The Psychological Record, 2003, 53, 617-634

DIFFERENTIAL PROBABILITY OF EQUIVALENCE CLASS FORMATION FOLLOWING A ONE-TO-MANY VERSUS A MANY-TO-ONE TRAINING STRUCTURE

ODDBJØRN HOVE
Helse Fonna HF

Twenty college students were randomly assigned to 2 groups. Dependent variables were presses on a touch screen, reaction time, and number of errors during training. An index of consistent responding to comparison stimuli during testing for equivalence was calculated. The subjects were also presented with a postexperimental task in which they sorted symbols used in the experiment. The main finding was that the many-to-one training structure was superior in producing equivalence outcome as compared to the one-to-many structure. There are indices that the 2 groups did not differ with respect to naming strategies. Reaction time measures differed in the 2 groups during the mix training but not during equivalence trials.

In a conditional discrimination task, as applied in a matching-to-sample procedure, the selection of a particular comparison stimulus (e.g., B1, not B2) is dependent upon a particular sample stimulus (e.g., A1) (A1B1 training). The minimal conditional discrimination training necessary for the testing of stimulus equivalence may include the following tasks in addition to the A1B1 training described above: A2B2 training, B1C1 (selection of comparison C1, not C2, is dependent on sample stimulus B1) and B2C2 training. Following this training, stimuli A1, B1, and C1 are said to be members of an equivalence class if a subject passes the tests for reflexivity (i.e., stimulus A is related to itself), symmetry (i.e., stimuli A and B are interchangeable), and transitivity (stimuli A and C are interchangeable if A relates to B and B relates to C) (Sidman, 1994). According to Sidman (1994) indices of equivalence are also obtained throughout the abbreviated or direct equivalence test. To pass this equivalence test, one must select comparison A1 (not A2) in presence of C1 and A2 (not A1) in presence of C2.

The study was conducted at University of Oslo, Norway. Oddbjørn Hove is currently at Helse Fonna HF, Rehabilitation Service for Adult and Children, Norway.
The author thanks Per Holth for helpful comments on an earlier draft of this paper.
Reprint requests may be sent to Oddbjørn Hove, Rehabilitation Service for Adults and Children, P.b. 2175, 5504 Haugesund, Norway. (E-mail: operant@online.no).

Any of the following matching-to-sample procedures may establish the conditional discrimination necessary for the testing of emergent equivalence relations: (1) one-to-many (OTM), in which the selection of comparison stimulus A1 (not A2) and C1 (not C2) are conditional upon sample stimulus B1, (2) many-to-one (MTO), in which the selection of comparison B1 (not B2) is conditional upon sample A1 and C1, and (3) linear series (LS), in which the selection of comparison B1 (not B2) is conditional upon sample A1 and selection of comparison C1 (not C2) is conditional upon B1 (Fields & Verhave, 1987; Saunders, Saunders, Williams, & Spradlin, 1993). All three types of training structures have been commonly used in the literature to produce two or more equivalence classes[1] (Saunders & Green, 1999). All of these procedures have been regarded as equally likely to produce emergent relations indicative of equivalence classes (e.g., Fields, Verhave, & Fath, 1984; Sidman & Tailby, 1982). The past few years have brought several studies that document differential probability of equivalence class formation following different training structures, both with mentally retarded subjects (Saunders et al., 1993; Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986) and intellectually normal subjects (Arntzen & Holth, 1997; Barnes, 1992, as cited in Barnes, 1994). Differential effects of training structure with respect to number of trials to criterion during training have been found in pigeons (Urcuioli & Zentall, 1993).

In general, the MTO structure has been regarded as superior in generating equivalence outcome as compared to the OTM and LS training structures (Barnes, 1992 as cited in Barnes, 1994; Saunders et al., 1988, 1993; Spradlin & Saunders, 1986). These studies typically made use of test sessions containing equivalence and symmetry test trials intermixed with trials from the training (baseline probes) following two-choice conditional discrimination tasks. None of these studies, however, were directly aimed at questions regarding differential effects of training structure. Rather, they typically compared different groups of subjects (Arntzen & Holth, 2000).

The only study that has investigated the relative effectiveness of different training structures directly is a study by Arntzen and Holth (1997). They found that, following three-choice conditional discrimination training, OTM was significantly more effective than MTO in creating the emergent relations indicative of equivalence class formations, and that the LS structure was the least effective in generating equivalence performance. They did not, however, check for sustained baseline performance. As pointed out by Saunders and Green (1999), the results reported in that paper might indicate that baseline performances engendered by the two training structures differed at the point where the isolated test trials were presented.

According to Sidman (1994), the OTM procedure is different from the

[1]The two training structures, MTO and OTM, are also described as comparison-as-node (CaN) and sample-as-node (SaN) respectively, in which the stimuli related to more than one other stimulus are node.

MTO procedure with respect to type of discrimination required during training versus those required during the test. The test for emergent relation following OTM training requires successive discrimination between samples that the subject previously had seen as comparisons, whereas a MTO training structure requires simultaneous discrimination between comparisons previously presented separately as samples. It has been suggested that the successive discrimination involved in equivalence test following OTM training is more difficult than the simultaneous discrimination involved in testing following MTO training (Saunders & Green, 1999). The results from Arntzen and Holth (1997) suggest otherwise. The findings from 1997 were partially replicated in Arntzen and Holth (2000). In Arntzen and Holth (2000), a within subject manipulation showed an equal likelihood of performance in accord with equivalence following OTM training as following MTO training. In both these papers, however, consecutive test blocks of 12 trials each were used. Other studies reporting MTO superiority have commonly used baseline trials intermixed with test trials.

