EUROPEAN JOURNAL OF BEHAVIOR ANALYSIS 2011, **12**, 157 - 172 NUMBER 1 (SUMMER 2011)

Single-Subject Withdrawal Designs in Delayed Matching-to-Sample Procedures

Christoffer Eilifsen and Erik Arntzen
Akershus University College

In most studies of delayed matching-to-sample (DMTS) and stimulus equivalence, the delay has remained fixed throughout a single experimental condition. We wanted to expand on the DMTS and stimulus equivalence literature by examining the effects of using titrating delays with different starting points during the establishment of conditional discriminations prerequisite for stimulus equivalence. In Experiment 1, a variation of a single-subject withdrawal design was used. Ten adults were exposed to one condition where the delay titrated between 0 and 3000 ms and another condition where the delay varied from 5000 to 8000 ms. Subsequently, participants were re-exposed to the condition they had first experienced. Results show that starting the titration of the delay at 5000 ms may have had a facilitatory effect on stimulus equivalence responding for some participants. For several participants, however, performance was stable throughout the experiment, apparently not affected by either the variations of the starting point of the titrating delay or by previous exposure to stimulus equivalence training and test procedures. In Experiment 2, we examined the effects of being exposed to the same stimulus equivalence procedure three times, again with adults as participants. Results show that such extended exposure had very limited effects on stimulus equivalence responding. This indicates that single-subject withdrawal designs may be an appropriate approach for studying stimulus equivalence.

Key words: delayed matching-to-sample, titrating delay, stimulus equivalence, single-subject design, withdrawal design, college students

---

Stimulus equivalence refers to a certain pattern of responding that has not been directly taught, but seems to emerge in certain contexts from sets of learned relations between stimuli. A typical stimulus equivalence experimental procedure involves teaching a participant several conditional discriminations that are not currently part of his or her behavioral repertoire, and subsequently testing for stimulus equivalence among stimuli involved in these newly established stimulus classes. To qualify as stimulus equivalence, the derived responding must have the properties reflexivity, symmetry, and transitivity. Reflexivity refers to unreinforced matching of identical stimuli included among the stimulus relations trained as prerequisite for testing stimulus equivalence. Symmetry refers to unreinforced responding in line with the established stimulus classes when the conditional and discriminatory functions of the stimuli used in training of the conditional discriminations are reversed. Transitivity refers to responding indicating a more complex recombination of stimuli related in the prerequisite training. It is possible to test transitivity and symmetry simultaneously. Such a test has been referred to as a global equivalence test or just an equivalence test (Sidman, 1986; Sidman & Tailby, 1982).

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

Stimulus equivalence has mainly been studied as direct or systematic replications of the phenomenon under different circumstances with a relatively small N, or by across group comparison with a large number of participants. Some studies have exposed the same participant to more than one condition for within-subject comparison (e.g., Arntzen, Grondahl, & Eilifsen, 2010; Imam, 2006; Smeets & Barnes-Holmes, 2005). As far as we know, however, there are no studies employing the experimental logic of single-subject designs to study stimulus equivalence, for example by using a withdrawal design. This is in contrast to the norm in behavior analysis, where single-subject experimental designs have been dominant and considered the most suitable way to study the subject matter at hand (Glenn, Ellis, & Greenspoon, 1992; Sidman, 1960). In the first experiment of the current study, we wanted to use single-subject experimental logic to examine stimulus equivalence by exposing participants to two different stimulus equivalence procedures with repeated exposure to one of the procedures.

Training structures concern differences in how the prerequisite conditional discriminations are sequentially presented to the participant and how stimuli in each stimulus class are "linked" in stimulus equivalence procedures. It is common to distinguish between the three training structures one-to-many (OTM), many-to-one (MTO), and linear series (LS). The LS training structure involves the training of a series of conditional discriminations where the comparison stimulus from one of the prerequisite conditional discriminations to be learned serves as the sample in another (Saunders & Green, 1999). The use of an LS training structure has been shown to lead to a lower probability of stimulus equivalence performance compared to the other two training structures, especially when combined with a simultaneous protocol (Arntzen et al., 2010; Arntzen & Holth, 1997, 2000; Buffington, Fields, & Adams, 1997; Fields et al., 1997). In the simultaneous protocol, the test trials are not introduced before the prerequisite stimulus classes have been established (Fields, Landon-Jimenez, Buffington, & Adams, 1995; Imam, 2006). This expected lower yield of stimulus equivalence can allow for the study of variation in performance not possible to detect under other circumstances due a ceiling effect. For this reason, an LS training structure in combination with a simultaneous protocol was used in the current experiments.

Most research on stimulus equivalence has employed a simultaneous matching-to-sample (SMTS) procedure both during the training of the prerequisite conditional discriminations and during the test for stimulus equivalence. In an SMTS procedure the sample stimulus remains accessible to the participant as the comparison stimuli are presented. Delayed matching-to-sample (DMTS) procedures have been used during the training of conditional discriminations in a few studies. Here the sample will disappear and a fixed amount of time will pass until the comparison stimuli are presented. Some studies have found that the use of delays during the training will increase the probability of participants responding according to the defining relations of stimulus equivalence (Arntzen, 2006; Saunders, Chaney, & Marquis, 2005). In addition, there are indications that longer delays produce more stimulus equivalence outcomes than shorter delays (Arntzen, 2006). In these types of DMTS procedures, the delay between sample disappearance and the appearance of the comparison stimuli has remained fixed to a specific length of time throughout the training procedure. An alternative to fixed delays is a titrating delayed matching-to-sample (TDMTS) procedure. In a TDMTS procedure, the length of the delay is a function of the performance of the participant. Typically the delay will increase if the participant responds according to the experimenter defined contingencies and decrease if the participant does not respond according to these contingencies.

Such a procedure will expose the participant to a range of delays, and these delays will be introduced gradually as the participant is responding correctly according to the experimenter defined stimulus classes (Ferraro, Francis, & Perkins, 1971; Jarrad & Moise, 1970).

