Response-to-Sample Requirements in Conditional Discrimination Procedures

Erik Arntzen, Live F. Braaten, Torunn Lian, and Christoffer Eilifsen
Oslo and Akershus University College

The present experiment examined number of training trials as a function of a required response to sample in establishing conditional discriminations and testing for responding accordance with stimulus equivalence. Reaction times were also examined. Twenty participants in 4 different experimental groups were compared in a between-groups design. The participants were trained to form 4 classes of 3 stimuli each with a linear series (LS) training structure. Unreinforced trials for directly trained relations were included in the test. Results from the present study show that a required response to sample increased chances to learn the discriminations more quickly. This may imply that the required response can increase the observing behavior. Required response to sample did not affect responding according to emergent relations in the test in any substantial way. Reaction time data show that there are systematic differences in reaction latency depending on the trial type in the test, and that incorrect responses may not be randomly selected.

Key words: required response to sample, stimulus equivalence, conditional discrimination, reaction time, adults

---

Any acquired behavior whose primary function is to affect the sensing of stimuli is called an 'observing response' (Donahoe & Palmer, 1994). When Wyckoff (1952) discussed the role of observing responses in discrimination learning, he referred to "any responses which result in exposure to the pair of discriminative stimuli involved" (p. 431). Furthermore, Wyckoff emphasized that it is important to distinguish these observing responses from the responses on which reinforcements are based. One can increase attending by altering the contingencies so that attention is directed to the relevant stimuli in the complex environment (McIlvane, Dube, & Callahan, 1996). Observing responses, and their effect on learning, can be studied in conditional discrimination training. In some discrimination training procedures the participants can be required to respond to the discriminative stimulus.

A possible outcome of conditional discrimination training is formation of equivalence classes. Stimuli within a class are said to be equivalent when they are interchangeable with one another (Green & Saunders, 1998). The properties used to determine whether the directly trained relations are equivalent are reflexivity, symmetry, and transitivity (Sidman & Tailby, 1982).

Three training structures used in conditional discrimination procedures are called one-to-many (OTM), many-to-one (MTO), and linear series (LS) (e.g., Saunders, Saunders, Williams, & Spradlin, 1993). The LS structure has shown to be the least effective in producing stimulus equivalence relations (e.g., Arntzen, Grondahl, & Eilifsen, 2010),

Author note: Correspondence concerning this manuscript should be addressed to Erik Arntzen, Oslo and Akershus University College, Department for Behavioral Science, PO Box 4 St. Olavs Plass, 0130 Oslo, Norway. E-Mail: erik.arntzen@equivalence.net.

and it is sometimes used in experiments to reduce ceiling effects. The combination of a simultaneous protocol and an LS training structure can be useful when studying variables that influence the formation of equivalence classes, as in the present study.

Conditional discrimination procedures could be arranged as either SMTS (simultaneous matching to sample) or DMTS (delayed matching to sample). In SMTS, the sample will remain present on the screen after the presentation of comparison stimuli, while in DMTS, sample stimuli disappear *n* seconds before the onset of comparison stimuli (Arntzen, 2006). Furthermore, in standard MTS and DMTS procedures, the sample observation period continues until the participants touch the sample stimulus, and is followed by the presence of the comparison stimuli, sometimes with different delays (Dube & McIlvane, 1999).

Whether or not a response to the sample stimulus is required has differed in the published experiments within the area of conditional discriminations and stimulus equivalence. When there is not a required response to sample, the sample is usually present *n* seconds before the comparisons appear (sample duration, SDur). Combining SMTS, 0-s delay matching, and requiring or not requiring an observing response can result in several combinations: (1) required response to sample (RRS) and SMTS, (2) RRS and 0-s delay, (3) no required response to sample (NRRS) and SMTS, (4) NRRS with *n* SDur and 0-s delay, and (5) NRRS *n* SDur and SMTS.

Touching of the sample is called an observing response (Sidman & Tailby, 1982). Research has been conducted to study the role of observing responses in discrimination learning, both in non-humans and in humans. Eckerman, Lanson, and Cumming (1968) studied the effect of observing responses in an MTS procedure in pigeons. In their first experiment, they examined matching without a required response to sample and compared the results with those of Cumming and Berryman (1965), who used the same procedure with a required response to sample. The results showed that MTS performance increased when an observing response was required, and that the pigeons needed fewer training sessions to establish matching with a required response to sample than without. Maki, Gillund, Hauge, and Siders (1977) studied the effect of extinction of observing responses on matching performance in pigeons and found that observing responses did have an effect on matching accuracy, that is, increasing the response requirement on the sample key might improve performance by increasing the duration of sample-stimulus presentation.

Spetch and Treit (1986) also studied the effect of manipulating the number of observing responses required in DMTS procedure with pigeons. During training, the pigeons learned to make a required response to an initiating stimulus (IS)—a white light—to produce the sample stimulus—a red or blue light. The sample stimulus terminated after 5 s independently of a response, and the comparison (red and blue light) followed immediately. In the test, Spetch and Treit examined whether an increase in the number of required responses to the IS or to the sample stimulus would have an effect on the accuracy of matching. They found that in both the first and the second experiment, accuracy was affected by the number of required responses to the sample stimulus but not by the number of required responses to the initiating stimulus. "The finding is … consistent with the interpretation that larger sample-response requirements facilitate accuracy because of the resultant increased exposure to the sample" (Spetch & Treit, 1986, p. 24). Similar results have been found with hens (Foster, Temple, Mackenzie, Demello, & Poling, 1995).

Carlin, Wirth, and Chase (1998) studied the effects of sample-response requirements on MTS performance with humans.

