

# On the role of trial types in tests for stimulus equivalence

Christoffer Eilifsen and Erik Arntzen  
Akershus University College

Some studies which have shown that differences in outcome on tests for stimulus equivalence dependent on different training structures, have run the tests as separate blocks without baseline trials interspersed in between test trials. Saunders and co-workers have argued that the differences in test outcome could be related to differences in the retention of trained discriminations during testing (R. R. Saunders, Drake, & Spradlin, 1999; R. R. Saunders & Green, 1999). In the current experiment, 20 adult participants were taught conditional relations by employing a linear series training structure. Following this training, non-reinforced trials of the directly trained relations were randomly interspersed in a mixed test for symmetry, transitivity, and global equivalence. After being exposed to the training procedure once, 17 of the participants did not perform in accord with stimulus equivalence, but 9 of these participants still responded in accord with the directly trained relations. After being exposed to the training procedure again, 10 participants still did not respond in accord with stimulus equivalence, while 7 out of these did respond consistent with the directly trained relations. This indicates that a "destroyed" baseline could not be responsible for these participants' failure to respond in accord with stimulus equivalence. In addition the reaction time between the appearance of comparison stimuli and subsequent responding during the test were recorded. Data show that the average reaction time varies as a function of which type of relation that is presented. There were also distinct differences in reaction time patterns for those participants who responded in accord with stimulus equivalence compared to those not responding in such a manner.

*Keywords:* Stimulus equivalence; linear series; direct trained trials in testing; adults

---

Stimulus equivalence can be defined as observed responding showing the emergence of novel conditional discriminations following the direct training of a series of interrelated conditional relations. This novel responding must show that the directly trained relations have the properties of reflexivity, symmetry, and transitivity. Reflexivity is an observation of identity matching among the stimuli involved in the prerequisite training. Symmetry is responding showing a reversal of the conditional and discriminatory functions of the stimuli used in the

training, while transitivity refers to responding showing a recombination of stimuli related in training by shared class membership (Sidman & Tailby, 1982). It is possible to arrange combined tests so that symmetry and transitivity can be assessed simultaneously. Such performance can be called global equivalence (Sidman, 1986). For explanatory purposes the stimulus classes established by training and tested for stimulus equivalence are often labeled by numbers, while individual members of each class are identified by letters. Stimuli in one class consisting of three members can be labeled by the alphanumeric codes A1, B1, and C1, while stimuli from a second class can be labeled A2, B2, and C2, and stimuli from a third class A3, B3, and C3. Individuals can, for example, be trained by using a conditional discrimination procedure to

---

Author note: Thanks to Felix Hognason and Richard Nartey for running some of the participants. Also thanks to Dag Erik Eilertsen for help with statistical analyses.

Correspondence concerning this manuscript should be addressed to Erik Arntzen, Akershus University College, PO Box 423, 2001 Lillestrom, Norway. E-mail: erik.arntzen@equivalence.net

respond to comparison stimulus B1, and not to comparison B2 or B3, given that stimulus A1 has been presented as the sample, and to comparison C1, and not to C2 or C3, given sample B1. In the same way individuals can be taught to match sample A2 to comparison B2, and not to B1 or B3, and to match sample B2 to comparison C2, and not to C1 or C3. They can then be trained to match A3 to B3, as opposed to B1 or B2, and B3 to C3 instead of C1 or C2. Following stable performance consistent with these experimenter designed contingencies, reflexivity can then be tested by arranging conditional discrimination test trials where the individual have the opportunity match sample A1 to comparison A1, with the choice of A1, A2, and A3 as comparisons. In the same manner the individual is given the opportunity to match sample B1 to comparison B1, with B2 and B3 as the other comparisons, C1 to C1, with C2 and C3 as the other comparisons, A2 to A2, with A1 and A3 as the other comparisons, B2 to B2, with B1 and B3 as the other comparisons, C2 to C2, with C1 and C3 as the other comparisons, A3 to A3, with A1 and A2 as the other comparisons, B3 to B3, with B1 and B2 as the other comparisons, and C3 to C3, with C1 and C2 as the other comparisons. Symmetry can be tested for by giving individual the opportunity to respond to comparison A1 given B1 as the sample, where A2 and A3 are the other comparison stimuli. In the same manner the individual is given the opportunity to match C1 to B1, with B2 and B3 as other comparisons, B2 to A2, with A1 and A3 as the other comparisons, C2 to B2, with B1 and B3 as the other comparisons, and B3 to A3, with A1 and A2 as other comparisons, and C3 to B3, with B1 and B2 as other comparisons. Transitivity is tested by arranging for the possibility to respond to comparison C1 given sample A1, with C2 and C3 as the other comparison stimuli. In addition, the opportunity is given to match A2 to C2, with C1 and C3 as the other comparisons, and A3 to C3, with C1 and C2 as the other comparisons. Global equivalence is assessed by arranging trials where it is possible to respond to comparison A1 given that C1 is the sample, with A2 and A3 as the other

comparison stimuli, to respond to comparison A2 given C2 as the sample, with A1 and A3 as the other comparisons, and to respond to comparison A3 given C3 as the sample, with A1 and A2 as the other comparisons. All tests are conducted in extinction conditions. If performance during the test then consistently shows the defining properties of stimulus equivalence, reflexivity, symmetry, and transitivity, it would be inferred, in the example described above, that three stimulus equivalence classes have been established (Green & Saunders, 1998).

