only the B–C, B–D, and B–E relations, as would be expected from the baseline training, but also the C–D, D–E, and C–E relations. None of the children of Group 2 passed the equivalence test, not before and not after symmetry. Although this difference (3/4 versus 0/4) was not statistically significant (Fisher test, $p = 0.071$), this finding is highly consistent with that (4/4 versus 0/2; Fisher test, $p = 0.066$) obtained by Saunders et al. (1988) under similar conditions (see Introduction section). Collectively, these findings (7/8 versus 0/6; Fisher test, $p = 0.002$) strongly suggest that some participants more readily demonstrate equivalence than others irrespective of the training protocol and the nodal stimuli that are used.

4. Discussion

The two training protocols did not affect the children's match-to-sample performances, not during training and not during testing. In Experiment 1, OTM training required a mean of 13 trial blocks, produced symmetry in seven children and equivalence in six children, three before and three after symmetry. MTO training required a mean of 14 trial blocks, produced symmetry in all eight children and equivalence in five children, four before and one after symmetry. In Experiment 2, eight children who had served in Experiment 1 served again, four who had passed and four who had failed the equivalence test. All children received training with the other protocol (e.g., MTO instead of OTM) and with different stimuli as nodes (e.g., abstract instead of familiar stimuli). The OTM and MTO protocols required about the same numbers of training blocks to complete the baseline training (same means as in Experiment 1). Both protocols produced symmetry in all children, produced equivalence in most children who also evidenced equivalence in Experiment 1, and did not produce equivalence in children who failed in Experiment 1. These findings are consistent with our aforementioned analysis of existing research indicating that the probabilities of the MTO and OTM protocols generating equivalence are the same. Furthermore, they support the repeatedly mentioned but systematically dismissed notion (Arntzen and Holth, 1997; Fields et al., 1999; Saunders et al., 1988; Spradlin and Saunders, 1986) that the performance discrepancies associated with these two protocols, if any, result from differences in the learning repertoires of the participants.

By implication, the present findings are inconsistent with the discrimination analysis of equivalence class formation (Saunders and Green, 1999; Saunders et al., 1999). This could be a problem if most of the existing protocol research was, in fact, consistent with this analysis, but this not the case. The discrimination analysis states that equivalence formation requires that the trained conditional discriminations require the participants to discriminate every stimulus from every other stimulus. With a MTO protocol, participants receive a more complete discrimination training than with a OTM protocol. For example, when training two 3-member equivalence classes, OTM participants need only make a single successive discrimination between the two samples (A1 versus A2), and simultaneous discrimination of each comparison from its paired counterpart (B1 versus B2, C1 versus C2). By contrast, MTO participants have to make a successive discrimination of each sample from every sample of the other class (B1 versus B2, B1 versus C2, C1 versus B2, C1 versus C2), and a single simultaneous discrimination between the comparisons (A1 versus A2). This analysis led to a set of predictions. First, the speed of learning the same number of conditional discriminations during MTO should be slower than with OTM. This prediction, however, is not generally supported. Some studies have reported that MTO protocols require more training units (trials or blocks of trials) than OTM protocols (Saunders et al., 1988, 1999). In other studies, the higher number of MTO training trials could be related to an order effect (Arntzen and Holth, 2000), both protocols required the same total numbers of trials (Arntzen and Holth, 1997; current study), or MTO required less trials than OTM (Hove, 2003).

Another prediction from the discrimination analysis is that MTO superiority should be more pronounced in children and persons with mental retardation than in adults and also more pronounced with larger classes than with smaller classes. As it stands, the evidence suggests the contrary. The strongest evidence in support for MTO superiority over OTM stems from the studies by Barnes (1994) and Hove (2003) in which adults were trained on 3-member classes. In both studies in which preschoolers were trained on larger 5-member classes (Saunders et al., 1999; the current study), the probability of the two protocols generating equivalence before symmetry was tested was the same.

Finally, there is little evidence for the assumption that the probability of equivalence is related to the proportion of trained simple discriminations. Apart from the aforementioned studies in which no differential effects between the two protocols were found, studies by Leader et al. (1996, 2000) and Smeets et al. (1997) have shown that equivalence class formation may come about without discrimination training. In each of these studies, adults and/or preschool children simply viewed series of pairs of successively presented stimuli (e.g., A1 then B1, A2 then C2, A1 then C1, A2 then B2). This respondent-type training was sufficient to induce equivalence (e.g., B1–C1, C2–B2) in the context of match-to-sample tasks, irrespective of the training protocol (MTO, OTM, LS) that was used. Therefore, the fact that, in one of these studies (Smeets et al., 1997), the MTO participants evidenced equivalence sooner than the OTM participants, cannot be seen as evidence in support of the discrimination analysis account, because there was no discrimination training. Furthermore, Hove (2003) reported that of the eight OTM participants who failed the equivalence test, six grouped the stimuli according to the designated classes during a subsequent sorting test. If the OTM training did not establish all discriminations for the emergence of stimulus equivalence, the class-consistent sorting performances would be difficult to understand. Clearly, the discrimination analysis of equivalence is not tenable.