Four college students were trained in conditional discrimination. Each participant was presented with two conditions. In one condition, the participants were not required to respond to the sample stimulus, and the sample and the comparison were presented simultaneously. In the other condition, the samples were presented after a variable amount of time (on average, 1.5 s), and a single response on the keyboard produced the comparison. The results from this study show that conditional-discrimination performance improved when a response to sample was required. The participants required fewer trials to reach 90% accuracy when they had to respond to the sample than when they were not required to respond to the sample.

Lian and Arntzen (2010) examined the function of response requirement versus no response requirement on number of trials to mastery the conditional discriminations and the formation of stimulus equivalence classes. In their study, six participants were exposed to two conditions, one with RRS and one with NRRS. Half of the participants started with the RRS condition, and the other half started with the NRRS condition. The results showed that number of trials to criterion was lower for the participants who started with the RRS condition than for those who started with the NRRS condition. There were no noticeable differences in responding according to equivalence relations between the two conditions.

When matching-to-sample performance and stimulus equivalence have been studied, the focus has usually been on establishing the conditional relations and the percentage correct responding of emergent relations in test trials. But as Dymond and Rehfeldt (2001) pointed out, it might be useful to look at other means of measuring and recording emergent stimulus relations. One way of measuring and recording responding is to examine reaction time, or response latency. Reaction time is measured from the appearance of the comparison stimuli to the selection of a comparison stimulus. In studies where reaction time have been recorded, it has been found that there are differences in reaction time in baseline and symmetry testing trials and differences in reaction time in transitivity and global equivalence trials, across response accuracy (Dymond & Rehfeldt, 2001). A number of studies have shown an increase in reaction time from baseline trials to trials of emergent relations, where there was an increase in reaction time from the directly trained relations to the symmetrical relations, and a further increase in reaction time from symmetrical relations to equivalence relations (Arntzen, Galaen, & Halvorsen, 2007; Arntzen et al., 2010; Eilifsen & Arntzen, 2009; Spencer & Chase, 1996). The same pattern has been shown in both the beginning and the end in tests, with an overall decrease in reaction time later in the tests. Research has also found that there are differences in reaction time depending on the different delays in DMTS, and that there is slower responding in longer delays, without an increase in errors (Baron & Menich, 1985). Various explanations for these systematic differences in reaction time, especially from the beginning to the end of tests for emergent relations, have been proposed. Some researchers explain that the test forces some sort of 'precurrent' behavior or problem-solving behavior, especially in the beginning of the test (Arntzen et al., 2007; Holth & Arntzen, 2000). Furthermore, the decrease in reaction time later in the test can be the result of a shift to more direct stimulus control (Arntzen et al., 2007).

As we mentioned above, procedures vary with regard to the use of a required response to sample in conditional discrimination training; therefore, the main purpose of the present experiment is to examine whether a required response to sample would have an effect on the outcome of matching and the number of training trials required to reach criterion. This experiment will also look at how—or if—a required response to sample affects responding in accordance with stimulus equivalence.

In addition, we asked if there is a difference between the results on the test in an MTS format compared to a sorting task involving the same stimuli. Reaction time differences will also be examined in the beginning and at the end of training, and in tests for emergent relations, within and across groups. Finally, reaction time differences between those who respond in accordance with stimulus equivalence and those who do not will be examined.

Method

Participants

The participants were 20 adults, 22 to 39 years old (mean age = 28 years), 5 males and 15 females, all acquaintances of the second author. None of the participants were familiar with this kind of experimental setting or the stimuli used, nor did they have any knowledge about stimulus equivalence. Participants were assigned to four different groups. Additionally, they were informed that they could withdraw from the experiment at any time. They were thoroughly debriefed about the experimental purpose after the experiment.

Settings

The experiment was conducted in a 6.45 m^2 (2.77 m × 2.33 m) study room. The room contained one table and chair on the left side, one table and chair in the back of the room for the experimental computer, and one table and chair on the right side for a computer that was turned off during the experiment. The participants faced the back wall and had their backs to the door during the experiment.

Apparatus and Stimuli

Two HP EliteBook 8740w personal computers were used in this experiment. Both computers had an Intel Core i5 2.40 GHz-520M processor. The computers had a 17.0-in. screen and an external mouse connected to it through one of the computer's USB ports. In the present experiment, we used a customized software package, developed in collaboration with the first author. Two different sets of stimuli were used, as shown in Figure 1. Both of the stimuli sets consisted of 12 arbitrary black stimuli on a white background, randomly arranged in potentially four classes, with three members in each class. The size of the stimuli varied from 0.4 in. to 0.9 in. in width and 0.5 in. to 1.3 in. in height. The sample stimulus was always presented in the middle of the screen, and the comparison stimuli were presented in the corners of the screen, in a random position for each trial. A click-sensitive area, invisible to the participants, measuring 1.7 in. x 1.7 in. surrounded all stimuli. In the lower-right corner of the screen, the number of correct responses was displayed for each correct trial, which was scheduled for programmed consequences.

Design

This experiment was arranged as a between-groups design, with five participants in four groups. Group 1 was exposed to an RRS (required response to sample stimulus) and SMTS (simultaneous matching to sample) procedure. The required response was to click on the sample stimulus that appeared in the middle of the screen. When the participant clicked on the sample stimulus, the four comparison stimuli appeared while the sample stimulus still remained on the screen. Group 2 was exposed to a 0-s delayed matching procedure, with an RRS. The participants were required to click on the sample stimulus, and then the sample stimulus disappeared and the comparison stimuli appeared after a 0-s delay. Group 3 was exposed to an SMTS procedure with NRRS (no required response to sample stimulus), where the sample stimulus and comparison stimuli appeared on the screen simultaneously. Group 4 was exposed to an NRRS with a 2-s SDur (duration of sample stimulus) and a 0-s delayed matching procedure.