Arntzen, Eilifsen, and Vaidya (2009) compared the use of 100 ms, 3000 ms, and 12000 ms fixed delays, and 0–100 ms, 0–3000 ms, and 0–12000 ms titrating delays during the conditional discrimination training prerequisite for stimulus equivalence testing in a group study with adult participants. Note that the 100 ms fixed condition and the 0-100 ms titration condition are thought to be identical and to have a behavioral effect similar to a 0 ms delay, as the delay is too short to be discriminated by participants. The procedure was involved an OTM training structure, a training structure correlated with high stimulus equivalence outcomes. The specific values of the delays were chosen because there have been several reports of changes in priming effects when delays of 2000–3000 ms have been used. Priming effects refer to the enduring effects of stimuli after they have been removed, typically examined by looking at response latencies to a target stimulus in relation to a previously presented priming stimulus (Donahoe & Palmer, 1994). Several studies of priming show that the priming effects on the target response is greatly reduced when the target stimulus is not introduced until 2000–3000 ms after the removal of the priming stimulus. It has been suggested in cognitively oriented literature on priming that this is the result of the limits of visualization strategies used in the task, and that when longer delays than 2000–3000 ms are employed other problem solving strategies are used. These strategies are said to involve verbal behavior (Phillips & Baddeley, 1971; Posner, Boies, Eichelman, & Taylor, 1969). We were for this reason interested to see whether using a delay of 3000 ms would increase or decrease the speed of acquisition of the prerequisite conditional discriminations in stimulus equivalence procedures compared to using higher delays. In addition we wanted to see whether the use of this delay value would facilitate or retard performance on tests for stimulus equivalence. When using fixed delays in Arntzen et al. (2009), the 3000 ms delay lead to fewer participants responding according to stimulus equivalence compared to the other conditions. In the other conditions all participants displayed stimulus equivalence responding. When using titrating delays the results were not systematic in the same way, with some participants in all titrating conditions not responding according to stimulus equivalence. One possibility is that this difference is found because the TDMTS procedure employed in the study started with a 0 ms delay and subsequently titrated upward. Longer fixed delays, on the other hand, by definition start and remain at the same relatively high value.

Experiment 1

The purpose of Experiment 1 was to investigate whether differential effects on performance could be obtained when using sample-comparison delays titrating between 0 and 3000 ms (Condition A) or 5000 and 8000 ms (Condition B) during the training of the discriminations prerequisite for testing for stimulus equivalence. The specific values of the delays were chosen so that the results could be compared to the results obtained in the previously mentioned study by Arntzen et al. (2009) and the studies on priming by Phillips and Baddeley (1971) and Posner et al. (1969). This research question was examined using a variation of a single-subject withdrawal design. To address possible order effects, half of the participants were exposed to the conditions in one order, while the other half was exposed to the conditions in the reversed order. Specifically, half the participants first experienced the A condition, followed by the B condition, which was again followed by a repetition of the A condition.

The other half were first exposed to the B condition, then to the A condition, followed by the B condition.

Method

Participants and setting. Ten adults between the age of 20 and 28, three females and seven males, took part in Experiment 1. The participants were assigned the numbers 4063, 4064, 4066, 4068, 4069, 4070, 4075, 4076, 4079, and 4080. All participants were naïve concerning the experimental manipulations and stimulus equivalence in general. Participants were recruited by advertising the need for paid help in psychology experiments at a local student housing complex, and through personal connections. All participants were paid 100 kroner (approximately € 12) for each hour of participation. The experiments took place in a laboratory setting, where participants sat in an office cubicle situated in a larger room. Participants were seated by a desk, facing a window covered with blinds that were shut.

Apparatus and stimuli. Two HP Compaq nc6320 portable laptop computers equipped with 35.6 cm screens and a two-button touchpad were used to conduct the experiment. Software developed by Psych Fusion, Ltd. in collaboration with the second author, and especially designed to be used in stimulus equivalence experiments, was used to present stimuli, record the responding of the participants, and to administer Programmed consequences to the participants. 45 stimuli measuring between 3.9 cm and 2.2 cm in height and between 1.9 and 3.1 cm in width were presented on the screen against a white background. All stimuli were black and surrounded by a clickable area measuring 3.7 cm in height and 8.7 cm in width. The stimuli used are depicted in Figure 1. One stimulus, the sample, would appear in the middle of the screen, while three others, the comparisons, would appear in three of the four corners randomly leaving one corner blank on each tria. Programmed consequences consisted of text appearing in the middle of the screen. This consequence interval also included the display of a number indicating the cumulative number of correct responses. This number was displayed in the right bottom corner of the screen. Prior to and after the experiment the participants were given printed copies of the stimuli in the same size as they appeared on the computer screen for a sorting task.

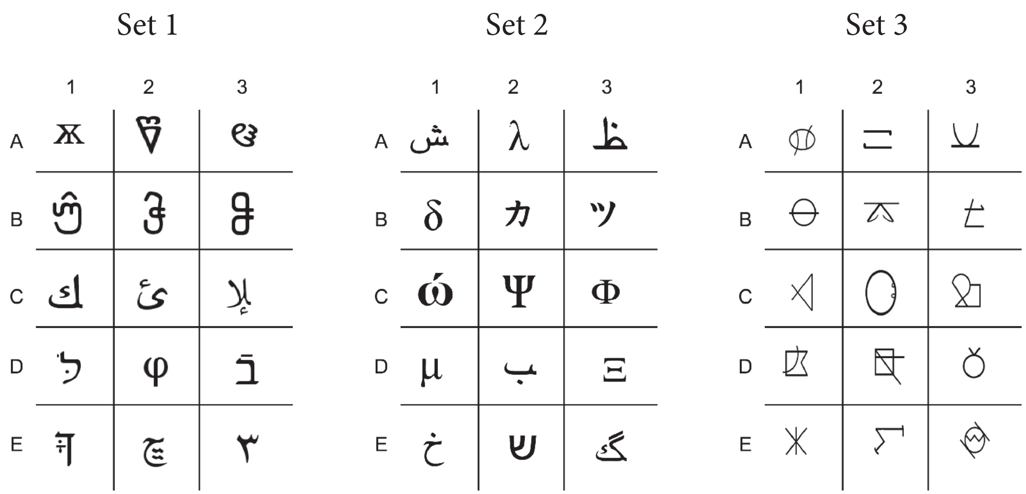

Figure 1. Stimuli used in the experiment. Set 1 was always employed in the first experimental condition. Set 2 was always used in the second condition, while set 3 was employed in the third condition. The numbers above each set indicate class membership, while the letters to the left of each set indicate members of each class.