The experimental procedure in stimulus equivalence experiments may vary according employment of different protocols. Fields et al. (1995) have distinguished between simple-to-complex, complex-to-simple, and simultaneous training protocols. The simultaneous protocol involves randomized presentations of the relations necessary to establish the prerequisite stimulus control for potential stimulus equivalence performance. This is usually followed by a separate test block randomly presenting trials testing for performance indicative of stimulus equivalence. Some studies have found the simultaneous training protocol to be less effective in producing stimulus equivalence compared to the two other protocols (Buffington, Fields, & Adams, 1997; Fields et al., 1997). In addition to the use of different protocols, stimulus equivalence experiments may be distinguished by the training structures used to establish the prerequisite conditional discriminations. The three different training structures linear series (LS), many-to-one (MTO), and one-to-many (OTM) have been used to train the prerequisite conditional discriminations for responding in accordance with stimulus equivalence. Although in the original analyses by Murray Sidman it was expected that the three training structures would lead to the same outcome on the equivalence tests (Sidman & Tailby, 1982), a number of later studies have shown that responding in accord with stimulus equivalence varies as a function of different training structures. When used in combination with a simultaneous protocol, the LS structure has been repeatedly been shown to be the least likely to lead to the formation of equivalence

classes (Arntzen, Grondahl, & Eilifsen, in press; Arntzen & Holth, 1997, 2000; Buffington et al., 1997; Fields et al., 1997). This training structure involves training a series of conditional discriminations where the comparison stimulus from one of the prerequisite conditional discriminations serves as the sample in the next. Individuals are trained to match A stimuli to B stimuli, and then B stimuli to C stimuli. Saunders and colleagues have proposed a discrimination analysis of the different training structures and the subsequent performance on tests for stimulus equivalence. This analysis suggests that the different training structures establish different combinations of successive and simultaneous discriminations which differentially prepare subjects for the discriminations necessary for successful performance on the test for stimulus equivalence. The analysis presupposes that successive discriminations are more difficult to learn than simultaneous discrimination, and that successive discrimination automatically leads to simultaneous discrimination, while simultaneous discrimination does not necessarily lead to successful performance when successive discrimination of the same stimuli is required. Saunders and colleagues concludes that as the MTO training structure establishes more successive discriminations than the OTM and the LS training structures, this training structure is superior to LS and OTM training in generating stimulus equivalence responding (K. J. Saunders & Spradlin, 1993; R. R. Saunders & Green, 1999). However, not all studies have supported this prediction (Arntzen et al., in press; Arntzen & Holth, 1997, 2000). In response to the study by Arntzen and Holth (1997), where utilization of the OTM training structure lead to a higher yield of stimulus equivalence responding than training by the MTO structure, Saunders and colleagues suggested that this may have been a result from the fact that the test for equivalence was conducted in a separate block from the training in this study. It is argued that this makes it possible that the directly trained relations were not maintained during the testing for the relations that define stimulus equivalence when the MTO structure was used (R. R. Saunders et al., 1999; R. R. Saunders & Green, 1999).

This argument raises the more general question of whether the failure to establish responding in accord with stimulus equivalence is caused by the experimental participant no longer being able to perform according to the contingencies established in the prerequisite baseline training. To examine this, the current experiment included the non-reinforced trials of baseline relations randomly interspersed among trials testing for symmetry, transitivity and global equivalence. As the current study aims to examine responding given the absence of responding in accord with stimulus equivalence, both an LS training structure and the simultaneous protocol was employed, so to decrease the chances of the participants exhibiting stimulus equivalence.

Reaction time data may be a useful additional measure employed when studying stimulus equivalence. This measure may be sensitive to different variables, even when accuracy on stimulus equivalence tests is not, and allows for a more fine grained analysis of the formation of stimulus equivalence classes (Dymond & Rehfeldt, 2001; Holth & Arntzen, 2000). Several studies have found marked differences in latency between the opportunity to respond to comparison stimuli and the occurrence of the actual response when examining different types of trials in the test for stimulus equivalence. Some have reported a difference in reaction times between trials involving directly trained relations and trials testing for emergent relations (R. R. Saunders & McEntee, 2004; Wulfert & Hayes, 1988). Others have in addition reported a distinct pattern of response latencies involving an increase in reaction time on symmetry trials compared to baseline trials, and a further increase in response latency between symmetry and transitivity or global equivalence trials (Arntzen, 2004; Arntzen et al., in press; Arntzen & Lian, in press; Bentall, Dickins, & Fox, 1993; Holth & Arntzen, 1998, 2000; Spencer & Chase, 1996). Studies have also shown that reaction times on all types of trials decrease during the course of the test (Arntzen, 2004; Arntzen et al., in press; Arntzen & Lian, in press; Holth & Arntzen, 1998, 2000). Several explanations for this systematic variation in reaction times have been proposed. Some have suggested some sort

of covert pre-current problem-solving behavior is taking place (Holth & Arntzen, 2000), while others have proposed that this pattern reflects varying degrees relatedness or substitutability among stimulus (Fields, Adams, & Verhave, 1993). Some have downplayed the relevance of this pattern of reaction times to stimulus equivalence responding (Tomanari, Sidman, Rubio, & Dube, 2006). The current experiment was conducted in part to try to replicate the pattern of systematic variations in reaction times according to which type of trial is presented and the decrease during the course of the test. It will also be examined if there are differences in reaction times between participants responding in accord with stimulus equivalence and participants not responding in such a manner.

In the current experiment, twenty adult participants were taught conditional relations using an LS-training structure. Following this baseline training, the participants were tested for the formation of a three 3-member classes. The purpose of the current study was (1) to arrange the tests for emergent relations with baseline probes interspersed with test trials to examine whether failure to respond in accord with stimulus equivalence corresponds with failure to respond according to directly trained baseline relations, (2) to replicate reaction time pattern concerning differences between trials involving directly trained relations, symmetrical relations, and transitive or global equivalence relations, (3) to compare reaction times of participants responding in accord with stimulus equivalence and participants not showing stimulus equivalence responding.

## Method

### *Participants and Setting*

20 adults between the age of 19 and 42 participated in the experiment. The experimental sessions were carried out in two different laboratories. Participants assigned the numbers 3001 through 3008, 6 female and 2 male health and social workers, were recruited in a professional setting. Participants indicated by the numbers 3009 through 3020, 6 males and 6 females, were recruited through personal contacts. The

experimental setting for the latter group was a small office, where participants were seated by a desk facing a window in front of them. For the former group the experiment was conducted in a small conference room, with office furniture lining the walls. Participants sat at one of these desks facing the wall. None of the participants had any experience with the current type of experimental setup or had any previous knowledge of stimulus equivalence.