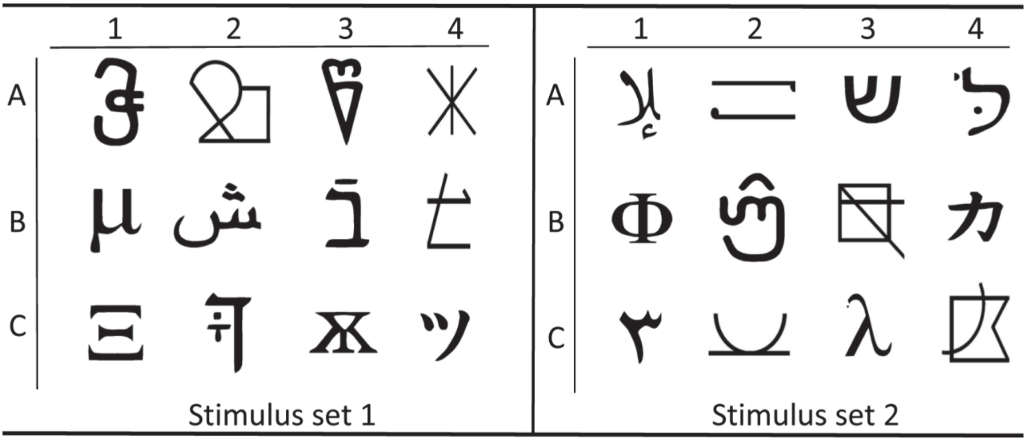

Figure 1. The two sets of stimuli used in the experiment. For each of the stimulus sets, the numbers indicate the experimenter-defined classes and the letters indicate the different members.

The sample stimulus appeared on the screen for 2 s, and the participants were not required to click on the sample. After the 2 s, the sample stimulus disappeared and the comparison stimuli appeared with a 0-s delay.

Procedure

General information to participants. When the participants arrived at the experimental setting, they were seated at the table to the left in the room and asked to read an information sheet with a consent form. This information sheet informed them that they were about to take part in an experiment within behavior analysis. They were also told that the experiment would involve sitting in front of a computer and being presented with different stimuli, and that the experiment would last approximately 1.5 hrs. After the participants read the information sheet, the experimenter emphasized verbally that the results were anonymous and that they could withdraw from the experiment at any time.

Categorization of stimuli cards by a sorting task. Before and after the experimental session, the participants were given laminated printouts of the stimuli and were asked to categorize them. The printouts were of the approximately the same size as the stimuli presented on the screen. The experimenter took a picture of the stimuli and noted the categorizations.

Instructions. The participants were seated in front of the computer and presented with instructions. The participants were given slightly different instructions, depending on the experimental group they were assigned to. Participants in Groups 1 and 2, experiencing RRS and SMTS and RRS and 0-s DMTS, respectively, were presented with the following text: "A stimulus will appear in the middle of the screen. Click on this by using the computer mouse"; this was followed by text below. Participants in Groups 3 and 4, experiencing NRRS and SMTS and NRRS, and 2 s SDur and 0-s DMTS, respectively, were presented with the following text**:** "A stimulus will appear in the middle of the screen. You don't need to click on this." followed by the text below:

Four other stimuli will appear. Choose one of these by clicking on it with the mouse. If you choose the stimuli that we have defined as correct, words like *good, excellent*, and so on will appear on the screen. If you press incorrectly, the word *wrong* will appear on the screen. At the bottom of the screen, the number of correct responses will be counted. During the experiment, the computer will not give you feedback as to whether your choices are right or wrong. However, based on what you have learned, you can get all of the tasks right. Please, try your hardest can to get everything right. Good luck!

If the participants did not have any questions after reading the instructions, they were told to press the "start" button to begin the experiment.

The procedure in the present experiment included three phases: (1) training of baseline conditional discriminations, (2) thinning of programmed consequences, and (3) testing for emergent relations. See Table 1 for an overview of the different experimental phases.

Phase 1: Acquisition—Training of baseline conditional discriminations. Participants were trained according to an LS training structure, and with a simultaneous protocol, implying that all the conditional discriminations were established before the test for emergent stimulus relations. The conditional discriminations were presented concurrently, implying that all trial types were presented from the beginning of the conditional discrimination training: A1B1B2B3B4, A2B1B2B3B4, A3B1B2B3B4, A4B1B2B3B4, B1C1C2C3C4, B2C1C2C3C4, B3C1C2C3C4, and B4C1C2C3C4. The first letter–number combination in each string represents the sample stimuli, and the underlined letter–number combination represents the correct comparison defined by the experimenter. All of these training trials were randomly presented, and the locations of correctly

Table 1. *Overview of Experimental Phases.*

Experimental phases	Trial types	Programmed concequences	Minimum # of trials
Cycle 1:			
Acquisition: Training of baseline conditional discriminations			
Trial types presented in random order	A1B1B2B3B4, B1C1C2C3C4, A2B1B2B3B4, B2C1C2C3C4, A3B1B2B3B4, B3C1C2C3C4, A4B1B2B3B4 and B4C1C2C3C4	100%	40
Maintenance: Thinning of programmed consequences			
75% probability	A1B1B2B3B4, B1C1C2C3C4, A2B1B2B3B4, B2C1C2C3C4, A3B1B2B3B4, B3C1C2C3C4, A4B1B2B3B4 and B4C1C2C3C4	75%	40
25% probability	A1B1B2B3B4, B1C1C2C3C4, A2B1B2B3B4, B2C1C2C3C4, A3B1B2B3B4, B3C1C2C3C4, A4B1B2B3B4 and B4C1C2C3C4	25%	40
0% probability	A1B1B2B3B4, B1C1C2C3C4, A2B1B2B3B4, B2C1C2C3C4, A3B1B2B3B4, B3C1C2C3C4, A4B1B2B3B4 and B4C1C2C3C4	0%	40
Test for derived stimulus relations	Directly trained trial types		
All trial types presented in random order	A1B1B2B3B4, B1C1C2C3C4, A2B1B2B3B4, B2C1C2C3C4, A3B1B2B3B4, B3C1C2C3C4, A4B1B2B3B4 and B4C1C2C3C4	0%	40
	Symmetry trials		
	B1A1A2A3A4, C1B1B2B3B4, B2A1A2A3A4, C2B1B2B3B4, B3A1A2A3A4, C3B1B2B3B4, B4A1A2A3A4 and C4B1B2B3B4	0%	40
	Transitivity trials		
	A1C1C2C3C4, A2C1C2C3C4, A3C1C2C3C4 and A4C1C2C3C4	0%	20
	Equivalence trials		
	C1A1A2A3A4, C2A1A2A3A4, C3A1A2A3A4 and C4A1A2A3A4	0%	20
Cycle 2:			
The experimental phases was repeated if participants did not respond in accordance with stimulus equivalence in Cycle 1			