All participants received instructions in English, with the exception of participant 4068, who were instructed in Norwegian."

Information and instructions given to participants. Upon arrival in the experimental setting, participants were instructed to read through a consent form that informed them about their anonymity and the possibility to withdraw from the experiment at any time. They were also told that data on their behavior during the experiment could be disseminated in the scientific community. Due to the comprehensiveness of the experimental procedure, participants were informed that they could choose whether they wanted finish their participation in one day or split it in into two consecutive days. They were also told that they could have short breaks at any time. The participants were told to consult the experimenter before leaving the experimental situation and ask whether it was a good time to do so. The experimenter would only recommend ending participation for the day between conditions. Short breaks would only be recommended in phases of the experiment that included programmed consequences, or between conditions. Following this information the participants were given printed copies of the stimuli to be used in the stimulus equivalence procedure and asked to categorize the stimuli as they liked. Participants were asked to categorize three sets of stimuli separately. Upon completion of this task the participants were instructed to have a seat in front of the computer. The following instructions were presented to the participants on the computer screen immediately prior to the start of conditional discrimination training:

> 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 told to read this text and to subsequently push a start button located on the bottom of the screen using the touch-pad.

Procedure. Each experimental condition included six phases. See Figure 2 for an overview of the experimental procedure. Phase 1 of the experiment was a TDMTS procedure where the delay between the disappearance of the sample and the appearance of the comparisons was titrated from a starting point to a maximum value. In Condition A the initial value was 0 ms and the maximum value 3000 ms, while in condition B the delay started at 5000 ms and had a maximum value of 8000 ms. Increases and decreases of the delay were 250 ms in both conditions, resulting in 12 steps of titration. See Figure 3 for an illustration of the titration procedure. Participants 4066, 4068, 4069, 4075, and 4080 were first exposed to Condition A, then to Condition B, followed by a repetition of Condition A (ABA-group). Participants 4063, 4070, 4076, and 4079 were first exposed to the B condition, then to the A condition, followed by a repetition of the B condition (BAB-group). Three different sets of stimuli were used, one set for each of the two different conditions and one for the re-exposure to the first condition. The three stimulus sets are depicted in Figure 1. An LS training structure was employed with a simultaneous protocol and a concurrent introduction of the stimulus relations. This meant that all AB, BC, CD, and DE trials were introduced randomly and mixed. Specifically, the following 12 trials were presented to the participants: A1B1B2B3, A2B1B2B3, A3B1B2B3, B1C1C2C3, B2C1C2C3, B3C1C2C3, C1D1D2D3, C2D1D2D3, C3D1D2D3, D1E1E2E3, D2E1E2E3, and D3E1E2E3.