### *Apparatus*

A HP Compaq nc6320 portable personal computer and a Dell Latitude D505 portable computer were used in the experiment. Both computers were equipped a built-in two-button touchpad and a 15 inch monitor. The experimental software was made by Psych Fusion Ltd in collaboration with second author. This software controlled stimulus presentations, recording of the responses, and administration of feedback to participants. Stimuli consisted of nine black abstract symbols, shown in Figure 1. These symbols were presented against a white background on the computer screen. The sample stimuli always appeared in the middle of the screen, while the comparisons randomly appeared in three of the four corners, always leaving one corner blank. During the training phase the total number of "correct" responses the participants had made was shown at the bottom of the screen. Prior to the start of the actual experiment participants were given printouts of the stimuli to be used in the experiment. These printed copies were made with laminated paper

	1	2	3
A			
B			
C			

Figure 1. The nine stimuli used in the experiment. The numbers above the stimuli indicates class membership, while the letters to the left of the stimuli indicate members of each class.

and were approximately the same size as they would later appear on the computer screen.

### Procedure

Upon arriving at the experimental laboratory all participants were given an information sheet which included a written consent form. This text let them know that they were about to take part in an experiment in the field of behavior analysis that involved doing tasks on a computer. Here they were also informed that the experiment would last for approximately one and a half hours, that there were no known harmful effects of participating in the study, and that they were free to withdraw from the experiment at any time without any negative consequences. After signing the consent form participants were given the small plastic-laminated pictures of the stimuli to be used in experiment, and told to categorize them. How they categorized the stimuli were then written down by the experimenter. Following this, participants were seated in front of the computer and presented with the following instructions on the computer screen:

A stimulus will appear in the middle of the screen. Click on this by using the computer mouse. Three other stimuli will then appear. Choose one of these by using the computer mouse. If you choose the stimulus we have defined as correct, words like very good, excellent, and so on will appear on the screen. If you press a wrong stimulus, the word wrong will appear on the screen. At the bottom of the screen, the number of correct responses you have made will be counted. During some stages of the experiment, the computer will not tell you if your choices are correct or wrong. However, based on what you have learned, you can get all the tasks correct. Please do your best to get everything right. Good Luck!"

Participants were then told that if they did not have any further questions they could press the start button located at the bottom of the screen. Any questions the participants had at this stage, or during the rest of the experiment were answered only with the information from the instructions above or the consent form. Following the response to the start button was the

appearance of a sample stimulus in the middle of the computer screen. Responding to this stimulus by using the computer mouse cursor to click on it was followed by the immediate appearance of the three comparison stimuli, while the sample stimuli remained on the screen. One of the three comparison stimuli appearing were part of the same experimenter designated class as the sample stimulus. Participants were exposed to a LS training structure, teaching the participants to match stimulus A1 to B1, B1 to C1, A2 to B2, B2 to C2, and A3 to B3 and B3 to C3. This was done by presenting the trials **A1**B1B2B3, **A2**B1B2B3, **A3**B1B2B3, **B1**C1C2C3, **B2**C1C2C3, and **B3**C1C2C3. The alphanumeric codes presented in bold letters are samples, while underlined alphanumeric codes indicate the experimenter defined "correct" comparisons. All these training trials were presented randomly and massed, with all possible types appearing three times in one block of trials. Consequently each block consisted of 18 trials. Each trial was followed by feedback lasting 1 s. When the participants pressed the correct comparison stimulus according to the experimenter designated classes the words *correct*, *excellent* and so on appeared on the screen. When a comparison stimulus from one of the other classes was chosen the word *wrong* appeared on the screen. Feedback was followed by a 1 s inter trial interval. Any given trial was repeated until the participant responded correctly according to the experimenter designated classes. When the participants matched 16 or more out of 18 trials in a block correctly, they proceeded to a training phases of reduced feedback. 16 out of 18 correct responses at a 100% feedback level were followed by phase of 75% feedback, where the feedback was randomly distributed among trials in a block. When the participant reached the criterion in the 75% feedback phase, feedback was reduced to 50%, then to 25%, and finally to 0%. Performance below the criterion in any of the phases, lead that particular phase being presented again. When the participants reached the mastery criterion in the training phase with no feedback, the test for stimulus equivalence was introduced. Here participants were exposed

to non-reinforced symmetry trials testing if they would match B1 to A1, C1 to B1, B2 to A2, C2 to B2, B3 to A3, and C3 to A3, transitivity trials testing if they would match A1 to C1, A2 to C2, and A3 to C3, and global equivalence trials testing if they would match C1 to A1, C2 to A2, and C3 to A3. In addition, non-reinforced trials of the directly trained relations were also included in the test, interspersed among the trials testing for symmetry, global equivalence, and transitivity. These directly trained trials, symmetry trials, transitivity trials, and global equivalence trials were all repeated three times, leading to the test block consisting of 54 randomly intermixed trials. See Table 1 for an overview of different phases of the experiment.

*Definitions of responding in accord with equivalence.* When tested, the participants had to match at least 17 out of 18 symmetry trials, 17 out of the 18 trials testing for transitivity or global equivalence, and 17 out of 18 directly trained relations “correctly” for performance to be considered in accord with stimulus equivalence. If the participants reached these criteria, the experiment ended after the test. If they did not, they were exposed to the training procedure again, followed by a second test. Once the participants had completed the second test, the experiment was over for all participants, even if

responding in accord with stimulus equivalence was not exhibited.

*Reaction time.* The reaction time from the presentation of the comparisons to a response to one of the comparisons was recorded. We calculated the mean median reaction time for all participants on the five last training trials, as well as for the five first and last five trials testing for directly trained relations, symmetry, and transitivity or equivalence.

*Categorization of stimuli.* After the actual experiment was over, participants were given the printouts of the stimuli used in the study and told to categorize them again. The categorization was noted down by the experimenter.

Participants were after the post-categorization task informed about the purpose of experiment, shown the data from their own experiment, and given an introductory article on the subject of stimulus equivalence.

#### Statistical analyses

The reaction time data were analyzed by repeated measures ANOVAs with one repeated factor (test types; direct trained, symmetry, and equivalence trials) and one group factor (test phase; first and last part of the test). The data were organized in 5 trial blocks – each block comprised one baseline measure for the five

Table 1. This table shows the different phases of the experiment, the different relations involved in each phase, the minimum number of trials per phase, and the percentage of feedback the participant received.