Verbal control has been suggested as relevant to the probability of equivalence outcome (Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Horne & Lowe, 1996). Spradlin and Saunders (1986) mentioned that differences in equivalence outcome might be a result of different kinds of verbal control. Following the assumption that the OTM structure exposes the subject to more difficult discriminations than the MTO structure, Saunders and Green (1999) suggest that subjects will be more likely to name sample and comparison stimuli during MTO training than during OTM. This, in turn, may affect the extent to which subjects with an initial OTM training are prepared for the new discriminations called for in tests.

It has been shown that in some cases the subject, although not responding in accord with equivalence, nevertheless shows systematic responding during equivalence testing (e.g., Holth & Arntzen, 1998). The subject may, for instance, always select A2 when C1 is the sample, A3 when C2 is the sample and A1 when C3 is the sample. Data indicative of a pattern of systematic responding is not available, however, in any of the papers reporting differential effects of training structure. Thus, it is not at all clear whether the two training structures differ with respect to probability of systematic responding.

Reaction time to comparison tends to decrease over the course of training with mixed symbols (e.g., AB training intermixed with BC training) (e.g., Spencer & Chase, 1996). Reaction time to comparison stimuli during equivalence trials are found to be slower than reaction time to comparison stimuli during baseline trials (Arntzen & Holth, 1997; Bentall, Dickens, & Fox, 1993; Holth & Arntzen, 1998; Spencer & Chase, 1996), and that reaction time to comparison stimuli tends to decrease over the course of testing (Arntzen & Holth, 1997; Bentall et al., 1993; Bentall, Jones, & Dickens, 1999; Holth & Arntzen, 1998). Among these studies, only Arntzen and Holth (1997) employed both OTM and MTO training structures. However, no group differences in reaction time to comparison

during training with mixed sets of stimuli or equivalence probes were reported. Thus, no data are available regarding differential reaction times during training and test. All in all, it is not clear how reaction time measures relate to differences in training structures because almost all of these studies used a linear series training structure.

The main purpose of the present study is to replicate Arntzen and Holth (1997) by examining the probability of equivalence outcome following three-choice one-to-many versus many-to-one training structures with equivalence and symmetry trials intermixed with baseline trials. Second, the present study examines (a) number of errors during training, (b) patterns of systematic responding to comparison during the equivalence tests, (c) mean reaction time during training and testing, and (d) postexperimental classification and labeling of symbols used in the experiment.

Method

Subjects

Twenty college students served as subjects and were successively assigned to two different groups. None of the participants reported any knowledge of the equivalence paradigm when asked. Arrangement was made so that the subjects could not talk to each other before the experiment was finished.

Apparatus

A personal computer controlled stimulus presentation and collected the data. A transparent touch screen was mounted in front of a 15" monitor. A cassette player controlled by the computer played music automatically following correct response during training.

Procedure

Stimulus material. Visual materials were Greek letters, as shown in Table 1, displayed on the monitor. The sample stimulus (7x7 cm) was presented on the left side of the monitor. Comparison stimuli (4x4 cm) were presented in six different positions (two columns and three rows) on the right side of the monitor. No more than three comparison stimuli were presented concurrently.

Table 1

Symbols Used in the Experiment

Sets	1	2	3
A	δ	φ	λ
B	γ	μ	ξ
C	ψ	ζ	θ

General information. When asked to join the project, the subjects were told that it was a learning experiment and that the task was to press on symbols on a monitor. They were told that the experiment would last about 1 hour, dependent on how rapidly and correctly they responded.

Seated in front of the monitor the subjects were given the following instruction:

> I want to know what is learned in a particular task. Your job in the experiment is to press on symbols on the transparent touch screen. First, press "Start". When you press this, a symbol will appear on the right-hand side of the monitor. At first only one symbol; later there will be more. One of these is the correct one. You will know that it is the correct one when you hear music from the cassette player following your response. If you press the wrong symbol, the monitor will turn black for some seconds and you have to repeat the last tasks. It is not possible to get to the next set of tasks without a certain number of correct in a row. After a while, the program will continue without music when you press the right symbol, but the monitor will turn black if you press a wrong symbol. The training will be followed by test tasks intermixed with training tasks. On the test tasks, no different consequences will follow correct or incorrect responses – no music and no blank screen. Your job is to get as much correct as possible.

Training and test. Each task started with the presentation of a sample stimulus, and immediately following a touch on the sample stimulus the comparison stimuli were presented on the right-hand side of the monitor. The sample stimulus remained on the monitor until a comparison stimulus was touched. To reduce the number of incorrect responses during the initial conditional discrimination tasks comparison stimuli were gradually introduced (training part 1). A touch on the sample stimulus was followed by the correct comparison stimulus only. Next, each correct comparison stimulus was presented together with one incorrect comparison stimulus, and then with the second incorrect comparison stimulus. At last, during Trials 10 through 12, each correct comparison stimulus was presented together with both of the incorrect comparison stimuli. The three comparison stimuli appeared in a random position on each trial except that there was never more than one in each row. Following the gradual introduction of three comparison stimuli requiring nine successive correct, BA (in OTM) or AB (in MTO) training required 21 successive correct responses to comparison stimuli and so did the following BC (in OTM) or CB (in MTO) training (training part 2). Finally, 24 correctly completed tasks were required when the two sets were quasi-randomly intermixed (training part 3) before testing began.

The test consisted of 20 symmetry trials and 20 equivalence trials quasi-randomly mixed with 40 baseline trials summing to 80 trials. No programmed reinforcement followed the test trials. To increase the probability of high baseline performance during the test, consequences for incorrect baseline trials were arranged according to a variable ratio 3 schedule. The correction consisted of a 5-s blank screen and repetition of the last trial. The experiment elapsed without pauses. The experimental condition for each of the two groups is summarized in Table 2.