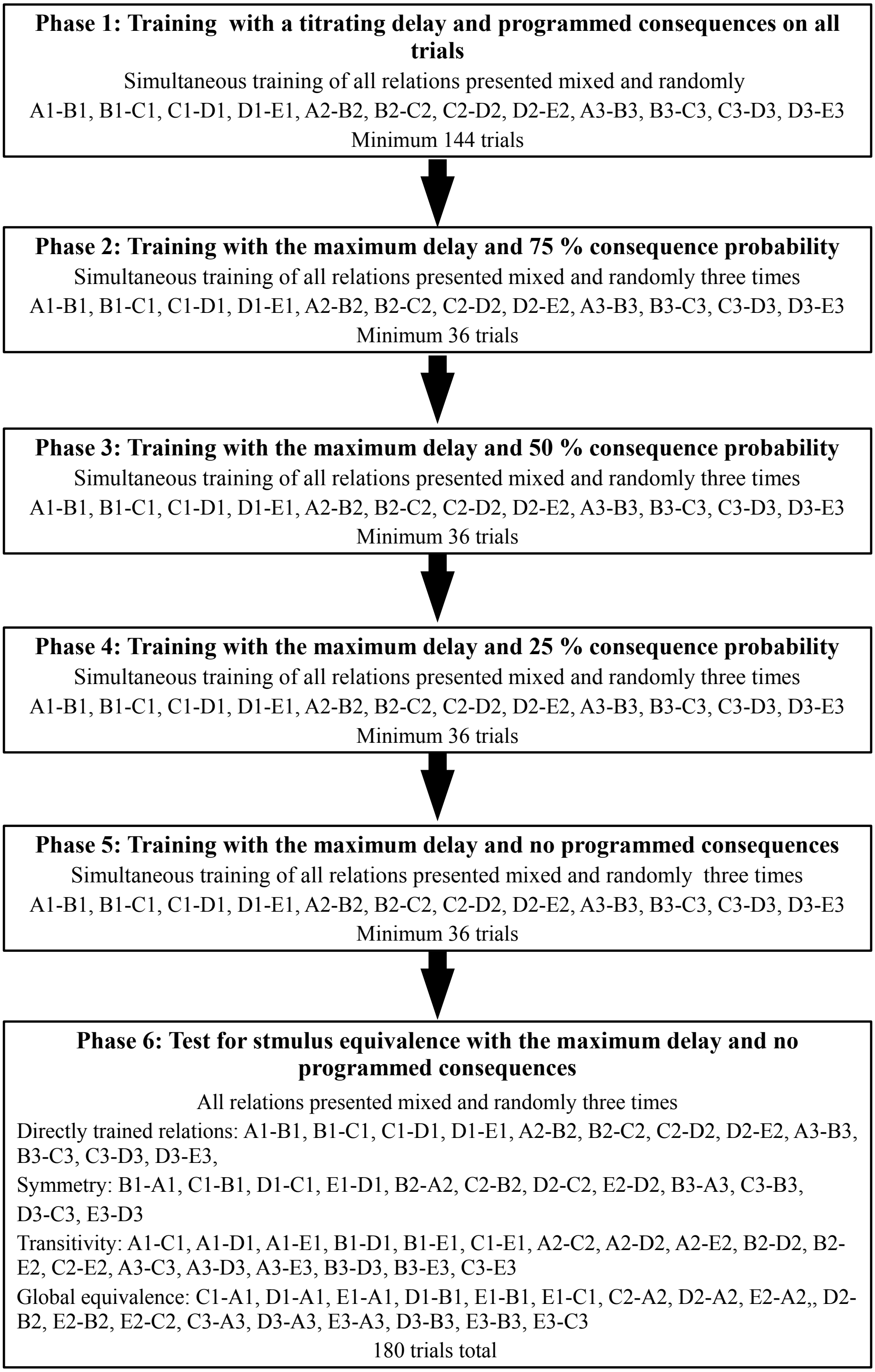

Figure 2. The figure shows the different phases of an experimental condition. The procedure was identical for all conditions, with the exception of variations in the starting point of the titrating delay between sample and comparisons.

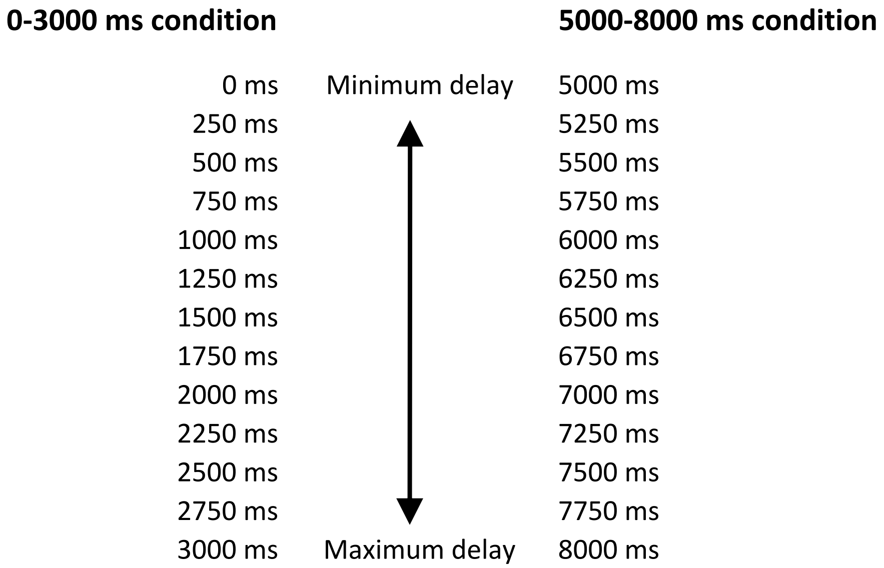

Figure 3. The steps of the delay for the two different experimental procedures are shown on the left and right of the figure. The double-headed arrow in the middle is meant to illustrate that both increments and decrements of the delay may occur as a function of participant performance.

The first alphanumeric code in each string of letters and numbers represents the sample stimulus on a given trial, while the underlined alphanumeric codes indicate the correct comparison stimulus. First, a sample stimulus appeared on the screen, and the participant had to respond to this stimulus by clicking on it using the computer mouse. Clicking on it would make the sample disappear and the comparisons appear after a certain amount of time. If the participant responded according to the stimulus classes meant to be established, the words "correct", "excellent", "super", "great", or "very good" appeared in the middle of the screen. If the participant made a response to a comparison stimulus not in the same experimenter-defined class as the sample, the word "wrong" appeared in this same spot as the positive feedback. In Phase 1 programmed consequences were provided on all trials. Both consequences indicating correct choices and consequences indicating incorrect responding lasted for 500 ms, and was followed by an inter-trial interval of 500 ms. Which comparison stimuli the participant responded to and response time to both the sample stimuli and comparison stimuli were recorded throughout the experiment. Experimenter defined class consistent performance was evaluated after one titration block, which consisted of a presentation of all 12 possible trials in a random order. Upward titration of 250 ms was initiated for one titration block if the participant responded according to the experimenter defined stimulus classes on 10 or more of the 12 trials during the previous block. Should performance in subsequent titration blocks fall below this criterion, there would be a 250 ms lower delay in the next block. The titration procedure continued until the maximum delay was reached (3000 ms or 8000 ms). When this maximum delay was obtained, the delay continued to be in effect on all trials for the remainder a training block. One training block consisted of three repetition of each possible trial type, resulting in 36 trials in each block. Specifically, this meant that the participant remained in Phase 1 for 0, 12, or 24 trials after reaching the maximum delay before performance was again evaluated, depending on what point in a training block the maximum titration value was reached. The next phase was initiated if the participant responded according to the experimenter defined stimulus classes on more that 90 % of trials (33 trials) over one training block of 36 trials. If not, Phase 1 would continue for another training block. During Phase 2 and in all subsequent phases, the maximum titration value remained in effect on all trials. In Phase 2, there was a 75 % probability of programmed consequences on any given trial. If participants continued to respond according to stimulus classes on over 90 % of trials in one training block, the next phase was started. If not, Phase 2 was continued for another training block of trials. In Phase 3, there was a 50 % probability of programmed consequences and in Phase 4 a 25 % probability. Phase shifts were also here initiated if the 90 % mastery criterion was reached, and the phase was continued if the mastery criterion was not met. Phase 5 included no programmed consequences. Mastery of the training trials with no programmed consequences over one training block, again meaning 90 % of trials corresponding to the experimenter defined classes, initiated the test for stimulus equivalence. The maximum delay value (3000 ms or 8000 ms) of the titration procedure remained in effect also in the test. In this test unreinforced trials that had been directly trained were interspersed in-between randomized presentations of trials testing for symmetry, transitivity, and global equivalence. All possible trials were repeated three times, resulting in 180 test trials. Stimulus equivalence responding was defined as responding in accordance with each of the relations tested for on more than 90 % of the trials testing for each relation. That meant that participants would have to respond in accordance with the directly trained relations on 33 out of 36 trials or more, respond according to symmetry on 33 out of 36 trials or more, and respond according to both transitivity and global equivalence on at least 98 out of the 108 trials testing for these relations for the performance to be considered an example of stimulus equivalence. Upon completion of the test, the participants were reminded of the possibility of continuing the next day and of the possibility of having a short break.