Phase	Trial types	Minimum # of trials	Feedback
Training phase 1 All trials trained simultaneously and randomly	A1-B1, B1-C1, A2-B2, B2-C2, A3-B3, B3-C3	18	100 %
Training phase 2 All trials trained simultaneously and randomly	A1-B1, B1-C1, A2-B2, B2-C2, A3-B3, B3-C3	18	75 %
Training phase 3 All trials trained simultaneously and randomly	A1-B1, B1-C1, A2-B2, B2-C2, A3-B3, B3-C3	18	50 %
Training phase 4 All trials trained simultaneously and randomly	A1-B1, B1-C1, A2-B2, B2-C2, A3-B3, B3-C3	18	25 %
Training phase 5 All trials trained simultaneously and randomly	A1-B1, B1-C1, A2-B2, B2-C2, A3-B3, B3-C3	18	0 %
Test phase All trials of all three types presented simultaneously and randomly	Relations previously trained	18	0 %
	Symmetry	18	0 %
	Transitivity/global equivalence	18	0 %
	A1-C1, A2-C2, A3-C3, C1-A1, C2-A2, C3-A3	18	0 %

last training trials and six test measures for test trials. Each measure was computed as a mean of five trials. Reaction times were measured in differences from baseline.

## Results

All participants finished the experiment, and they all did so in one session. The number of training trials each participant used to complete the training and their performance on trials testing for directly trained relations and symmetry relations are shown in Table 2. Performance on trials testing for transitivity or global equivalence is also shown in Table 2, in the column marked EQ. In addition, performance during the first test phase for all participants is also visually depicted in Figure 2. Here, black bars indicate responding on test trials involving the baseline relations, grey bars indicate respond-

ing on trials testing for symmetry, striped bars indicate responding on transitivity trials, and white bars indicate responding on trials testing for global equivalence. Responding is in Figure 2 depicted as percentage of responding in accord with each type of relations. Three out of twenty participants, participants #3002, #3004, and #3011, responded in accord with stimulus equivalence on the first test. These participants were therefore not exposed to a second round of training and testing. The remaining participants did not, on the first test, meet the criteria set for determining if stimulus equivalence classes had formed. Participant #3005 did not respond consistent with any of the three types of relations on either the first or the second test. Participant #3007 reached the criterion for symmetry relations on the first test, and the criteria for both symmetry and transitivity and global equivalence relations on the second test,

Table 2. Individual test data for the twenty participants. The left portion of the table shows data from the first part of the experiment, while the right portion shows the second part. NA is written in boxes where the part of the experiment indicated was not applicable for this particular participant. The number of training trials and the number of responses consistent with directly trained, symmetry, and transitivity and global equivalence relations are reported for both parts of the experiment. Bold text indicates that the participant's responding is within the mastery criterion set for that particular type of relation.

Participant number	Part 1				Part 2			
	# of training trials	Test trials			# of training trials	Test trials		
		DT	SYM	EQ		DT	SYM	EQ
3001	432	<b>17/18</b>	12/18	11/18	90	<b>18/18</b>	<b>17/18</b>	14/18
3002	162	<b>18/18</b>	<b>18/18</b>	<b>18/18</b>	NA	NA	NA	NA
3003	162	<b>17/18</b>	13/18	4/18	90	<b>18/18</b>	15/18	0/18
3004	144	<b>17/18</b>	<b>18/18</b>	<b>18/18</b>	NA	NA	NA	NA
3005	324	14/18	13/18	7/18	126	16/18	12/18	8/18
3006	252	<b>18/18</b>	<b>18/18</b>	16/18	90	<b>18/18</b>	<b>18/18</b>	<b>18/18</b>
3007	180	16/18	<b>17/18</b>	12/18	90	15/18	16/18	<b>18/18</b>
3008	162	16/18	15/18	11/18	90	<b>18/18</b>	<b>18/18</b>	<b>18/18</b>
3009	324	<b>18/18</b>	<b>18/18</b>	12/18	90	<b>18/18</b>	<b>18/18</b>	<b>18/18</b>
3010	198	16/18	16/18	<b>18/18</b>	90	16/18	<b>17/18</b>	<b>18/18</b>
3011	252	<b>18/18</b>	<b>17/18</b>	<b>18/18</b>	NA	NA	NA	NA
3012	162	15/18	16/18	9/18	90	<b>18/18</b>	<b>18/18</b>	2/18
3013	180	<b>17/18</b>	<b>17/18</b>	1/18	90	<b>17/18</b>	<b>18/18</b>	0/18
3014	216	<b>18/18</b>	<b>18/18</b>	15/18	90	<b>18/18</b>	<b>18/18</b>	<b>18/18</b>
3015	180	14/18	10/18	3/18	90	<b>18/18</b>	<b>18/18</b>	14/18
3016	216	<b>17/18</b>	<b>18/18</b>	7/18	90	<b>18/18</b>	<b>18/18</b>	16/18
3017	270	14/18	15/18	14/18	108	<b>18/18</b>	<b>18/18</b>	<b>18/18</b>
3018	144	<b>18/18</b>	<b>18/18</b>	16/18	108	<b>18/18</b>	<b>18/18</b>	<b>17/18</b>
3019	162	<b>17/18</b>	<b>17/18</b>	16/18	90	<b>18/18</b>	<b>18/18</b>	<b>18/18</b>
3020	198	14/18	15/18	5/18	108	<b>18/18</b>	<b>18/18</b>	16/18

but not the criterion for directly trained relations on either of the tests. Similar results are seen in the performance of participant #3010, who responded according to transitivity and global equivalence on the first test, while, like participant #3007, reaching the criteria for symmetry and transitivity and global equivalence on the second test. This participant did not consistently respond according to the directly trained relations on any of the tests. Participant #3001 and #3003 respond to criterion only on the test for directly trained relations on the first test. While participant #3003 also reaches the criterion set for determining appropriate stimulus control on the symmetry trials on the second test, participant #3001 again only responds consistent with the directly trained relations. After being exposed to the training procedure once, participant #3012, #3015, and #3020, display none of the discriminations under study on the first test. However, after going through training again, they reach the set criteria for directly trained relations and symmetry relations. Participant #3013 and #3016 reach the criteria for both directly trained relations and

symmetry relations both on the first and on second test. When being tested for the first time, participant #3008 and #3017 do not reach the criteria for any of the types of relations tested for. After the training procedure is administered again these participants do, however, perform consistent with all these three types of relations when tested. Participant #3006, #3009, #3014, #3018, and #3019 respond according to criteria for determining appropriate stimulus control on directly trained relations and symmetrical relations in test 1, while displaying all three types of in the second test. Altogether, 10 out of 20 participants displayed responding consistent with stimulus equivalence defined by a minimum of systematic performance on trials involving directly trained relations, symmetrical relations, and transitivity and global equivalence relations, by the end of the second test.