Note. The table shows the different experimental phases, the trials types presented, the probability of programmed consequences and the minimum number of trials in each phase. For each trial type the stimulus serving as sample is presented first and the correct comparison stimulus is the one underlined.

defined comparisons were randomly located in the four corners. Each of the trial types was presented five times; consequently, each block consisted of 40 trials. When the participants clicked on the correct comparison defined by the experimenter, words like *good*, *excellent,* and *correct* appeared on the screen. When they clicked comparisons from one of the other classes, the word *wrong* appeared on the screen. Each programmed consequence lasted 500 ms and was followed by an ITI lasting 1,000 ms

The criterion to proceed to the next phase of the experiment was a minimum of 90% correct comparison choices in one training block, constituting a minimum of 36 out of 40 correct comparison choices. If the participants did not reach this criterion, the training phase was repeated until the criterion was met.

Phase 2: Maintenance—Thinning of programmed consequences. Programmed consequences were scheduled for all comparison choices in the first experimental phase. The participants remained in one block until they reached performance criterion, defined by the experimenter as 36 out of 40 correct trials (90%). When this criterion was reached, the participants continued with training blocks with reduced programmed consequences. First, there was a 75 % chance of programmed consequences on any given trial. When the participants reached criterion on this phase, they received 25% programmed consequences, and after that, 0% programmed consequences. Not meeting the criterion in one of the blocks led the participants to go through the same block again. When the participants reached the mastery criterion on the last block with no programmed consequences, the test for stimulus equivalence was introduced.

Phase 3: Test for emergent relations. In the test for emergent relations, the participants experienced symmetry trials, transitivity trials and equivalence trials in random order with the directly trained conditional discriminations interspersed. Each trial type was presented five times implying that participants experienced a total of 120 trials during the test, whereof 40 were directly trained conditional discriminations, 40 were symmetry trials, 20 were transitivity trials and 20 were equivalence trials. The different trials are shown in Table 1. No programmed consequences were scheduled for test trials.

Definitions of responding in accordance with stimulus equivalence. When tested, the participants had to meet a performance criterion to be said have to responded in accordance with stimulus equivalence. They had to master 36 out of 40 trials involving directly trained relations, 36 out of 40 on the symmetry trials, and 36 out of 40 trials involving transitivity and global equivalence relations combined. If the participants reached the test criterion, the experiment was over; if they did not reach criterion, participants were exposed to a new cycle with training and test. If they did not reach the performance criterion the second time, the experiment was ended and they were not exposed to a new training and test cycle. Another experimental criterion was that if the participants did not reach mastery criterion on the directly trained relations in the first test, their data were not included in later analyses. If the participants did not have the directly trained relations intact, the probability of responding correctly to the emergent relations would be low.

After finishing the experiment the participants were debriefed and were told the purpose of the experiment. They were also shown the data from their experiment and offered an introductory article on stimulus equivalence written by the first author.

Dependent variables and recording. The dependent variables in this study were number of trials in conditional discrimination training, number of participants responding in accordance with stimulus equivalence, and reaction time to comparison stimuli.

Furthermore, responses in the card sorting tasks. All data, except for the categorization tests, were automatically registered by the software and summarized by the second author. Reaction time to comparison stimuli was measured as the time from the presentation of the comparison stimuli to when a response to a comparison stimulus occurred. The median reaction time for the last five training trials, the first five test trials, and the last five test trials was calculated for each participant and averaged across participants. In addition, the median reaction times for the first five training trials in the second training block and for the last five training trials before thinning of programmed consequences were calculated for each participant and averaged across participants.

Results

Participants 4618 and 4627 did not reach criterion for the directly trained relations in the test and will not be included in the analysis of the results. The other participants did reach criterion for the directly trained relations in the test and finished the experiment with one or two cycles of training and test (see Table 2). In Group 1, RRS and simultaneous matching, the median number of responses in the acquisition phase was 400 responses. Four of 5 participants had 320 to 480 training trials in the acquisition phase, while Participant 4614 had 720 training trials. In Group 2, RRS and 0-s delay, the median of responses in the acquisition phase was 520.