Table 1. *From the left to the right, the participant number, gender, and age of each participant is reported. This is followed by the number of training trials used during the first training procedure. Subsequently the number of responses according the directly trained relations, symmetry, and transitivity/equivalence out of all opportunities to respond in such a manner is presented. The number of training trials and test results are subsequently reported for the next two training and test procedures for each participant. Test results reported in bold text indicate responding within the experimenter defined mastery criterion for this particular relation.*

			A (0-3000 ms)				B (5000-8000 ms)				A (0-3000 ms)			
#	Gender	Age	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.
4066	M	26	612	**36/36**	**35/36**	**106/108**	648	32/36	**34/36**	93/108	720	**35/36**	**35/36**	92/108
4068	M	20	756	**33/36**	17/36	40/108	612	**34/36**	15/36	41/108	1908	31/36	12/36	40/108
4069	F	27	1620	25/36	16/36	53/108	792	21/36	18/36	54/108	684	**34/36**	27/36	69/108
4075	M	27	540	**35/36**	**36/36**	**105/108**	540	**35/36**	**35/36**	**105/108**	504	30/36	31/36	49/108
4080	M	24	1368	29/26	18/36	41/108	1080	24/36	18/36	38/108	468	**33/36**	32/36	42/108
			B (5000-8000 ms)				A (0-3000 ms)				B (5000-8000 ms)			
#	Gender	Age	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.
4063	F	23	540	**35/36**	30/36	84/108	504	**33/36**	31/36	81/108	468	**33/36**	**35/36**	**105/108**
4064	F	26	576	**34/36**	**35/36**	**103/108**	468	**36/36**	**36/36**	**106/108**	360	**35/36**	**35/36**	**101/108**
4070	M	27	864	**36/36**	**35/36**	**105/108**	432	**36/36**	**34/34**	79/108	396	32/36	32/36	77/108
4076	M	28	432	**35/36**	22/36	32/108	432	31/36	24/36	37/108	468	**33/36**	24/36	28/108
4079	M	23	504	**35/36**	**34/36**	**102/108**	396	**36/36**	**34/36**	**106/108**	504	**36/36**	**36/36**	**107/108**

If they chose neither, the next condition was started. When all three conditions were completed, the participants were given the physical copies of the stimuli again, and asked to categorize the three sets as they felt like now.

Results and Discussion

In the ABA-group Participant 4066 chose to split his participation into two days, going through the first A condition in one day and the subsequent B and A condition on the next day. Participants 4069 and 4080 also chose to divide their participation into two days, but these participants finished condition A and B on the first day, and then came back and went through re-exposure to condition A the next day. Participant 4068 and 4075 finished the experiment in one day. In the BAB-group Participant 4064 chose to go through the first exposure to condition B on one day, and then came back and finished the two other conditions on the next day. All other participants in this group concluded their participation in one day. In the sorting task of at the start of the experiment none of the participants categorized the stimuli according to any of the experimenter defined stimulus classes.

Table 1 displays the results for all participants and all the three conditions each participant was exposed to. One way to look at the results is to consider participants exposed to the different variations in the order of conditions as two different groups. 2 out of 5 participants responded according to stimulus equivalence when starting with the A condition (0–3000 ms titration), while 3 out of 5 did when starting with the B condition (5000–8000 ms titration). In the third condition, the reversal condition, no participants exposed to the A condition responded according to stimulus equivalence, while 3 out 5 participants in the B condition did so.

As the experiment was run as a withdrawal design, one may also look at the effect of the manipulated variables in the same subject. Participants 4066, 4068, 4069, 4075, and 4080 were exposed to the conditions in the ABA order. Participant 4066 used between 612 and 720 trials to complete each training procedure. In the first test the participant displayed stimulus equivalence responding. In the second and third test, however, the participant failed to respond according to stimulus equivalence. Participant 4068 used 756 and 612 trials to complete the first and second training procedure, respectively, while it took the participant 1908 trials to reach the test during the third training procedure. This participant did not respond according to stimulus equivalence on any of the tests. Participant 4069 used 1620 trials to complete the first training procedure, while finishing the training in less than half the number of trials in the second and third procedure. Participant 4069 did not respond according to stimulus equivalence on any on the tests. Participant 4075 used between 504 and 540 trials to reach the mastery criterion set for the training procedures. While this participant displayed stimulus equivalence performance on both the first and the second test, the participant did not do so on the third test. Participant 4080 completed the first and second training procedure in 1368 and 1080 trials, respectively. In the last training procedure the test was reached in 468 trials. This participant did not respond according to stimulus equivalence on any of the tests. Participants 4063, 4064, 4070, 4076, and 4079 were exposed to the conditions in the BAB order. Participant 4063 completed training in between 468 and 540 trials. No derived performance was detected on the first and second test, but on the third test she responded in line with the defining properties of stimulus equivalence. In the first, second, and third condition, participant 4064 used 576, 468, and 360 trials, respectively, to complete each training procedure. This participant responded according to stimulus equivalence on all the three tests. Participant 4070 completed the training procedure in the first condition in 864 trials, finished the second training procedure in 432 trials, and the third in 396 trials. In the first condition this participant performed according to stimulus equivalence, but he failed to do so on subsequent tests. Participant 4076 finished the training procedures in between 432 and 468 trials. No responding according to derived relations was observed during any of the tests. Participant 4079 used 504 trials to meet the mastery criterion for the training procedure in the first condition, 396 trials in the second procedure, and again 504 trials in the training procedure of the third condition. This participant responded according to stimulus equivalence in all tests.