Across both experiments, equivalence before symmetry was seen nine times, seven times after baseline training with all abstract stimuli, and twice after baseline training with pictorial nodes (Fisher test, $p = 0.011$). Although this discrepancy diminished after symmetry was tested, this finding (a) suggests that, initially at least, the pictorial nodes interfered with equivalence formation; and (b) is consistent with Sidman's position (1994) that familiar stimuli may confound the relations resulting from the explicitly arranged experimental contingencies with relations established during the participants' pre-experimental histories. Yet, our current findings are inconsistent with those reported by

Holth and Arntzen (1998). In that study, adults were trained on three 3-member classes with a LS protocol (A–B, B–C), some with all abstract stimuli (Greek letters), others with abstract stimuli and with pictures serving as A, B, or C stimuli. Most participants did not evidence equivalence (C–A), unless the A and C stimuli or the B stimuli (nodes) were pictures. Although it might be tempting to attribute this discrepancy between these two studies to different populations (adults versus children), it should be noted that in the literature reported differential effects of easy versus difficult-to-pronounce, meaningful versus not meaningful, and familiar versus abstract stimuli are far from coherent (e.g., Arntzen, 2004; Liddy et al., 2000; Mandell and Sheen, 1994).

Finally, the present findings provide additional evidence for the reversibility of equivalence relations. Equivalence reversal is important because the trained relations are held to be the basis for equivalence-class performances. The evidence on the reversibility of equivalence, however, is divided. Some studies reported that equivalence is difficult to reverse, far more so than symmetry, thereby implying that the equivalence relations become independent from the trained relations from which they emerged (Pilgrim et al., 1995; Pilgrim and Galizio, 1990, 1995; Roche et al., 1997; Spradlin et al., 1973). Other studies reported equivalence reversal in adults, children, and persons with mental retardation. In some studies, this was achieved by switching the class-specific reinforcers (Dube and McIlvane, 1995; Dube et al., 1987, 1989), in other studies by reversing the contingencies of nonspecific reinforcers (Garotti et al., 2000; Saunders et al., 1999, Experiment 2; Smeets et al., 2003). In the present study, this was achieved by establishing a relation between one member (B) of each class and a new stimulus (A) that was inconsistent with the relation between the new stimulus and the other class members (A–C, A–D, A–E). Clearly, equivalence reversal can be achieved through various procedures.

In sum, the present findings indicate that (a) both protocols (MTO, OTM) were equally effective in establishing trained and derived match-to-sample performances; (b) equivalence occurred more readily when all abstract stimuli were used than when familiar stimuli were used as nodes; and (c) the only apparent differences were those among individual children. Some children evidenced equivalence while others did not, irrespective of the training protocol and the nature of the nodal stimuli that were used.

Although the current study indicates that both the MTO and OTM protocols are equally effective in generating equivalence with young normally developing children, as noted in the Introduction research with non-humans has previously demonstrated a relatively clear MTO superiority effect (Hall et al., 1993, Urcuioli, 1996; Urcuioli et al., 1995). In fact, whereas research with humans has produced inconsistent findings when comparing the effects of OTM versus MTO, animal studies have been relatively straightforward, the latter protocol is better. Thus, there could still be some basis for the MTO superiority effect, but it must be acknowledged that it remains extremely difficult to capture with humans.

References

Arntzen, E., 2004. Probability of equivalence formation: familiar stimuli and training sequence. Psychol. Rec. 54, 275–291.

Arntzen, E., Holth, P., 1997. Probability of stimulus equivalence as a function of training design. Psychol. Rec. 47 (2), 309–320.

Arntzen, E., Holth, P., 2000. Equivalence outcome in single subjects as a function of training structure. Psychol. Rec. 50, 603–628.

Barnes, D., 1994. Stimulus equivalence and relational frame theory. Psychol. Rec. 44, 91–124.

Dube, W.V., McIlvane, W.J., 1995. Stimulus-reinforcer relations and emergent matching to sample. Psychol. Rec. 45, 591–612.

Dube, W.V., McIlvane, W.J., Mackay, H.A., Stoddard, L.T., 1987. Stimulus membership established via stimulus-reinforcer relations. J. Exp. Anal. Behav. 47, 159–175.

Dube, W.V., McIlvane, W.J., Maguire, R.W., Mackay, H.A., Stoddard, L.T., 1989. Stimulus class formation and stimulus-reinforcer relations. J. Exp. Anal. Behav. 51, 65–76.