Postexperimental task. Immediately following the experiment a postexperimental task was conducted. Here, the subjects were asked to

Table 2

An Overview of the Experimental Design

Group	Training 1, 2, & mix	Intermixed test trials		
		Symmetry	Equivalence	Baseline
1. OTM	A ← B → C	AB and CB	CA	BA and BC
2. MTO	A → B ← C	BA and BC	CA	AB and CB

sort nine cards (4x4 cm) with an exact copy of the symbols used in the experiment. Spontaneously reports of labeling the stimuli were recorded. If they did not give any indications of labeling the symbols or of thinking about them in any specific manner, they were asked whether they actually had done so during the experiment.

Dependent Measures

Key presses on the touch screen in front of the monitor, reaction times to both sample and comparison stimuli, and the number of trials to criterion were recorded. Performance in accord with baseline, symmetry, and equivalence was defined as 90% correct responses to comparison or more.

Labeling of symbols during the postexperimental task were scored according to the following classification: 1 point for unique ascribing of labels to symbols within a class (e.g., A1 = "Y"; A2 = "E"); 2 points for ascribing of one label to two symbols within a class (e.g., both A1 and A2 labeled as "DY"); 3 points for ascribing one label to all three symbols within a class (e.g., all symbols A1, A2, and A3 labeled as "TOP").

Statistical Analysis

A two-tailed Fischer exact probability test was used to test differences in probability of equivalence outcome. Both the symmetry and equivalence test trials were divided in two for detection of delayed emergence of stimulus equivalence (test-half 1 and test-half 2). Each test half contained 10 symmetry and 10 equivalence probes. The 40 baseline trials were divided in two for detection of change in performance during testing (test-halves 1 and 2).

For statistical purpose, the reaction time to each press on comparison stimuli during training and testing was transformed logarithmically. This resulted in a near perfect normal distribution. Prior to the statistical analysis, mean reaction time of the first and last five responses to comparison stimuli during training parts 1 and 2, and during training with mixed set of tasks were calculated for each subject. Also, mean reaction time to comparison stimuli for each subject during the test phase was calculated. For easy detection of changes within the test, data was divided in three parts with the following number of trials in each test part: test part 1 consisted of seven equivalence trials, six symmetry trials, and 14 baseline trials; test part 2 consisted of six equivalence trials, six symmetry trials, and 13 baseline trials; test part 3 consisted of seven equivalence trials, seven symmetry trials, and 13 baseline trials. For statistical analysis of reaction time data, repeated measures ANOVA with one group factor was used.

Patterns of systematic responding during the equivalence test were evaluated throughout the test. Indices of systematic responding were calculated for each block of six consecutive trials, starting with Trials 1-6 and ending with Trials 15-20. Within any of the six-trial blocks, at least one of each trial type was included. An index of 0.0 indicates that no sample stimulus occasioned the same comparison selection more than once within that particular block of six trials. Each time a sample occasioned the same comparison as in a previous choice within a block, the index increased by 0.33. Indexes of 1.0 indicate that each sample occasioned its same comparison choice within a block (e.g., C1 always occasioned A1 (C1A1); C2A3; C3A2).

Results

During the first test-half, 1 of the 10 subjects in the OTM group responded in accord with equivalence. Following MTO training, 7 of 10 subjects showed equivalence. Two of the subjects showed symmetry performance in the first test-half following OTM training. Following MTO training, 8 subjects responded in accord with symmetry. During the first 20 baseline trials, 9 subjects in the OTM group emitted between 90% and 100% correct responses. In the MTO

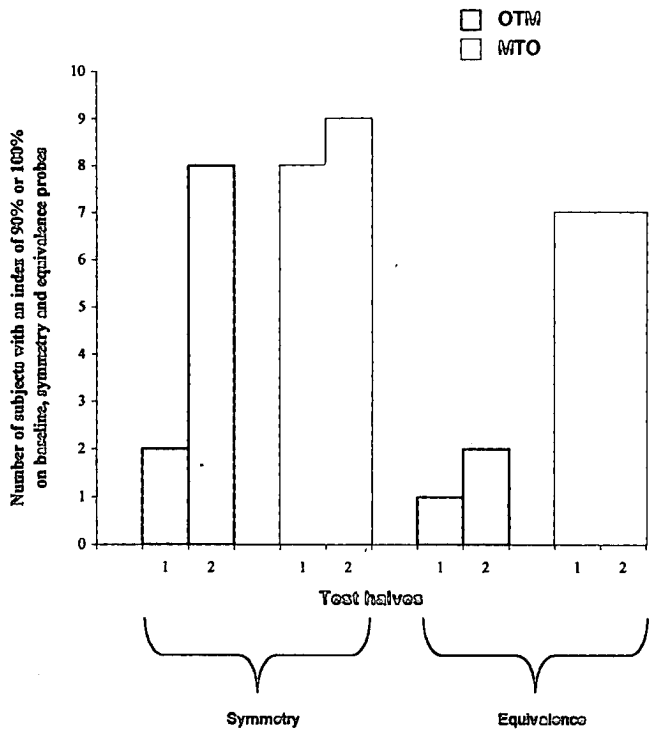

Figure 1. Number of subjects in the two groups with an index of 90% or 100% on symmetry and equivalence probes. OTM is illustrated in white bars. MTO is illustrated in gray bars.

group, 8 subjects showed sustained baseline performance during the first 20 baseline probes. A summary of the results on symmetry and equivalence probes from each test half is illustrated in Figure 1.