The participants used between 144 and 432 trials to complete the first training phase. Participants #3002, #3004, and #3113 who reached the test criteria for all three relations tested for on the first test used an average of 186 trials to complete the training procedure, while

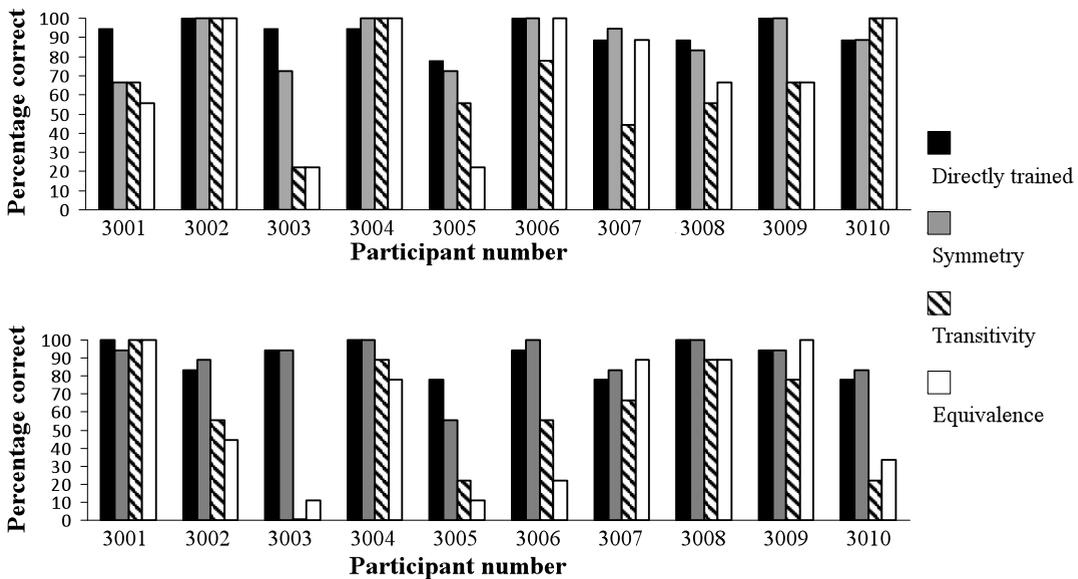


Figure 2. Graphic display of individual test results shown as percent of responses consistent with the different types of relations tested for. Black bars show performance on trials testing for directly trained relations, grey bars show trials testing for symmetry, striped bars show transitivity trials, and white bars indicate performance on trials testing for global equivalence. The upper portion of the figure shows the results for participants #3001 to #3010. The lower portion of the figure shows the results for participants #3011 to #3020.

participants #3005, #3008, #3012, #3015, #3017, and #3020, who did not perform in accordance with criteria for any of the relations used an average of 216 trials. Participants #3001, #3003, #3006, #3009, #3013, #3014, #3016, #3018, #3019, who met the criteria for the directly trained relations or the directly trained and symmetry relations used an average of 232 trials complete the training. Participant #3007 and #3010, who were the ones meeting the criterion for one of the derived relations, but not for directly trained relations, used an average of 225 trials to complete training. The participants who continued to a second training phase after the first test for the most part completed the second training phase in the minimum number of trials possible, which were 90 trials, or 5 blocks. The exception was participants #3005, who used 126 trials (7 blocks), and participants #3017, #3018, and #3020, who all completed the second training phase using 108 trials (6 blocks).

Figure 3 shows the mean median reaction

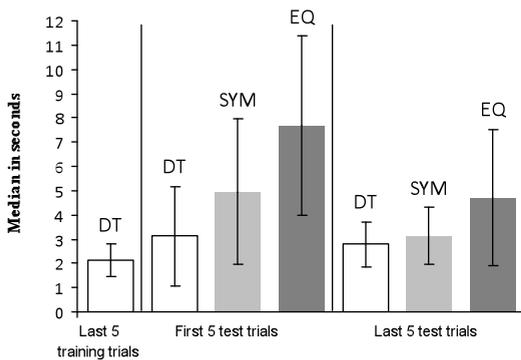


Figure 3. The figure shows group-data of reaction times to comparison for all participants on the first test for stimulus equivalence. Starting from the left, light grey bars show the mean median reaction time of the last five training trials and the first five trials testing for directly trained relations. Reaction times for the first five symmetry trials are indicated by dark grey, followed by the first five transitivity or global equivalence trials in black. The mean median reaction time for last five test trials testing for these three types of relations are then shown using the same colors indicating the different relations. The top and bottom of each error bar indicates plus and minus one standard deviation.

time to comparison for all participants for the last five trials of the training phase, the first five trials testing for directly trained relations during the test, the first five trials testing for symmetry and the first five trials testing for transitivity or global equivalence during the first test phase. Light grey bars indicate directly trained relations, while dark grey bars indicate symmetry trials. The black bars indicated trials testing for transitivity or global equivalence. The mean median for the last five trials of training was approximately 2.14 s, while it was approximately 3.12 s for the first five trials involving the directly trained relations when they appeared interspersed test trials for the emergent relations. For the first symmetry trials the mean median reaction time to comparison was approximately 4.95 s, while the average median on the transitivity and global equivalence trials was 7.68 s. The measures of the last five test trials testing for either directly trained relations symmetry, or transitivity and global equivalence show a decrease in reaction times, to an mean median of 2.8, 3.13, and 4.7 s, respectively.