Table 2. *An Overview of the Results*

				1st Cycle				2nd Cycle			
	Participant			Trials	Number of correct choices in test			Trials	Number of correct choices in test		
	#	Gender	Age	Aquisition	DT	SY	TRANS/EQ	Aquisition	DT	SY	TRANS/EQ
Group 1	4616	F	26	320	**40**	**40**	**40**	NA	NA	NA	NA
RRS and	4615	F	26	400	**40**	**40**	**38**	NA	NA	NA	NA
Simultaneous	4621	F	22	360	**40**	**40**	**36**	NA	NA	NA	NA
Matching	4623	F	34	480	**39**	**39**	**39**	NA	NA	NA	NA
	4614	F	29	720	**40**	**39**	13	160	**40**	**36**	15
Median:				400							
Group 2	4613	F	26	520	**39**	**40**	**36**	NA	NA	NA	NA
RRS and 0 s	4617	F	27	360	**38**	**38**	27	160	**40**	**40**	35
DMTS	4624	M	32	560	**40**	34	35	160	**40**	**40**	**39**
	4620	M	32	400	**40**	14	19	160	**40**	**40**	**40**
	4625	M	33	1000	**40**	14	19	160	**39**	**40**	**38**
Median:				520							
Group 3	4610	F	26	520	**39**	**40**	**40**	NA	NA	NA	NA
NRRS and	4626	F	39	640	**39**	**40**	**37**	NA	NA	NA	NA
Simultaneous	4605	F	26	480	**36**	**39**	30	160	**40**	**39**	**39**
Matching	4612	M	29	680	**39**	**39**	29	160	**40**	**40**	33
	4628	M	22	760	**39**	23	10	200	29	32	28
Median:				640							
Group 4	4619	F	26	560	**40**	**40**	**40**	NA	NA	NA	NA
NRRS, 2 s SDur	4603	F	30	640	**40**	**38**	**40**	NA	NA	NA	NA
and 0 s DMTS	4611	F	25	520	**39**	**40**	**38**	NA	NA	NA	NA
	4608	F	27	560	**39**	**37**	**38**	NA	NA	NA	NA
	4622	F	26	320	**39**	**39**	**39**	NA	NA	NA	NA
Median:				560							

Note. The table shows the individual results in the different Groups. Test results are reported as correct comparison choices for the different trial types. Maximum number of each trial type in the test is 40. DT denotes directly trained conditional discriminations when interspersed in test, SY denotes symmetry trials and TRANS/EQ denotes combined transitivity and equivalence trials. The bolded numbers indicate that the mastery criteria were met.

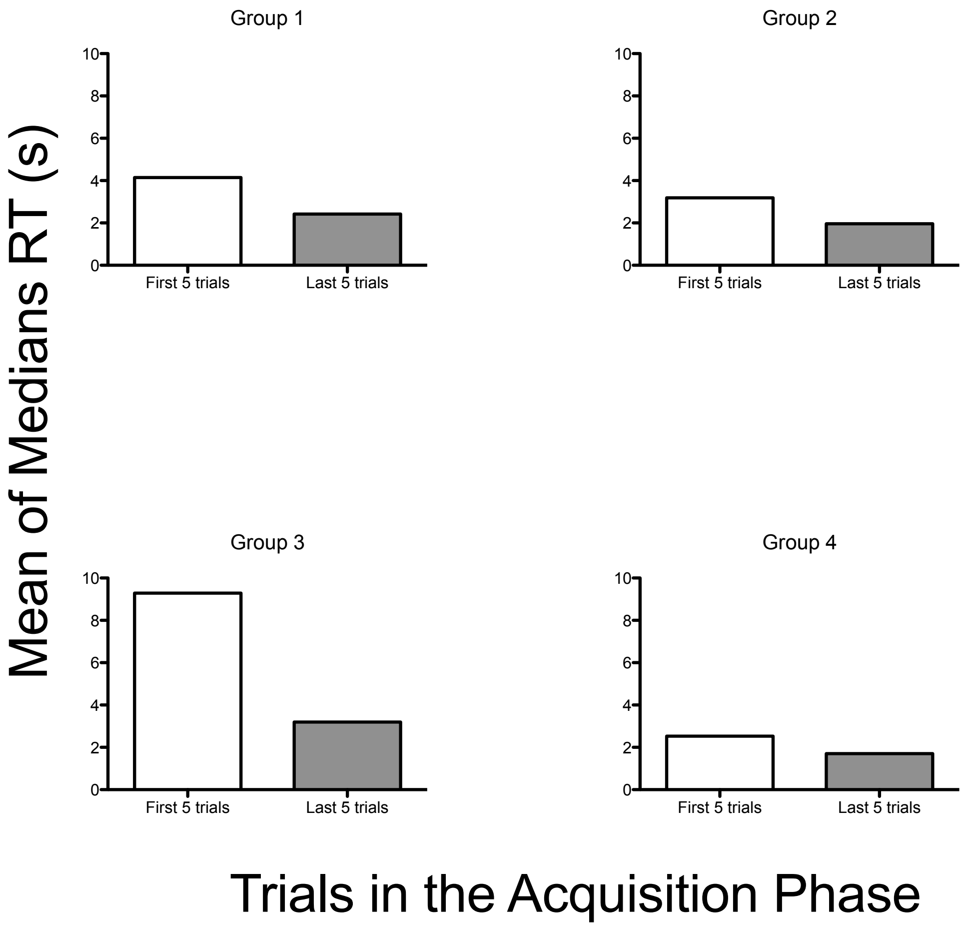

Figure 2. The mean of medians reaction time to comparison stimuli in training for all groups. Group 1 is RRS, simultaneous. Group 2 is RRS, 0-s delay. Group 3 is NRRS, simultaneous. Group 4 is NRRS, SDur 2 s. Reaction time was calculated from the first five trials in the second training block and the last five trials in the last training block before reducing the probability of programmed consequences.

Participant 4625 had 1,000 training trials; the other participants used between 360 and 560 trials. In Group 3, NRRS and simultaneous matching, the median number of responses was 640 in the acquisition phase. Participant 4628 had the highest number of responses, 760, in the acquisition phase, while the rest of the participants had 520 to 680 responses. In Group 4, NRRS and 2-s SDur and 0-s delay, the median was 560. Participant 4622 had the lowest number of baseline conditional discrimination trials across all groups, with 320 trials to reach the mastery criterion for baseline conditional discriminations. Participant 4603 had the highest number of trials in this group, with 640 training trials.