Fields, L., Hobbie-Reeve, S.A., Adams, B.J., Reeve, K.T., 1999. Effects of training directionality and class size on equivalence class formation by adults. Psychol. Rec. 49, 703–724.

Garotti, M., De Souza, D.G., De Rose, J.C., Molina, R.C., Gil, M.A., 2000. Reorganization of equivalence classes after reversal of baseline relations. Psychol. Rec. 50, 35–48.

Hall, G., Ray, E., Bonardi, C., 1993. Acquired equivalence between cues trained with a common antecedent. J. Exp. Psychol.: Anim. Behav. Proc. 19, 391–399.

Holth, P., Arntzen, E., 1998. Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. Psychol. Rec. 48, 81–110.

Hove, O., 2003. Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. Psychol. Rec. 53, 617–634.

Leader, G., Barnes, D., Smeets, P.M., 1996. Establishing equivalence relations using a respondent-type training procedure. Psychol. Rec. 46, 685–706.

Leader, G., Barnes-Holmes, D., Smeets, P.M., 2000. Establishing equivalence relations using a respondent-type training procedure III. Psychol. Rec. 50, 63–78.

Liddy, F., Barnes-Holmes, D., Hampson, P.J., 2000. The effect of stimulus meaningfulness on the formation of equivalence classes. Eur. J. Behav. Anal. 1, 71–87.

Mandell, C., Sheen, V., 1994. Equivalence class formation as a function of the pronounceability of the sample stimulus. Behav. Proc. 32, 29–46.

Pilgrim, C., Chambers, L., Galizio, M., 1995. Reversal of baseline relations and stimulus equivalence. II. Children. J. Exp. Anal. Behav. 63, 239–254.

Pilgrim, C., Galizio, M., 1990. Relations between baseline contingencies and equivalence probe performances. J. Exp. Anal. Behav. 54, 213–224.

Pilgrim, C., Galizio, M., 1995. Reversal of baseline relations and stimulus equivalence. I. Adults. J. Exp. Anal. Behav. 63, 225–238.

Roche, B., Barnes, D., Smeets, P.M., 1997. Incongruous stimulus pairing contingencies and conditional discrimination training: effects on relational responding. J. Exp. Anal. Behav. 68, 143–160.

Saunders, R.R., Drake, K.M., Spradlin, J.E., 1999. Equivalence class establishment, expansion, and modification in preschool children. J. Exp. Anal. Behav. 71, 195–214.

Saunders, R.R., Green, G., 1999. A discrimination analysis of training-structure effects on stimulus equivalence outcomes. J. Exp. Anal. Behav. 72, 117–137.

Saunders, R.R., Wachter, J., Spradlin, J.E., 1988. Establishing auditory control over eight-member equivalence class via conditional discrimination procedures. J. Exp. Anal. Behav. 49, 95–115.

Saunders, K.J., Saunders, R.R., Williams, D.C., Spradlin, J.E., 1993. An interaction of instructions and training design on stimulus class formation: extending the analysis of equivalence. Psychol. Rec. 43, 725–744.

Sidman, M., 1987. Two choices are not enough. Behav. Anal. 22, 11–18.

Sidman, M., 1994. Equivalence Relations and Behavior: A Research Story. Authors'Cooperative, Boston, MA.

Sidman, M., Tailby, W., 1982. Conditional discriminations versus matching-to-sample: an expansion of the testing paradigm. J. Exp. Anal. Behav. 37, 5–22.

Smeets, P.M., Barnes-Holmes, Y., Akpinar, D., Barnes-Holmes, D., 2003. Reversal of equivalence relations. Psychol. Rec. 53, 91–119.

Smeets, P.M., Leader, G., Barnes, D., 1997. Establishing stimulus classes in adults and children using a respondent-type training procedure: a follow-up study. Psychol. Rec. 47, 285–308.

Spradlin, J.E., Saunders, R.R., 1986. The development of stimulus classes using match-to-sample procedures: sample classification versus comparison classification. Anal. Interv. Dev. Disabil. 6, 603–628.

Spradlin, J.E., Cotter, V.W., Baxley, N., 1973. Establishing a conditional discrimination without direct training: a study of transfer with retarded adolescents. Am. J. Ment. Defic. 77, 556–566.

Urcuioli, P.J., 1996. Acquired equivalence and mediated generalization in pigeons. In: Zentall, T.R., Smeets, P.M. (Eds.), Stimulus Class Formation in Humans and Animals. Elsevier, Amsterdam, pp. 55–70.

Urcuioli, P.J., Zentall, T.R., DeMarse, T., 1995. Transfer of derived sample-comparison relations by pigeons following many-to-one versus one-to-many matching with identical training relations. Q. J. Exp. Psychol. 48B, 158–178.