Figure 4 shows the reaction times during the first test for participants #3001, #3003, #3005, #3006, #3007, #3008, #3009, #3010, #3012, #3013, #3014, #3015, #3016, #3017, #3018, #3019, and #3020. These are the participants who failed to reach the criteria set for mastery on one or more of the three types of relations tested for, and therefore did not respond in accord with the experimenter set definition of stimulus equivalence. As in Figure 3, light grey indicate trials of relations that have been directly trained, dark grey indicate symmetry trials, and black indicate transitivity and global equivalence trials. The mean median reaction time for the last five training trials was approximately 2.26 s, while the mean of the median of the first five trials of the directly trained relations appearing in the test phase was 3.33 s. On the first five trials testing for symmetrical relations the mean median was approximately 5.12 s, while it was 8.5 s on the first five trials testing for transitivity or equivalence. On the last five trials of the test, the mean median reaction time was approximately 2.91 s on trials involving previously directly trained relations, 3.19 for

trials testing for symmetry, and 5.15 s for trials testing for transitivity or global equivalence. The statistical analyses showed that there was an interaction effect of test type (direct trained, symmetry and equivalence trials) and test phase (first and last part)  $F(2,18) = 10,08, p = .001$ . The t-test showed that there was an effect of test phase both for symmetry trials and equivalence trials,  $t(1,19) = 2,7, p = .01$  and  $t(1,19) = 5,07, p = .001$ , respectively, which means that reaction time decreased during the test.

Figure 1 shows the mean median of reaction times for participants #3002, #3004, and #3012 on trials in the first test. These are the participants who responded in accord with the experimenter set criterion for stimulus equivalence. The colors of the bars have the same connotations as in Figure 3 and Figure 4. The mean median reaction time of the last five training trials was approximately 1.48 s, while for the first five trials testing the same type of relation during the test it is 1.97 s. Average median on the first five symmetry trials was approximately

4.0, while the first five transitivity and global equivalence trials have a mean median of 2.9. The last five trials of the test for each of the three types of relations have a mean median of approximately 2.14 s for directly trained relations, 2.74 for symmetry, and 2.14 s for transitivity or global equivalence.

None of the participants categorized stimuli according to the experimenter defined classes prior to the start of the experiment. During the post-categorization 14 out of 20 participants, #3002, #3003, #3004, #3006, #3007, #3008, #3013, #3014, #3015, #3016, #3017, #3018, #3019, and #3020, sorted the printouts of the stimuli according to the experimenter designated classes. 6 of these participants, #3003, #3007, #3013, #3015, #3017, and #3020, did not, by the end of the second test, respond to criteria for one of more of the three types of relations tested for in the actual experiment. 4 out of these 6, #3013, #3015, #3016, and #3020 met the criteria for both the directly trained relations and the symmetry relations, while 1 (#3003) reached only the criterion for the

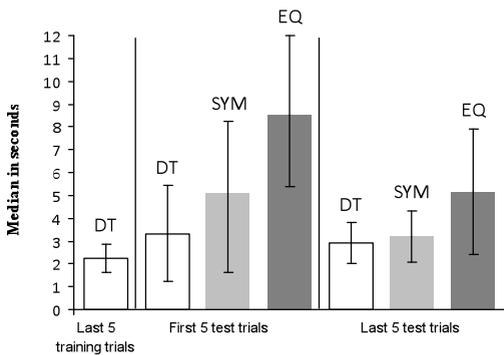


Figure 4. The figure shows group data for reaction time to comparison for participants failing to reach the criterion for stimulus equivalence. From the left, light grey bars show the mean median reaction time of the last five training trials and the first five trials testing for directly trained relations. Reaction times for the first five symmetry trials are indicated by dark grey, followed by the first five transitivity or global equivalence trials in black. The mean median reaction time for last five test trials testing for these three types of relations are then shown using the same colors indicating the different relations. The top and bottom of each error bar indicates plus and minus one standard deviation.

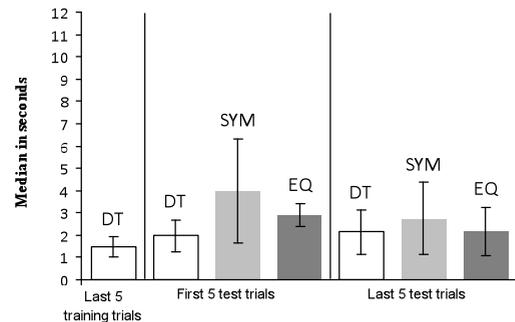


Figure 5. The figure shows group data of reaction time to comparison for participants reaching the criterion for stimulus equivalence on the first test. From the left, light grey bars show the mean median reaction time of the last five training trials and the first five trials testing for directly trained relations. Reaction times for the first five symmetry trials are indicated by dark grey, followed by the first five transitivity or global equivalence trials in black. The mean median reaction time for last five test trials testing for these three types of relations are then shown using the same colors indicating the different relations. The top and bottom of each error bar indicates plus and minus one standard deviation.

directly trained relations, and 1 (#3007) only for the transitivity and global equivalence relations. 1 out of the 10 participants, participant #3009, that responded consistent with all three types of relations during the test for stimulus equivalence did not sort the stimuli according to the experimenter designated classes in the categorization task.

### Discussion

The current experiment was conducted to examine whether baseline relations would be displayed or not by participants who did not perform in accord with stimulus equivalence. This was done by using an LS training structure and including interspersed probes for directly trained relations in the test for stimulus equivalence. In addition, we wanted to examine reaction time patterns, both by comparing reaction times to comparison stimuli between different types of trials in the test, and by comparing participants who responded in accord with stimulus equivalence and participants who exhibit stimulus equivalence responding.

Our findings replicated earlier studies on the effectiveness of producing equivalence responding with the LS training structure (Arntzen et al., in press; Arntzen & Holth, 1997, 2000; Buffington et al., 1997; Fields et al., 1997). Only 3 out of 20 participants responded in accord with stimulus equivalence on the first test, while an additional 7 participants responded consistent with the definitional requirement for stimulus equivalence on the second test. In other words, after being exposed to the training procedure twice, only 50% of the participants responded in accord with stimulus equivalence. This can be considered a fairly low yield of stimulus equivalence responding when using adults as participant. Several of the participants who did not respond in accord with all the requirements for stimulus equivalence showed that the baseline performance was intact during the first or second sequence of training and testing. On the first test, 2 of the 17 the participants not responding according to stimulus equivalence still responded to criterion on the directly trained relations, while 7