Four of 5 participants, 4615, 4616, 4621, and 4623, responded in accordance with stimulus equivalence in the first training cycle in Group 1. Participant 4614 did not respond in accordance with stimulus equivalence in either Cycle 1 or 2. In Group 2, one Participant, 4613, responded

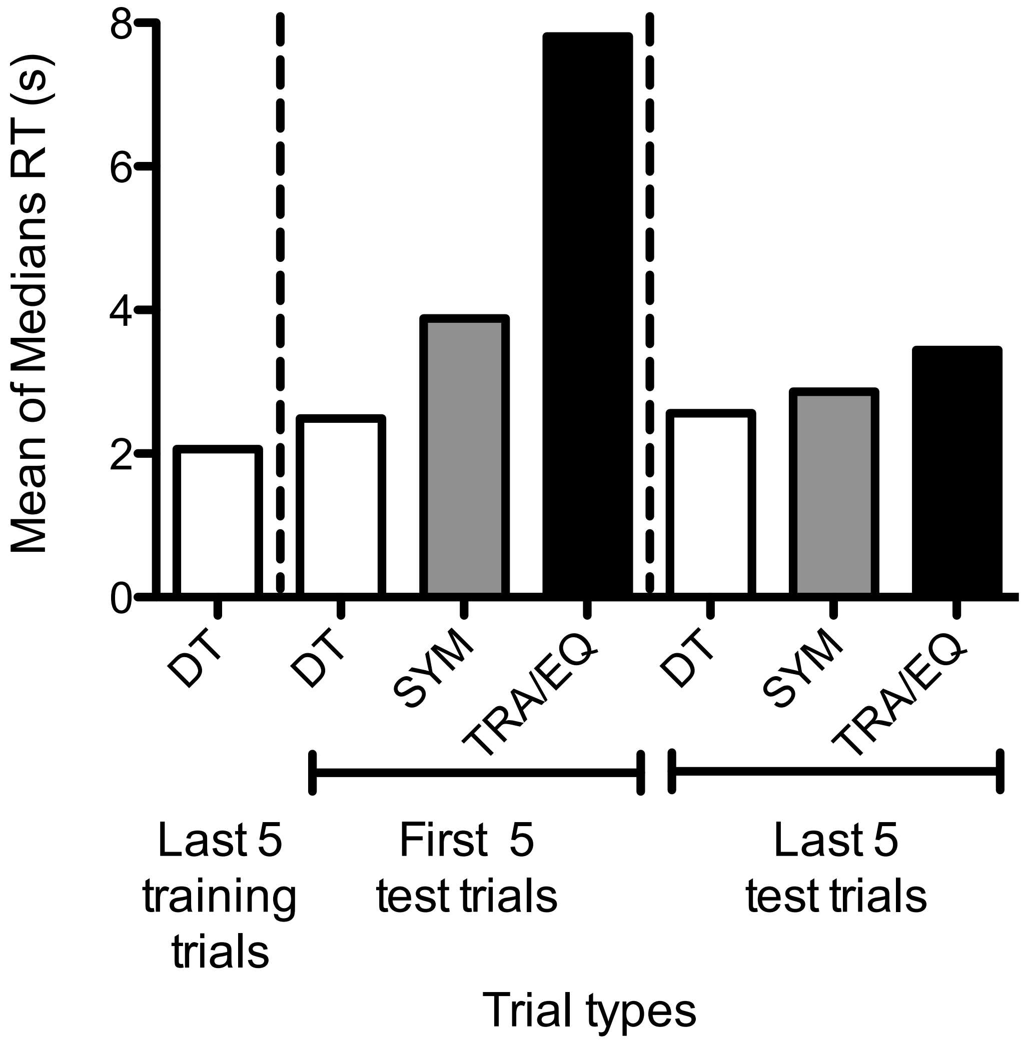

Figure 3. The mean of medians reaction time to comparison stimuli, across groups, for the last 5 training trials (DT), directly trained trials in test (DT), symmetry trials (SYM), and transitivity/ equivalence trials (TRA/EQ) in test.

in accordance with stimulus equivalence. When training and test were repeated in Cycle 2, Participants 4625, 4620, and 4624 responded in accordance with stimulus equivalence, while Participant 4617 did not. In Group 3, 2 of 5 participants, 4610 and 4626, responded in accordance with stimulus equivalence. When training and test were repeated in Cycle 2, Participant 4605 responded in accordance with stimulus equivalence, while Participants 4612 and 4628 did not. In Group 4, 5 of 5 participants responded in accordance with stimulus equivalence.

Figure 2 shows the mean of medians of reaction times to comparison stimuli for the first five training trials in the second training block and the last five trials before thinning of programmed consequences. The reaction time to comparison stimuli in early training phases in Group 1 was 4.15 s; in Group 2 was 2.53 s; in Group 3 was 9.29 s, and in Group was 4, 3.19 s.