Only two out of ten participants responded according to stimulus equivalence all three times they were exposed to a stimulus equivalence test. This can be considered a low yield of stimulus equivalence responding, and replicates earlier findings concerning the use of an LS training structure with a simultaneous protocol (Arntzen et al., 2010; Arntzen & Holth, 1997, 2000; Buffington et al., 1997; Fields et al., 1997).

The results may be used to argue that including a higher titrating delay in the prerequisite conditional discrimination training increase chance of stimulus equivalence responding. However, differences in behavior as a function of different experimental manipulations cannot be said to be very large. Participants exposed to the delay titrating from 5000 to 8000 ms in the first condition display stimulus equivalence responding to a larger extent than their counterparts starting with the 0 to 3000 ms delay (3 out of 5 participants, compared to 2 out of 5, respectively). In addition, after both groups are re-exposed to the starting condition in the last stimulus equivalence procedure, the results on the test for stimulus equivalence diverge even more, with none of the 0–3000 ms participants responding according to stimulus equivalence, while 3 out 5 still do so in the 5000–8000 ms condition. One may also argue that the prerequisite conditional discriminations are established quicker when participants start with a high titrating delay, and subsequently are exposed to a lower delay in the following condition. None of participants starting with the 5000–8000 ms condition used more that 1000 trials to reach the test, while this outcome occurred several times and with different participants in the group starting with 0–3000 ms condition. The average number of training trials is also markedly higher in all conditions for the participants starting with the 0–3000 ms condition. The results can be said to replicate the results of Arntzen (2006) where the inclusion of higher fixed delays during the conditional discrimination training was shown to be more effective in generating stimulus equivalence performance compared to lower delays. As the current study indicates that titrating delay procedures starting a 0 ms is less effective in generating stimulus equivalence outcomes than titrating delay procedures starting at a relatively high delay, the current results also support a conclusion that the different effects of fixed and titrating delays observed in Arntzen et al. (2009) were related to variations in the participants' experience with longer delays. Compared to procedures using a high fixed delay or titration procedures starting at a high delay, the experience with high delays is usually less in a titrating delay procedure starting at 0 ms, even when the procedure ends at a relatively high delay. This could explain the lower stimulus equivalence yields in the 0–12000 ms titrating condition compared to the 12000 ms fixed condition in the study by Arntzen et al. (2009). In both the current study and in Arntzen et al. (2009) the increases and decreases of the delay was governed by performance in blocks consisting only one repetition of each possible trial, resulting in relatively short number of trials with each delay of the titration procedure in effect. Results may be different if each titration block consists of more trials as this would increase the chance of the participants coming into contact with higher delays over more trials.

The data can hardly support an unambiguous conclusion about the superiority of using a titrating delay procedure starting at a 5000 ms compared to titrating delay procedure starting 0 ms in either generating stimulus equivalence performance or in establishing conditional discriminations quickly. For such a conclusion to be firmly drawn out of the single-subject experimental logic of the current study, a considerable change of behavior should have been seen in the second condition for the group starting with the 5000 ms to 8000 ms delays. That is, after being exposed to the 0–3000 ms delay training one should have seen less stimulus equivalence responding and a higher number of training trials compared to the preceding and subsequent conditions with the higher titrating delays. Similarly, participants starting with the 0–3000 ms delay should have performed according to stimulus equivalence on the second condition with the 5000 to 8000 ms titrating delay to a larger extent and the number of training trials should have been lower in this high-value titrating delay condition. A clear picture of considerably changed performance as a function of experimental conditions cannot be seen in the data. The most remarkable thing about these results is that there is no systematic positive effect on stimulus equivalence performance of previously being exposed to cycles of conditional discrimination training and tests for stimulus equivalence. This seems to be the case independently of which order of conditions the participants were exposed to. It is not unreasonable to assume that previous exposure to stimulus equivalence procedures would lead to higher stimulus equivalence outcomes on subsequent test. Only Participant 4063 responds in a way that can be argued to represent such a pattern. Most participants respond very similar on all three tests for stimulus equivalence, either by responding according to stimulus equivalence on all tests, as Participants 4064 and 4079, or by responding according to none of the relations tested for or only according to the directly trained relations, as Participants 4068, 4069, 4076, and 4080. The rest of the participants, 4066, 4070, and 4075, display a negative trend of stimulus equivalence performance in the sense that a stimulus equivalence outcome is observed in one or more of the first two conditions, but not during the final condition they were exposed to. The effects of several exposures to stimulus equivalence procedures alone may be obscured in Experiment 1, as the experiment included manipulations of the temporal value between sample disappearance and comparison appearance. To examine the effects of repeated exposure to stimulus equivalence procedures directly, a second experiment was conducted.

Experiment 2

It has been suggested that stimulus equivalence is difficult to study using single-subject designs, as there is a potential for carry-over effects from one condition to the other (Arntzen & Vaidya, 2008). The term carry-over effect is a very general term and can be defined as the influence exposure to one experimental condition has on experimental conditions that are introduced at a later point in time. Carry-over effects can make it difficult to assess the effects of manipulated variables in single-subject experimental designs, as the effects of the independent variable cannot be sufficiently isolated. The presence of carry-over effects have been pointed out as shortcoming of single-subject designs, especially by proponents of group-designs (Barlow, Nock, & Hersen, 2009). It does seem possible that something similar to a carry-over effect will also occur when exposing the same participant repeatedly to an identical stimulus equivalence procedure. Experiment 2 was carried out to examine effects of repeated exposure to stimulus equivalence procedures by having each participant take part in an identical test and training procedure three times.