other participants responded to criteria on both the directly trained relations and symmetry. Altogether, 9 out of 17 participants not showing stimulus equivalence class formation during the first test still responded in accordance with the directly trained relations. On the second test, 6 of the 10 participants who did not reach the set criterion for determining responding to be in accord with stimulus equivalence reached the set criteria for both the directly trained relations and symmetry, while 1 participant only reached the criterion for the directly trained relations. Altogether, 7 out of the 10 participants not showing stimulus equivalence responding still performed to criterion on the trials testing for directly trained relations on the second test. The data obtained in the current experiment show that lack of responding in accordance with the relations that define stimulus equivalence is not in general accompanied by the participant not being able to perform according to the directly trained discriminations. These results indicate that the failure to respond in accord with stimulus equivalence is not in general caused by the loss of the ability to perform according to the baseline contingencies. This retracts from Saunders and co-workers argument that Arntzen and Holth's (1997) demonstration of higher yields of stimulus equivalence using an OTM training structure compared to an MTO training structure came about because of "destroyed" baseline relations (R. R. Saunders et al., 1999; R. R. Saunders & Green, 1999). However, the argument of Saunders and colleagues referred to the use of an MTO training structure, while the current experiment employed an LS training structure. This training structure was employed in the current experiment partly to limit the number of participants showing stimulus equivalence responding. Most experiments employing MTO and OTM training structures have yielded positive outcomes on the test for stimulus equivalence when adults have been used as participants (R. R. Saunders & Green, 1999). This made it likely that results obtained using one of those training structures would be of limited use for answering the research questions of the current experiment, which required examination of responding in the absence of

stimulus equivalence performance. The LS training structure was also employed because it can be argued that Saunders and colleagues raised the question of whether “destroyed” baseline discriminations in general is the cause of failure to respond according to stimulus equivalence. This can be tested by using any training structure. However, it remains possible that the use of an MTO training structure will yield different results from those obtained in the current experiment. In general, future research should include the use of different protocols and training structures, and increase the number of stimuli per class and number of classes used, to examine if similar results can be obtained under different circumstances than those examined in the current experiment.

Reaction time data replicate previous findings showing a distinct pattern when reaction times to comparison are categorized according to which type of relation they involve. When looking at all participants, there is on average a slight increase in median reactions time of the five last training trials compared the median of the five first trials presenting directly trained relations in the test. The mean median reaction time on early symmetry trials is higher than on the trials involving the directly trained relations, while an additional increase is observed on the mean median of the five trials testing for transitivity or global equivalence in the beginning of the test. Average median reaction times have decreased on trials involving all three relations by the end of the test, but the relative differences in reaction times when comparing the three different kinds of relations remain similar to what they were in the beginning. These results are comparable to those obtained in a number of other studies (Arntzen, 2004; Arntzen et al., in press; Arntzen & Lian, in press; Holth & Arntzen, 1998). The pattern of low reaction times on trials involving directly trained relations, higher reactions times on symmetry trials, and an additional increase in reaction time on trials testing for transitivity and global equivalence trials, is even more distinct when one looks at the participants only reaching the set criteria for directly trained relations or both directly trained relations and symmetry. As have been

suggested elsewhere (Pilgrim & Galizio, 1996), these reaction time differences may indicate that performance seen in relation to the different trial types involved in stimulus equivalence experiments differ in significant ways. However, when looking only at the participants responding in accord with stimulus equivalence on the first test, this characteristic pattern is not present. For these participants the reaction times are much lower, and the mean median response latencies on the trials testing for different types of relations are similar to each other, with symmetry trials having the on average highest mean reaction time both in the beginning and at the end of the test. The variations in reactions described above are not observed in the performance of these three participants at all. Such distinctive and systematic differences in reaction time between participants responding in accord with stimulus equivalence and those that do not are difficult to ignore. The current results can be looked at in the light of similar to results obtained in a study by Bentall, Dickins, and Fox (1993). In this study reaction times were lower and differences between types of trials were minimal for participants trained and tested for stimulus equivalence with stimuli assumed to be pre-associated, compared to participants who were trained and tested with only abstract stimuli divided into arbitrary classes. The latter group showed a reaction time pattern similar to the participants in the current experiment who did not respond in accord with stimulus equivalence. Although the relationship between reaction time and stimulus equivalence responding is not directly examined in the study by Bentall, Dickins, and Fox, the pre-association group are also the group who has the highest yield of stimulus equivalence responding. Bentall and co-workers conclude that this difference in reaction times is most likely the result of the availability of a common name for each class of stimuli for the pre-associated group. Therefore a common labeling strategy is more likely to produce stimulus equivalence responding than either individual labeling of each stimulus or not responding verbally to stimuli at all. However, the fact that the relationship between stimulus equivalence and reactions times are

not directly examined in this study, in addition to Bentall and colleagues only examining the reaction times of responding deemed "correct" makes comparisons between their study and the current experiment somewhat difficult.

As none of the participants categorized the printouts of the stimuli according to the experimenter designed classes during the pre-experiment sorting task, the possibility of the stimuli being pre-associated according to these classes was clearly ruled out. In the post-categorizations task two of the participants who responded in accord with all the three relations involved in the set mastery criterion for stimulus equivalence, did not categorize the stimuli consistent with the experimenter designated classes. 6 out of the 10 participants who failed to reach the criteria for either the directly trained relations, the symmetry relations, and/or the transitivity and global equivalence relations, still categorized the stimuli according to the experimenter defined classes. Thus, stimulus sorting did not converge well with responding on the test for stimulus equivalence. This is consistent with at least some of the previous experiments that have included such a test of stimuli sorting after the test for stimulus equivalence (Smeets, Dymond, & Barnes-Holmes, 2000). However, the sorting task is considerably different from the test for stimulus equivalence. In the current experiment, this task involved all stimuli being present simultaneously, so that is possible for the participants to scan back and forth between the different stimuli. Furthermore, the sorting of stimuli into categories was done in a single trial.