During the second test-half the number of subjects responding in accord with equivalence increased by one to two in the OTM group. In the MTO group Subject #011 showed an increase of correct responses from 80% to 100% and thus showing indices of equivalence. Subject #014 dropped from an equivalence index of 100% to 70% and thus did not meet the criterion for equivalence performance in the second half. Following MTO training, 9 subjects showed symmetry performance in the second test-half. The index for Subject #017 dropped slightly below the criterion for symmetry. This, however, did not seem to interfere with the equivalence performance as indicated by a stable 90% correct during both test-halves. Following OTM training, the number of subjects responding in accord with symmetry increased by five, to a total of seven. Two of the subjects dropped in baseline performance during the last 20 baseline probes. In the MTO group, all 10 subjects showed a solid baseline performance during the last 20 baseline probes. One of the subjects (#001) in the OTM group did show experimenter-defined symmetry performance despite a sloppy baseline. In the MTO group, 3 subjects (#011, #013 and #015) showed perfect or near-perfect symmetry performance within the first test-half and Subject #020 in the second test-half, although the baseline was weak. Two of these subjects also responded in accord with equivalence on the same first test-half. Results from individual subjects are shown in Tables 3 and 4.

Table 3

Summary of Individual Subject's Response during Training and Testing in the OTM Group

Sub	No. of trials during training			No. of errors during training			Baseline probes (%)		Symmetry (%)		Equivalence (%)		No. of corrected
	1	2	Mix	1	2	Mix	1-20	21-40	1-10	11-20	1-10	11-20	
#001	44	33	39	2	0	2	85	**90**	**90**	**90**	50	60	2
#002	77	33	61	7	0	7	**95**	**95**	80	**100**	40	70	0
#003	103	37	24	10	1	0	**100**	**100**	**100**	**100**	**100**	**100**	0
#004	74	35	77	4	2	6	**100**	**100**	80	**100**	70	**100**	0
#005	81	56	50	9	2	3	**95**	**95**	80	**90**	60	70	0
#006	888	94	64	149	10	5	**95**	**95**	80	**90**	70	80	1
#007	33	59	123	0	2	10	**95**	**100**	80	70	40	30	0
#008	296	33	24	69	0	0	**95**	70	80	60	40	50	4
#009	102	33	33	10	0	2	**90**	**90**	80	**90**	60	70	2
#010	57	45	26	3	2	2	**90**	75	50	80	20	40	6

Note. Responding in accord with experimenter-defined baseline, symmetry, or equivalence is shown in boldface number.

There was a significantly higher probability of equivalence outcome following MTO than OTM during the first test-half. When the subject with a weak baseline was removed from the sample (one in each of the two groups), there was still a significantly higher probability of equivalence

Table 4

Summary of Individual Subject's Response during Training and Testing in the MTO Group

Sub	No. of trials during training			No. of errors during training			Baseline probes (%)		Symmetry (%)		Equivalence (%)		No. of corrected
	1	2	Mix	1	2	Mix	1-20	21-40	1-10	11-20	1-10	11-20	
#011	57	41	68	4	2	7	**90**	**95**	**100**	**100**	80	**100**	2
#012	55	56	67	1	3	7	**90**	**95**	80	**90**	50	70	0
#013	79	50	36	5	3	1	**90**	**95**	**100**	**100**	**90**	**100**	1
#014	39	44	24	1	1	0	**100**	**100**	**100**	**100**	**100**	70	0
#015	45	39	128	2	2	20	**90**	**100**	**90**	**100**	**90**	**100**	1
#016	91	44	26	9	1	2	**95**	**100**	**100**	**100**	**100**	**90**	1
#017	56	97	89	4	11	11	**95**	**95**	**90**	80	**90**	**90**	1
#018	69	33	50	12	0	2	**100**	**100**	**100**	**100**	**100**	**100**	0
#019	78	33	24	5	0	0	**100**	**100**	**100**	**100**	**100**	**100**	0
#020	80	33	64	7	0	10	85	**95**	60	**90**	60	50	4

Note. Responding in accord with experimenter-defined baseline, symmetry, or equivalence is shown in boldface number.

outcome in the MTO group as compared to the OTM group in the first test-half. During the second test-half; no significant difference between the two groups was detected with respect to equivalence performance. In the OTM group, data from 2 subjects were removed from the test due to weak baseline performance. Table 5 shows *p* values from the two-tailed Fischer exact probability test.

Table 5

P Values from Two-Tailed Fischer Exact Probability, Pair-wise Group Comparison for Both Groups

Groups	*p* value	
	First test-half	Second test-half
OTM vs. MTO (all Ss)	$p < 0.05$	$p = 0.0697$
OTM vs. MTO (solid baseline only)	$p < 0.05$	$p = 0.1534$

There were no significant differences between the two groups over training parts (group x training part interaction) with respect to number of incorrect responses to comparison, $F(2, 36) = 1.19$, $p > 0.315$.

Consistency indices (Figures 2 and 3) throughout testing for subjects in the OTM group show that 4 of the 8 subjects who did not respond in accord with equivalence nevertheless responded consistently. The remaining 4 subjects responded inconsistently. Subject #009 responded systematically during the initial part of testing. The index dropped, however, in the last part of testing. In the MTO group, 2 subjects did not respond in accord with equivalence but showed consistent responding. Thus, all 10 subjects in the MTO group responded consistently to comparison stimuli. The two-tailed Fischer exact probability test showed that there was no significant difference between the two groups with respect to probability of systematic responding, $p = 0.6284$.