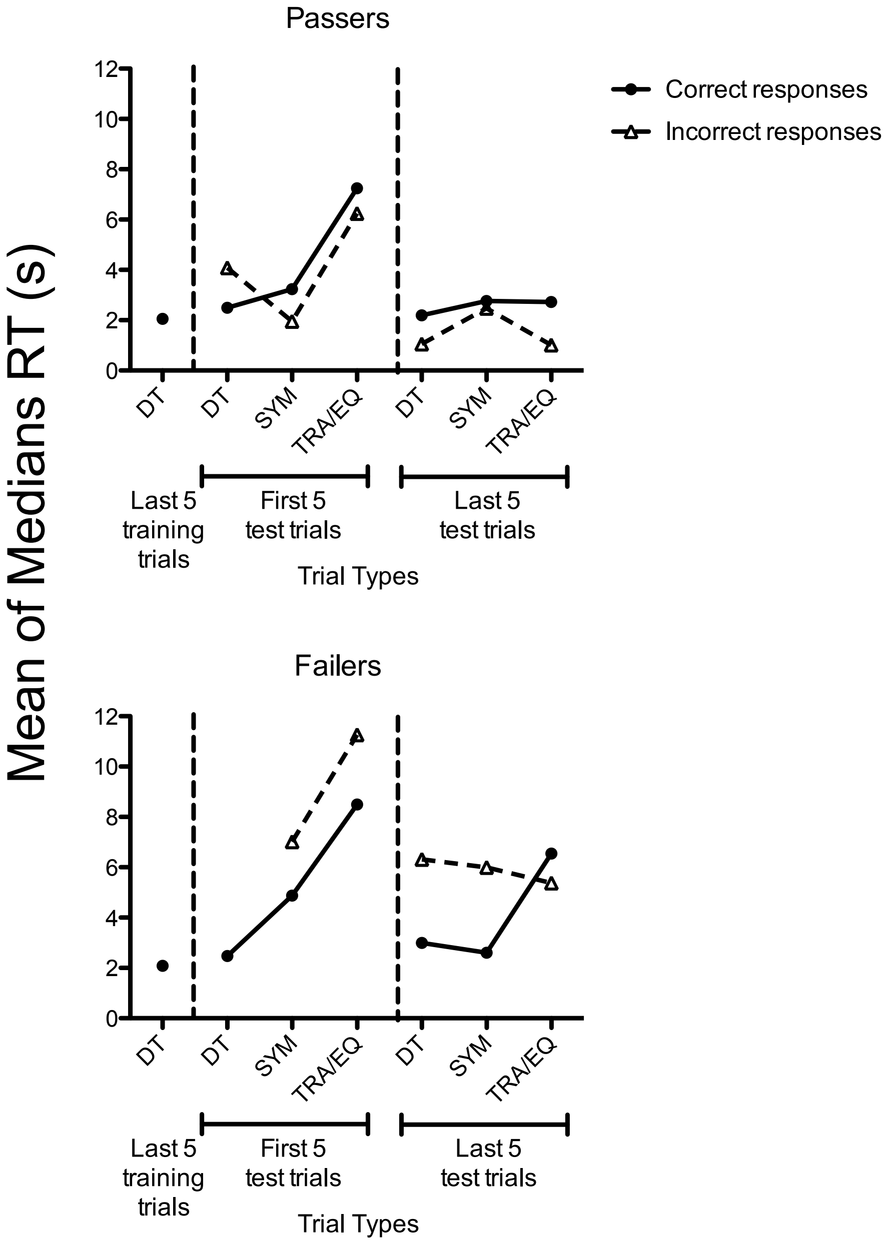

Figure 4. The mean of medians reaction time to comparison, across groups, for the five last training trials (DT), the directly trained trials (DT), symmetry trials (SYM), and transitivity/equivalence trials (TRA/EQ). The upper panel shows reaction times for participants who responded in accordance with stimulus equivalence, and the lower panel shows reaction times for participants who did not respond in accordance with stimulus equivalence. There are no datapoints for incorrect responses for directly trained trials in training and for "failers" for DT for first 5 trials because there were no incorrect responses.

The reaction times to comparison stimuli in late training in Group 1, 2, 3, and 4 were 2.42 s, 1.7 s, 1.2 s, and 1.96 s, respectively.

Figure 3 shows the mean of medians of reaction times to comparison stimuli for the last five training trials, the first five test trials, and the last five test trials for all participants, regardless of group. The results are presented across groups, because the pattern was the same for all groups. The reaction time to comparison stimuli for the last five trials in training was 2.1 s. The figure shows that there is an increase in reaction time from training trials to first test trials, for both symmetry trials and the combined transitivity and equivalence trials. Symmetry trials increased to 3.88 s, and combined equivalence trials increased to 7.81 s. At the end of the test, the reaction time to comparison stimuli decreased to 2.86 s for symmetry trials and to 3.44 s for the combined transitivity and equivalence trials. The reaction time to comparison stimuli for the directly trained conditional discriminations increased to 2.49 s for the first five test trials and to 2.86 s for the last five test trials.

Figure 4 shows the mean of medians of reaction times to comparison stimuli for correct and incorrect responses, regardless of experimental group. The data are summarized for the last five training trials, first five test trials, and last five test trials in the first training and test cycle and is in addition presented for each type of relational type involved in the test. Among the participants who responded in accordance with stimulus equivalence, their reaction time to comparison stimuli was shorter at correct directly trained trials than at incorrect directly trained relations in the first part of the test. For all other trial types, the participants responded more quickly at incorrect choices than at correct choices in both the first and the last part of the test. The reaction time to comparison stimuli of the participants who did not respond in accordance with stimulus equivalence, was shorter at correct trials than at incorrect ones. The only exception is the equivalence trials in the last part of the test, where reaction time to comparison stimuli was shorter for the incorrect comparison choices.

None of the participants categorized the laminated stimuli in accordance with the experimenter-defined classes in the pre-categorization test. In the post-categorization task all but one participant, Participant 4608, categorized the laminated stimuli in accordance with the experimenter-defined classes. Participant 4608 did, however, respond in accordance to stimulus equivalence.

Discussion

In the present experiment, we asked whether a required response to sample stimulus would have an effect on the number of training trials to criterion in conditional discrimination training and responding in accordance with stimulus equivalence. Conditional discriminations were established in an LS training structure, examining two conditions with a required response to sample and two conditions without required response to sample, and two of those conditions with simultaneous matching and two conditions with 0-s delayed matching. Furthermore, we wanted to examine reaction time for correct and incorrect responses for those who responded in accordance with stimulus equivalence and those who did not.