Method

A 57-year-old woman and a 24-year-old man participated in Experiment 2. These two participants were assigned the numbers 4078 and 4082, respectively. All procedural variables were identical to Experiment 1 with one exception: In Experiment 2, the two participants were exposed to same titrating delay procedure three times. Participant 4078 was exposed to the 0–3000 ms titrating delay procedure three times, while Participant 4082 was exposed to the 5000–8000 ms titrating delay procedure. Note that three different sets of stimuli were used also in this experiment, one unique set for each time the procedure was run for each participant.

Results and Discussion

Participant 4078 went through the first stimulus equivalence procedure in one day and concluded her participation by completing the two additional test-training procedures the following day. Participant 4082 finished the experiment in one day. Neither of the two participants sorted the printed copies of the experimental stimuli according to any of the experimenter defined stimulus classes.

The results from Experiment 2 can be seen in Table 2. Participant 4078 completed the first training procedure in 2124 trials, the second in 864 trials, and the third in 1296 trials. This participant did not respond according to stimulus equivalence on any of the tests. Participant 4082 used between 324 and 432 trials to complete the training procedures. The participant responded according to stimulus equivalence on the second and the third test.

It can be argued that Experiment 2 replicated the results from Experiment 1 concerning the differential effects of high and low starting points of the titrating delay. Participant 4082, who experienced the 5000–8000 ms titration procedure three times, has a lower number of training trials in all procedures compared to Participant 4078, who was trained using the 0–3000 ms procedure. In addition, while derived responding is evident in all tests for the participant being trained and tested with the higher delay, such responding is completely absent for the participant experiencing the lower delay. The results from Experiment 2 also replicate the results from Experiment 1 in the sense that repeated exposure to stimulus equivalence training procedures and tests have limited influence on responding according to the defining properties of stimulus equivalence. For both participants, the same pattern of responding on stimulus equivalence tests is repeated all three times they experience the experimental procedure. With the exception of the lack of transitivity/equivalence responding on the first test of participant 4082, and the subsequent stimulus equivalence outcome on the latter tests, the repeated exposure to stimulus equivalence procedures does not seem to have either a negative or a positive effect on the derived performance of these two participants.

General Discussion

In Experiment 1 we examined whether differential effects on the behavior of the same participant could be obtained as a function of the starting point of titrating delays between sample disappearance and comparison presentation when establishing conditional discriminations prerequisite to tests for stimulus equivalence.

Table 2. See text for Table 1.

			A (0-3000 ms)				A (0-3000 ms)				A (0-3000 ms)			
#	Gender	Age	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.
4078	F	57	2124	28/36	25/36	58/108	864	**35/36**	31/36	71/108	1296	25/36	23/36	42/108
			B (5000-8000 ms)				B (5000-8000 ms)				B (5000-8000 ms)			
#	Gender	Age	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.	# of tr. trials	DT	Sym.	Trans./Global eq.
4082	M	24	432	**36/36**	**36/36**	95/108	360	**33/36**	**34/36**	**107/108**	324	**36/36**	**35/36**	**106/108**

The delays varied either from 0–3000 ms or from 5000–8000 ms. Results from Experiment 1 give some support to the notion that starting a titration procedure at a high delay has a facilitative effect on the speed of the establishment of conditional discriminations and on the formation of stimulus equivalence classes. The results cannot, however, be interpreted as an unambiguous demonstration of this, as considerable changes in the behavior of participants do not occur as a result of the exposure to the different experimental manipulations. The most interesting finding in Experiment 1 was that the performance of most of the participants was similar on all three tests for stimulus equivalence. Experiment 2 was set up in an effort to look directly at the effects of repeated exposure to the same stimulus equivalence procedure. The results indicate that previous exposure to the same stimulus equivalence procedure has limited impact when looking at performance in tests for stimulus equivalence in single-subject experimental procedures.

As there are some indication that DMTS training procedures with relatively high fixed delays facilitate stimulus equivalence responding (Arntzen, 2006), such procedure may be useful to apply as an alternative to SMTS in cases where stimulus equivalence does not easily emerge. Although there are no studies that directly explore this issue, it is not unreasonable to assume that some learners, as children or persons with learning disabilities, may have difficulties in initially mastering DMTS tasks with high delays. A gradual and individualized approach to high delay values in the form of titration may therefore be a very potent way to facilitate stimulus equivalence performance in these groups of learners. For this reason we think it could be a fertile venture to further explore how to maximize a facilitative effect of titrating delays on stimulus equivalence performance. Specifically we think that further studies should explore the effects of increasing the number of trials participants are exposed to at each step of the titration procedure in order to secure stable performance on a wide range of delay values at the pace of the individual learner.

Several studies have looked at the impact of previous exposure to stimulus equivalence procedures on subsequent stimulus equivalence responding. Some studies have used a simple-to-complex protocol to establish stimulus equivalence classes prior to training a second set of conditional discriminations using a simultaneous protocol. These studies have used a LS training structure. When employing this training structure the simple-to-complex protocol is correlated with a high number of participants responding according to stimulus equivalence, while the simultaneous protocol is correlated with low stimulus equivalence yields in combination with an LS training structure. Results from these studies with adult participants indicate that prior establishment of stimulus equivalence classes with a simple-to-complex protocol increase the probability of subsequent stimulus equivalence performance following simultaneous protocol training (Buffington et al., 1997; Fields et al., 1997). This phenomenon can be considered a carry-over effect. One related study also shows that the critical prerequisite for the establishment of stimulus equivalence classes was previous exposure to a test for derived responding that involved transitivity trials. Effects were greatly diminished when pre-training only consisted of conditional discrimination training (Fields et al., 2000). There have also been investigations into the development of conditional discriminations and stimulus equivalence responding without differential reinforcement following prior exposure to prerequisite training and tests for stimulus equivalence. When children and developmentally disabled adults have performed this task, the general finding is that responding consistent with the development of conditional discriminations does occur, and that stimuli involved in these unreinforced selections may be part of stimulus equivalence classes (Saunders, Drake, & Spradlin, 1999; Saunders, Saunders, Kirby, & Spradlin, 1988).