During training with both sets AB and CB (for the MTO group) and BA

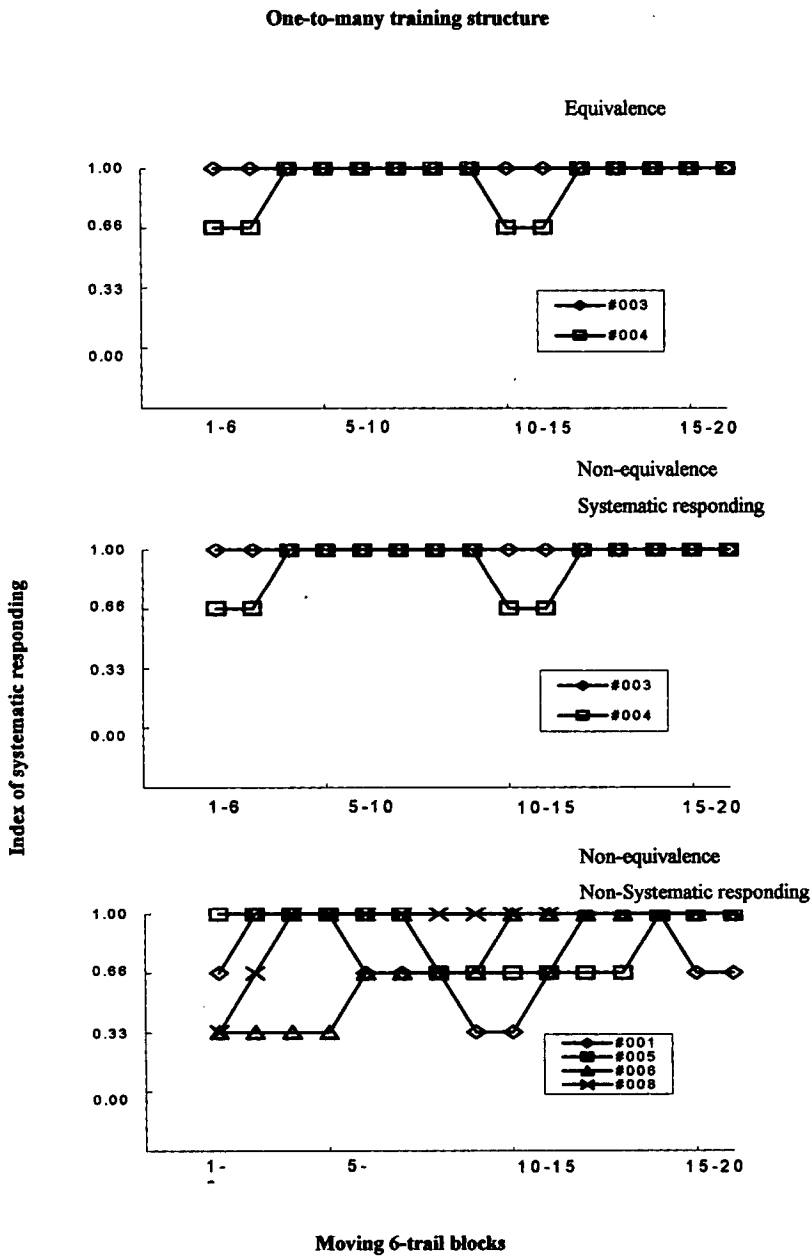

Figure 2. Individual curves for each subject in the OTM group according to different patterns of responding. The upper panel shows subjects responding in accord with equivalence, the middle panel shows systematic nonequivalence, and the lower panel shows nonsystematic responding.

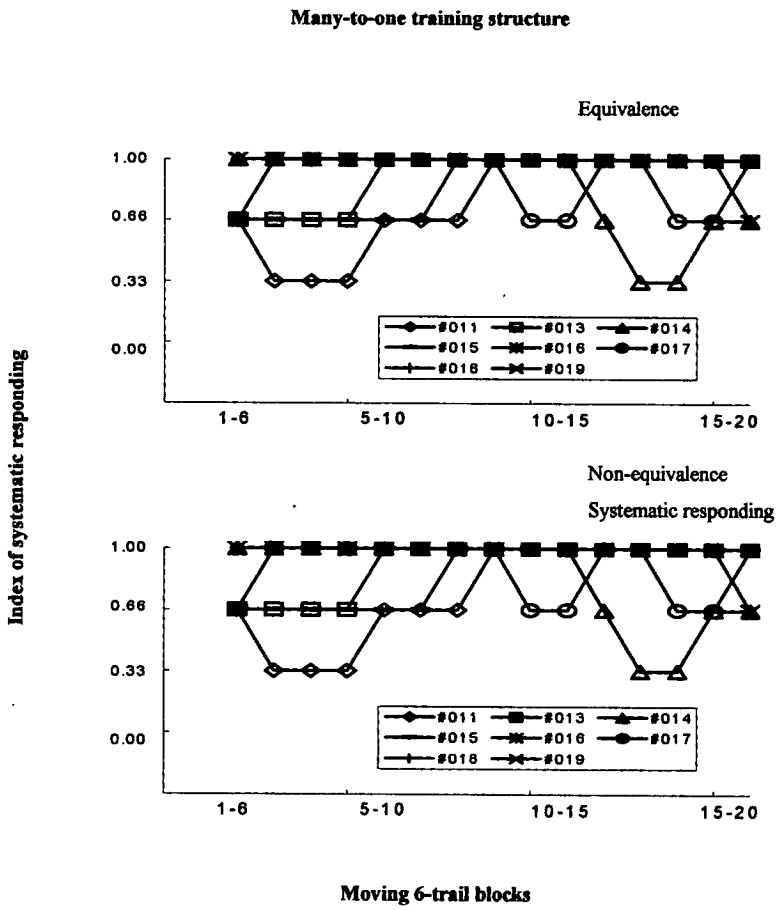

Figure 3. Individual curves for each subject in the MTO group according to different patterns of responding. The upper panel shows subjects responding in accord with equivalence, and the lower panel shows nonsystematic responding.

and BC (for the OTM group), that is the mixed training, the OTM group emitted responses to comparison stimuli significantly slower than the MTO group, $F(1, 18) = 7.96$, $p < 0.05$. There was a significant decrease in reaction time from the first to the last five responses during the mixed training, $F(2, 36) = 9.22$, $p < 0.001$ in both groups, and an increase from the last five during training to test phase one for the baseline probes, $F(1, 18) = 15.00$, $p < 0.001$ in both groups. The group difference in reaction time to comparison during baseline probes endured throughout the test, $F(1, 18) = 9.09$, $p < 0.01$. Figure 4 shows the logarithm transformed reaction time data.