The results of participants #3007 and #3010 are difficult to interpret, as they do not reach the criterion set for determining sufficient stimulus control on the trials interspersed in the test involving previously directly trained relations, but do reach the set criteria for either symmetry or transitivity and global equivalence, or both symmetry and transitivity and global equivalence. The two participants display such responding on both of the test for stimulus equivalence they were exposed to. Such results have rarely been published in the stimulus equivalence literature. An exception is Saunders, et al. (1999), where one child participant fails to perform according

to symmetry, but reach the mastery criteria for transitivity. A closer look at the responding of this participant revealed that the majority of "errors" were made on trials involving samples from only one of the stimulus classes. Because of this Saunders, et al. (1999) concluded that the unusual responding during the test most likely was caused by negative stimulus control being responsible for some of the "correct" responding during the training, resulting in the prerequisite baseline class never actually being established according to the experimenter designed contingencies (Carrigan & Sidman, 1992). This subsequently led to the participant's failure to respond according to all the relations that define stimulus equivalence on trials involving one of the stimulus classes, and explains the unusual response pattern. However, in the current experiment the responding that do not reflect the development of stimulus equivalence are not systematically made on trials testing within one particular class of stimuli. In addition the "errors" are distributed throughout the test, for both participants and on both test, ruling out the possibility of delayed emergence of stimulus equivalence (Bush, Sidman, & de Rose, 1989; Sidman, 1994). Such performance contradicts the view that the relations that define stimulus equivalence are a coherent whole originating from the demands of the reinforcement contingency (Sidman, 1990, 1994, 2000). There is a possibility that the performance of these two participants is the result of procedural issues or even lack of experimental control in some way. One possibility is that the responding may reflect the fact that stimulus equivalence and the percentage correct measure used to determine if such performance is present is not one and the same. The responding of participants #3007 and #3010 show a fairly high number of "correct" responses, as they do not respond with less than 15 out of 18 of the trials in accord with any of the three types relations tested for during the second test. Interestingly, while participant #3007 sorts the stimuli according to the experimenter defined classes during the post-categorizations task, participant #3010 does not. So, while it may be possible to argue that the unexpected performance of participant

participants #3007 and #3010 is a result of procedural artifacts, this argument is, as far as we can see, much more difficult to apply to the performance of all the participants who responded according to the directly trained relations, but not according to all the relations that define stimulus equivalence. The responding of these participants indicates that the necessary prerequisite discriminations were established and maintained during the course of the test. Yet stimulus equivalence fails to emerge, with some participants responding only in accord with the directly trained relations, while others respond in accord with the directly trained relations and symmetry. These results seems difficult to explain if stimulus equivalence is to be viewed as a necessarily consistent set of relations arising from the reinforcement contingencies established during the training.

In summary the current study showed that in the majority of the cases the participants' baseline performance was intact even when these individuals did not respond in accordance with the relations that define stimulus equivalence. A response pattern of low reaction times on trials by the end of the training, slightly higher reaction times on the first trials of involving directly trained relations in the test, higher reactions times on symmetry trials, and still higher reaction time again on trials testing for transitivity and global equivalence was replicated from earlier studies. We also found that there were large differences between reaction times when comparing those participants who responded in accord with stimulus equivalence and those who did not.

## References

- Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*, 275-291. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- Arntzen, E., Grondahl, T., & Eilifsen, C. (in press). The effects of different training structures in the establishment of conditional discriminations and the subsequent performance on the tests for stimulus equivalence. *The Psychological Record*.
- Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*, 309-320. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record, 50*, 603-628. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- Arntzen, E., & Lian, T. (in press). Training conditional discrimination with and without familiar picture-stimuli as node. *The Psychological Record*.
- Bentall, R. P., Dickins, D. W., & Fox, S. R. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology, B: Comparative and Physiological Psychology, 46B*, 187-214. doi:10.1080/14640749308401085
- Buffington, D. M., Fields, L., & Adams, B. J. (1997). Enhancing equivalence class formation by pretraining of other equivalence classes. *The Psychological Record, 47*, 69-96. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior, 51*, 29-45. doi:10.1901/jeab.1989.51-29
- Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 183-204. doi:10.1901/jeab.1992.58-183
- Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8-12. Retrieved from <http://www.eahb.org/NewSitePages/BulletinHomepage.htm>
- Fields, L., Adams, B. J., & Verhave, T. (1993). The effects of equivalence class structure on test performances. *The Psychological Record, 43*, 697-712. Retrieved from <http://thepsy->

- chologicalrecord.siuc.edu/index.html
- Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior*, *64*, 129-145. doi:10.1901/jeab.1995.64-129
- Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J. et al. (1997). Using simultaneous protocol to study equivalence class formation: The facilitating effect of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior*, *67*, 367-389. doi:10.1901/jeab.1997.67-367
- Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of Research Methods in Human Operant Behavior* (pp. 229-262). New York: Springer.
- Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent non-equivalence. *The Psychological Record*, *48*, 81-110. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- 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. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- Pilgrim, C., & Galizio, M. (1996). Stimulus equivalence: A class of correlations, or a correlation of classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus Class Formation in Humans and Animals* (pp. 173-195). Amsterdam: Elsevier.
- Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior*, *60*, 571-585. doi:10.1901/jeab.1993.60-571
- Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior*, *71*, 195-214. doi:10.1901/jeab.1999.71-195
- 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. doi:10.1901/jeab.1999.72-117
- Saunders, R. R., & McEntee, J. E. (2004). Increasing the probability of stimulus equivalence with adults with mild mental retardation. *The Psychological Record*, *54*, 423-435. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and Integration of Behavioral Units* (pp. 213-245). Hillsdale, NJ: Erlbaum.
- Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejeune (Eds.), *Behaviour Analysis in Theory and Practice: Contributions and Controversies* (pp. 93-114). Hillsdale, NJ: Erlbaum.
- Sidman, M. (1994). *Equivalence Relations and Behavior: A Research Story*. Boston: Authors Cooperative.
- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*, 127-146. doi:10.1901/jeab.2000.74-127
- 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. doi:10.1901/jeab.1982.37-5
- Smeets, P. M., Dymond, S., & Barnes-Holmes, D. (2000). Instructions, stimulus equivalence, and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record*, *50*, 339-354. Retrieved from <http://thepsychologicalrecord.siuc.edu/index.html>
- Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, *65*, 643-659. doi:10.1901/jeab.1996.65-643
- Tomanari, G. Y., Sidman, M., Rubio, A. R., & Dube, W. V. (2006). Equivalence classes

with requirements for short response latencies. *Journal of the Experimental Analysis of Behavior*, 85, 349-369. doi:10.1901/jeab.2006.107-04

Wulfert, E., & Hayes, S. C. (1988). Transfer

of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, 50, 125-144. doi:10.1901/jeab.1988.50-125

---