This suggests that there is a carry-over effect of the previous exposure to stimulus equivalence procedures in the sense that a behavioral pattern is continued in the unreinforced conditional discrimination task and the test for stimulus equivalence. These results can, however, be contrasted with a study by Holth and Arntzen (2000) where an adult participant was exposed to an identical stimulus equivalence procedure involving a simultaneous protocol a total of nine times. Following this, new conditional discriminations were trained using a different set of stimuli with an otherwise identical procedure, and a test for stimulus equivalence was again administrated. This participant did not respond according to stimulus equivalence on any of the tests despite the extensive experience with the procedure. Other studies that have not specifically looked at extended exposure to stimulus equivalence procedures, but have examined other experimental variables by exposing the same participant to different experimental conditions, have also found little effect of re-exposure. For example, Smeets and Barnes-Holmes (2005) exposed children to different combinations of training structures and stimulus materials in a stimulus equivalence procedure. Subsequently some of the children were trained and tested again using a different training structure and a different type of stimuli than they had previously been exposed to. All children who failed to show stimulus equivalence responding during the first test, also failed to do so after a second training procedure, and only 3 out of 4 children who responded according to stimulus equivalence during the first test still did so in the second procedure.

One question is whether the current results have come about because effects of extended exposure to stimulus equivalence procedures are limited or because some behavior is established during the first test that leads to a lack of necessary behavior change on the subsequent tests for stimulus equivalence. One line of research that may suggest the latter is a series of studies by Pilgrim and Galizio (1990, 1995) on the effects reversing directly trained contingencies on later stimulus equivalence tests. In these studies adult participants were exposed to changed contingencies in the prerequisite conditional discriminations following the establishment of stimulus equivalence classes. After the reversal training, which involved the same stimuli as in the previously established stimulus equivalence classes, particpants were again tested for stimulus equivalence. While novel conditional responding was established and symmetry performance changed according to the reversed contingencies, responding remained consistent with the original stimulus equivalence classes on trials testing for transitivity. Pilgrim and Galizio (1995) suggests that the results may have occurred because some sort of rule-governance established during the first test made participants insensitive to prevailing contingencies during the second test for stimulus equivalence. If so, one is actually dealing with a carry-over effect that may inhibit stimulus equivalence class formation for some participants and not a lack of effect of extended exposure to stimulus equivalence training and test procedures. The contingency reversal studies must necessarily, however, employ the same stimulus set in all training procedures and tests for stimulus equivalence. Because of this the results may not be relevant to take into account to explain the results in current experiments, where new stimulus classes involving novel stimuli are established in each of the stimulus equivalence procedures.

Further studies on the use of sample-comparison delays in stimulus equivalence procedures should include using fixed delays of similar delay values as the spectrum of titrating delays used in the current study. This would illuminate the possible differential effects of using titrating and fixed delays in stimulus equivalence procedures. One may also conduct studies that include both conditions with titrating delays and conditions with fixed delays in a within-subject experimental design.

One obvious way to further approach the issue of the effects of repeated exposure to stimulus equivalence procedures is to run direct replications of Experiment 2 to see if the outcome will be similar. A systematic replication should also be conducted where the same stimulus set is used in all three training-test repetitions. In general, those participants who do not respond according to stimulus equivalence in stimulus equivalence experiments should more often be re-exposed to the training procedure and then tested again for derived relations. This will provide further knowledge of the effects of repeated exposure to stimulus equivalence procedures, which would be of great value if one wishes to study the phenomenon using single-subject designs.

References

Arntzen, E. (2006). Delayed matching to sample: Probability of responding in accord with equivalence as a function of different delays. *The Psychological Record, 56*, 135–167. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

Arntzen, E., Eilifsen, C., & Vaidya, M. (2009). *On titrating of delays in DMTS and SE.* Paper presented at the Annunal Convetion of the Association for Behavior Analysis, Phoenix, AZ.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). 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, 60*, 437–462. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

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., & Vaidya, M. (2008). The effect of baseline training structure in equivalence class formation in children. *Experimental Analysis of Human Behavior Bulletin, 29*, 1–8. Retrieved from http://www.eahb.org/NewSitePages/Bulletin-Homepage.htm

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for behavior change*. Boston, MA: Allyn and Bacon.

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

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior (V. P. Dorsel, Ed.)*. Boston, MA: Allyn and Bacon.

Ferraro, D. P., Francis, E. W., & Perkins, J. J. (1971). Titrating delayed matching to sample in children. *Developmental Psychology, 5*, 488–493. doi:10.1037/h0031598

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 facillitating 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

Fields, L., Varelas, A., Reeve, K. F., Belanich, J., Wadhwa, P., Derosse, P. et al. (2000). Effects of prior condtional discrimination traning, symmetry, transitivity, and equivalence training on the emergence of new equivalence classes. *The Psychological Record, 50*, 443–466. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

Glenn, S. S., Ellis, J., & Greenspoon, J. (1992). On the revolutionary nature of the operant as a unit of behavioral selection. *American Psychologist, 47*, 1329–1336. doi:10.1037/0003-066X.47.11.1329

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

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107–124. doi:10.1901/jeab.2006.58-04

Jarrad, L. E., & Moise, S. L., Jr. (1970). Short-Term Memory in the Stumptail Macaque: Effect of physical restraint of behavior on performance. *Learning and Motivation, 1*, 267–275. Retrieved from http://www.elsevier.com/wps/find/journaldescription.cws_home/622909/description#description

Phillips, W. A., & Baddeley, A. D. (1971). Reaction-time and short-term visual memory. *Psychonomic Science, 22*, 73–74. Retrieved from http://www.psychonomic.org/psp/publications-resources.html

Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior, 54*, 213–224. doi:10.1901/jeab.1990.54-213

Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63*, 225–238. doi:10.1901/jeab.1995.63-225

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology: General, 79*, 1–16. doi:10.1037/h0026947

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizens. *The Psychological Record, 55*, 539–559. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

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., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior, 50*, 145–162. doi:10.1901/jeab.1988.50-145

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books.

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., & 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., & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes, 69*, 281–293. doi:10.1016/j.beproc.2004.12.009