In both groups, there was a significant decrease in reaction time from the initial to the final part of testing, both during equivalence and during

symmetry trials. The ANOVA showed the following values: $F(2, 36) = 24.35$, $p < 0.0001$. Both groups responded significantly slower to comparison during testing for equivalence than for baseline probes, $F(2, 36) = 61.37$, $p < 0.0001$. No significant differences in reaction time during equivalence and during symmetry testing were found.

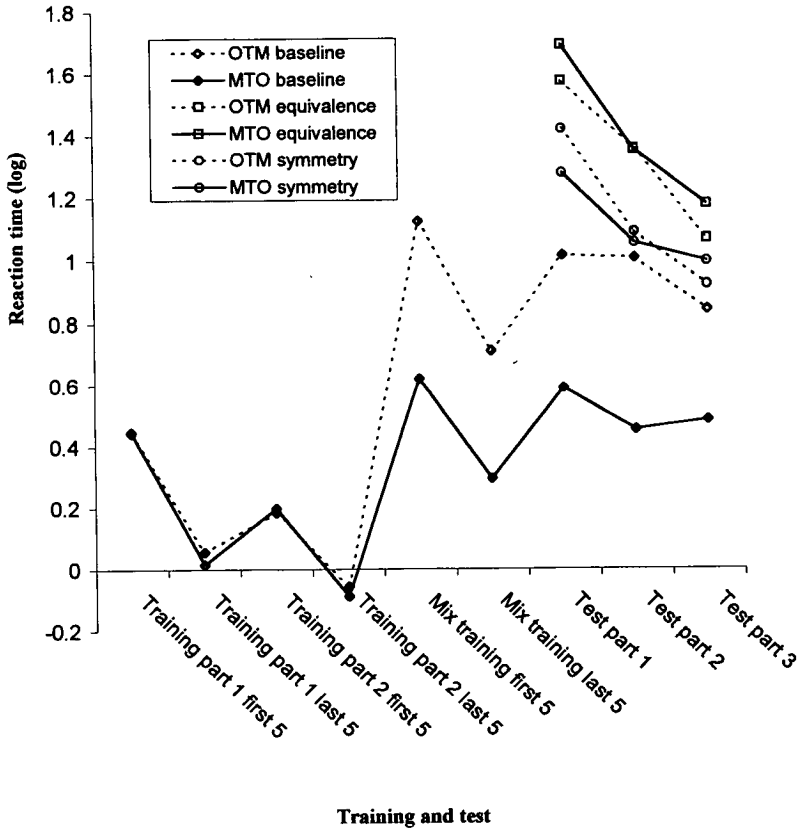

Figure 4. Average reaction time to comparison in the OTM and the MTO group during training and testing. Training parts refer to the first five and last five responses to comparison stimuli during the different training parts. Test part 1 refers to the first seven equivalence trials, the first six symmetry trials, and the first 14 baseline trials; test part 2 refers to the middle six equivalence trials, the middle six symmetry trials, and the middle 13 baseline trials; test part 3 refers to the last seven equivalence trials, the last seven symmetry trials, and the last 13 baseline trials.

On the postexperimental task, 6 of 10 subjects in the OTM group sorted the cards in accord with the experimenter-defined classes. Four of these did not respond in accord with equivalence during the experiment. All 10 subjects in the MTO group sorted the cards in accord with the experimenter-defined classes. Two of these subjects did not respond in accord with stimulus equivalence during the experiment.

All except 1 subject labeled some of the symbols during the postexperimental interview without being explicitly asked. The 1 subject (from the OTM group) who did not report any labels did not have the time to participate in this task. There were no differences in the two groups with respect to labeling. The average scores for labels in the OTM group were 7.44 as compared to 7.7 in the MTO group.

Discussion

The main finding was that the MTO conditional discrimination training was significantly more effective than the OTM procedure in producing the predicted equivalence outcome. The two groups did not differ with respect to probability of systematic responding during equivalence testing. It is also shown that although the two groups differed in accuracy of responses during testing, the two groups did not differ with respect to reaction time measures during the same test. The two groups differed with respect to average reaction time to comparison on training trials. Subjects in the OTM group responded significantly slower than subjects in the MTO group. This difference was evident from the start of the training part three, in which all different training types were intermixed consistently throughout the test. In the OTM group, the average reaction time to baseline trials during the test part was equal to average reaction time to symmetry trials in the same group.

The MTO superiority found in the present experiment is in accord with the findings reported in the majority of studies that have investigated this matter. The only exception seems to be the report by Arntzen and Holth (1997) which showed a OTM superiority and Arntzen and Holth (2000), in which the likelihood for equivalence performance was approximately equal in the two groups. The present study differs from the previous studies reporting MTO superiority by using three-choice rather than two-choice tasks, and from the Arntzen and Holth (1997, 2000) studies by using baseline probes intermixed with tests for emergent relation.

The main finding in this paper, that the MTO training structure is more likely to produce performance in accord with equivalence, supports the Saunders and Green (1999) analysis of the effect of differences in simple discrimination in the two training structures. They suggest that the simple discriminations presented during training account, at least, for some of the differential effects of training structure observed. During MTO training (AB, CB), the subject is exposed to successive discrimination of all sample stimuli (A1, A2, A3, C1, C2, and C3) and to simultaneous discrimination of all comparison stimuli (B1, B2, B3). In this case, the subject must discriminate A and C stimuli. According to Saunders' and Green's analysis, the only new discrimination during the subsequent equivalence test (CA) would be the assumed easier simultaneous discrimination of the comparison stimuli (A1, A2, and A3). Following the same analysis, the OTM training (BA, BC) exposes the subject to the successive discrimination of B1, B2, and B3, and the simultaneous

discrimination of all comparison stimuli (all As and all Cs). The subsequent equivalence test (CA) would expose the subject to a new successive discrimination of sample stimuli (C1, C2, and C3). The problem then, as illustrated by Barnes (1994), is that the simple discriminations involved in CA relations following OTM training structure, at least in some cases, may not be pitted against each other. Hence, the partition of preexisting classes into the experimenter defined-equivalence classes is more or less left to chance when the subject is exposed to training according to an OTM structure. One solution to this would be to present symmetry tasks before equivalence tasks. Following a positive symmetry test, all the subjects, regardless of which training structure would have the same simple discriminations prior to the equivalence test. In the present paper, 7 out of 10 subjects in the OTM group did respond in accord with symmetry during the second test-half. A subsequent equivalence test would give the answer whether or not this alters the probability of equivalence outcome when the OTM structure is employed. In the present experiment, however, the 1 subject who did respond in accord with symmetry during the first (and second) test-half did not respond in accord with equivalence as one might predict from the Saunders and Green (1999) analysis.

The basic prediction from the Saunders and Green (1999) analysis is that without simple discrimination, there is no way in which a subject on a reliable basis can hit upon the correct A stimulus when presented for any of the three C stimuli. In the present study, 6 of the subjects who did not respond in accord with equivalence during testing nevertheless classified the symbols into the according classes during the postexperimental task. This may indicate at least partial lack of simple discrimination. Contrary to the test arrangement on the computer, where stimuli were presented both simultaneously and successively, in this part of the experiment, all symbols were presented simultaneously. Thus, the stimuli had to be discriminated from each other. It is not possible with the current arrangement, however, to decide whether these subjects classified pairs as they were trained (e.g., A1=B1 in training part 1, and C1=B1 in training part 2) or as equivalence classes (A1=B1=C1). Unfortunately, none of these subjects were trained and tested for equivalence on a new set of symbols, following the postexperimental task. It is possible, however, that such an arrangement, following training with one set of stimuli, would make the subject differentially prepared for the CA problem with a new set of stimuli as compared to a subject without that particular history. If the OTM training structure does not establish all discriminations necessary for the emergence of stimulus equivalence, one would expect that the subjects responded nonsystematically during testing of these relations. It does not explain, however, why some of the subjects in the OTM group did show systematic nonequivalence, and others did show equivalence (and thus systematic responding). Thus, it seems that simple discrimination is a necessary but not sufficient condition for the emergence of stimulus equivalence.

In Arntzen and Holth (1997) only the direct equivalence test was used and, therefore, this test was used in the present experiment. In addition, this study made use of symmetry tests. There is a possibility that the subjects would be better prepared for the direct equivalence test if the test phase also contained transitivity test (AC) in that emergence of transitivity may precede emergence of equivalence on a direct test (Fields, Adams, Newman, & Verhave (1992). This, however, was not confirmed by Holth and Arntzen (2000). Rather the subjects' responses during transitivity and direct equivalence test developed into patterns which were in discordance with the experimenter-defined classes. In Holth and Arntzen (2000), a linear series (LS) training structure was used. This structure is different from the OTM and MTO structure in that a direct equivalence test in a LS structure requires the emergence of two symmetry relations, whereas the OTM and MTO structure requires only one symmetry relation. Further, the OTM and MTO structure differs with respect to directionality. Technically, a direct equivalence test following OTM training structure is equal to a transitivity test following a MTO structure. Sidman (1994) argues, however, that there is no way in which direction of training can have an effect on the subject in an experiment. There is still a possibility that test of transitive relations in the present experiment would alter the results on the direct equivalence test used here.

Saunders and Green (1999) suggested that the higher number of successive discriminations during MTO training might increase the likelihood of subjects ascribing labels to the stimuli used. Data from the present study, in which no differences in labeling were found, indicate otherwise. Although naming has been reported, and extensively debated in the literature on equivalence (e.g., Sidman, 1994), there are no standard methods for recording or scoring verbal reports. Often, subjects are asked whether they gave any of the stimuli a name. This raises the possibility that the response is dependent upon the subject being asked for names, and not that naming necessarily occurred during training and/or testing. In the present experiment, the subjects were not explicitly asked whether they ascribed labels to the symbols during the experiment. Rather the labeling was recorded as the subjects sorted symbol cards during the postexperimental part and thus eliminating the possibility that labels were assigned to the symbols dependent upon the subject being asked to do so. Still there is work to be done with respect to the effect of labeling stimuli. There may, for instance, be a difference in equivalence outcome in those subjects ascribing labels to classes of stimuli as compared to subjects ascribing labels to the individual stimuli.

Perhaps the most conspicuous part of the present study is that the opposite result with respect to probability of equivalence outcome was found as compared with the study conducted by Arntzen and Holth (1997). What is particularly interesting is that this study was conducted using the same computers as the present one, with the same program, the same training procedure, and the same stimuli. The only remaining obvious difference is that in the present study, baseline probes were intermixed with symmetry and equivalence probes whereas Arntzen and

Holth (1997) used blocks of equivalence probes without checking sustained baseline performance during or after the equivalence test. It is a possibility that the result in Arntzen and Holth (1997) reflects detoriated baseline performances in subjects trained according to the MTO structure. This, however, seems unlikely (e.g., Arntzen & Holth, 2000; Holth & Arntzen, 2000). Hence, the probability of equivalence outcome seems to be differentially affected in the two training structures when tests are conducted in isolated test blocks as compared to tests in which equivalence, symmetry, and baseline probes are intermixed. Why is this so? When testing is conducted in blocks, the subject may respond to this as a new training (CA), comparable to training parts 1 and 2, in which the sets (AB and BC) are trained separately. When equivalence testing is intermixed with symmetry and baseline probes, however, the testing may not appear as yet another novel block. Thus, it is possible that testing for equivalence in blocks induces a different kind of contextual control over responses than tests with equivalence, symmetry, and baseline probes intermixed. Contextual control would in this case be established over a number of trials. Thus, there should be an equal probability of equivalence outcome on the first equivalence test when MTO and OTM tested with blocks of trials and MTO and OTM tested with intermixed trials are compared. When this question came up, Per Holth supplied me with the original data from Arntzen and Holth (1997). The first equivalence trial for each of the subjects in the two groups from Arntzen and Holth (1997) was compared with the first equivalence trial for each of the subjects in the two groups in the present experiment. In the present experiment, 8 of the subjects in the MTO group selected the experimenter-defined correct comparison stimulus, whereas 7 of the subjects selected the correct comparison stimulus in the OTM group. In Arntzen and Holth (1997), 5 of the subjects selected the correct comparison in the MTO group, and 10 of the subjects selected the correct comparison in the OTM group. The Pearson Chi-Square showed that there was no significant difference in the four groups with respect to responses to the experimenter-defined correct comparison stimulus: $p = 0.074$. This indicates that differential outcome on equivalence tests may be directly related to the test procedure employed.

It has been suggested that contextual control by exclusion might explain some of the results indicating MTO superiority (Sidman, 1994). According to Arntzen and Holth (2000), this is unlikely. They suggest that another side effect of this procedural change from a two- to a three-choice task might be a differential change in task difficulty, as for example measured in numbers of errors during training (Arntzen & Holth, 1997). In the present study, no difference was found between the two groups with respect to errors during training. There was, however, a difference in reaction time during training with a mixed set of symbols. This difference was evident throughout the test. During training, when the two sets (BA and BC for OTM; AB and CB for MTO) were mixed, the OTM group responded significantly slower to comparison than the MTO group. According to Saunders and Green (1999), when presented for mixed

training, subjects in the OTM group must successively discriminate comparison stimuli A from comparison stimuli C. In the MTO group, however, no new successive discrimination of comparison stimuli is needed, because B stimuli in both tasks (AB; CB) serve as comparison. If the successive discrimination is more difficult than simultaneous discrimination, one may indeed predict that reaction time to comparison stimuli during training with mixed sets of symbols is slower during an OTM training structure as compared to a MTO training structure. Accordingly, one might also expect that reaction time to comparison during testing for equivalence would be different in the two groups. This, however, is not the case in the present experiment. An alternative interpretation of the group differences in reaction time during the mixed training is that, in the MTO group, two different samples occasion the same comparison. If the conditional discrimination between each of the samples and the corresponding comparison is established, the subject can predict what comparison to press even before the comparison is presented. In the OTM group, in contrast, the subject must wait for the comparison to be presented, simply because the same sample occasions two different comparison stimuli.

References

ARNTZEN, E., & HOLTH, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record,* 47, 309-320.

ARNTZEN, E., & HOLTH, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record,* 50, 603-628

BARNES, D. (1994). Stimulus equivalence and relation frame theory. *The Psychological Record,* 44, 91-124.

BENTALL, R. P, DICKENS, D. W., & FOX, S. R. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology,* 46B, 187-214.

BENTALL, R. P., JONES, R. M., & DICKENS, D. W. (1999). Error and response latencies as a function of nodal distance in 5-member equivalence classes. *The Psychological Record,* 49, 93-115.

DUGDALE, N., & LOWE, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 115-138). Hillsdale, NJ: Erlbaum.

EIKESETH, S., & SMITH, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior,* 58, 123-133.

FIELDS, L., ADAMS, B. J., NEWMAN, S., & VERHAVE, T. (1992). Interaction among emergent relations during equivalence class formation. *The Quarterly Journal of Experimental Psychology,* 45B, 125-138.

FIELDS, L., & VERHAVE, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior,* 48, 317-332.

FIELDS, L., VERHAVE, T., & FATH, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior,* 42, 143-157.

HOLTH, P., & ARNTZEN, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record,* 48, 81-110.

HOLTH, P., & ARNTZEN, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record,* 50, 305-337.

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

SAUNDERS, R. R., & GREEN, G. (1999). A discrimination analysis of training structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior,* 72, 117-137.

SAUNDERS, K. J., SAUNDERS, R. R., WILLIAMS, D. C., & SPRADLIN, J. E. (1993). An interaction of instruction and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record,* 43, 725-744.

SAUNDERS, R. R., WACHTER, J. A., & SPRADLIN, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedure. *Journal of the Experimental Analysis of Behavior,* 49, 95-115.

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

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.

SPENCER, T. J., & CHASE, P. N. (1996). *Journal of the Experimental Analysis of Behavior,* 65, 643-659.

SPRADLIN, J. E., & SAUNDERS, R. R. (1986). The development of stimulus classes using match-to-sample procedure: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities,* 6, 41-58.

URCUIOLI, P. C., & ZENTALL, T. V. (1993). The test of comparison-stimulus substitutability following OTM matching by pigeons. *The Psychological Record,* 43, 645-759.