The results from the present experiment show that in the groups with a required response to sample, fewer training trials were required to reach the 90%-accuracy criterion, compared to the groups without a required response. The results also show that simultaneous matching with an RRS involved a lower number of training trials to criterion than 0-s delayed matching with an RRS. Simultaneous matching with NRRS gave a slightly higher number of training trials to criterion, compared to 0-s delayed matching with NRRS. When comparing across all groups, simultaneous matching with a required response was most efficient in establishing the conditional discriminations, and the groups with simultaneous matching without required response to sample were least efficient. These results replicate the findings of Lian and Arntzen (2010), and they also support earlier findings from research done with both humans (Carlin et al., 1998) and nonhumans (Eckerman et al., 1968; Foster et al., 1995; Spetch & Treit, 1986). An interpretation of these results is that when participants were required to touch the sample stimulus, visual contact with the conditional stimulus was more likely, as compared with groups where there was no required response to sample. This does not, however, necessarily imply that the sample stimulus was observed for a longer period of time in the RRS groups than in the NRRS groups. For example, Palmer (2010) emphasized the importance of expanding experimental procedures in the science of human behavior, and one of the expansion tools he discussed was eye-tracking. Dinsmoor (1985) described *observing behavior* as any auxiliary behavior that increases the participant's contact with the discriminative stimuli, and an auxiliary behavior can be eye movements. While eye movements cannot explain our responses, they do cause the image of the sample stimulus to reach the fovea and can be considered one of the "behavioral events that might serve as more invariant elements of relational responding" (Palmer, 2010, p. 40). Eye-tracking can therefore provide a more straightforward means of studying the relations between MTS accuracy and the participant's observing behavior than in the present study. For example, Dube et al. (2006) studied DMTS accuracy and eye movements and found, among other things, that participants with high and low accuracy made similar numbers of observations per trial, but participants who had high accuracy had a longer sample-stimulus observing duration. Dube et al. (2010) studied observing behavior in conditional discrimination training and found that participants with high accuracy scores had reliable observing responses to all stimuli. Participants with intermediate accuracy scores showed failure to observe the sample stimuli and had relatively brief observation durations. When given different interventions to improve observing behavior, the observing duration and matching accuracy increased.

Even though a response to a sample stimulus is an indirect way to measure observing behavior, the present results show that responding to sample stimulus does increase chances to learn the discriminations faster, and the required response may increase observing behavior. Thus, altering contingencies to increase observing behavior increased the participants' contact with the sample stimulus and enhanced discrimination. Interestingly, the experimenter observed in the present study that some of the participants in Group 3 did show overt observing behavior where no observing response was required. In this group, the sample stimulus and the comparison stimuli were presented simultaneously. The participants were observed to move the mouse cursor over the sample stimulus before moving the mouse out to select a comparison stimulus. This overt observing behavior was repeated on the next trial. This new overt behavior was topographically the same as for those who were required to respond to a sample stimulus. Unfortunately, this was observed by coincidence and was not systematically reported, but it should be further examined when using a simultaneous matching procedure without a required response to sample stimulus.

Reaction time data, displayed in Figure 2, show that there was higher reaction time to comparison stimuli in the beginning of the training in Group 3, NRRS simultaneous matching, than in all the other groups. At the end of training, the reaction time was approximately the same in all four groups. In Group 3, the sample stimulus and the comparison stimuli were presented on the screen simultaneously, while in the other groups the sample stimulus was presented alone on the screen before the comparison stimuli were presented. The participants in Groups 1, 2, and 4, therefore, had the opportunity to observe the sample stimulus for *n* second before the comparison stimuli were presented. Because reaction time to comparison stimuli was measured as the time from when the comparison stimuli were presented on the screen to when a comparison choice occurred, the amount of time the sample stimulus was presented was not included for Groups 1, 2, and 4 and is likely to have influenced the higher reaction times to comparison stimuli in Group 3. One question that arises from this is whether the differences in reaction time in the beginning of the training were due only to procedural differences in the amount of contact the participant had with the sample stimulus before the appearance of comparison stimuli. Even though 2 s were subtracted from the reaction time in the beginning of the training in Group 3, to equalize sample observation duration, a difference of a little more than 6 s was still found. Therefore, one might speculate that participants in Group 3 used more time to direct attending to the relevant stimulus than participants in the other groups. This can be studied more closely with eye-tracking, to obtain information about how long the participant fixates on one stimulus and about eye movement between the stimuli, as mentioned above.

When looking at reaction time in training and test across all groups, shown in Figure 3, there is a slight increase in mean of medians of reaction times from the last five trials in training to the first five trials in testing for the directly trained relations in the test. The mean of medians of reaction times on the trials testing for symmetrical relations was higher than on trials testing directly trained relations. A further increase was observed in the mean of medians of reaction times in trials testing transitive and global equivalence relations. In the last five trials testing the same relation, there was an average decrease in mean of medians of reaction times for all relations. However, the differences were the same when comparing the three relations in the last trials, where the mean of medians of reaction times was slightly higher on trials testing symmetrical relations than on directly trained relations, and lower than trials testing for transitivity and global equivalence. These data are a replication of other studies examining reaction time in trials testing for stimulus equivalence relations (e.g., Arntzen et al., 2007; Arntzen et al., 2010; Arntzen & Hansen, 2011, in press; Eilifsen & Arntzen, 2009; Spencer & Chase, 1996).

Also, when looking at those participants who responded in accordance with stimulus equivalence (see Figure 4), the same patterns are displayed, especially in the first five trials testing the directly trained and the emergent relations. The same pattern can also be seen in participants who did not respond in accordance with stimulus equivalence, in the beginning of the test. Figure 4 also shows the differences in reaction time for correct and incorrect responding. Interestingly, those who failed the test had longer reaction times to comparison in incorrect responding than correct responding except on the last five trials testing equivalence relations. This finding is in accordance with other research (e.g., Dixon, Rehfeldt, Zlomke, & Robinson, 2006). This might indicate that the participant does not randomly select a comparison stimulus when responding incorrectly. If one assumes that there is some sort of 'precurrent' behavior going on when reaction times increase, this would also account for those who respond incorrectly. One can interpret reaction time for incorrect responding in three ways: The behavior can be under stimulus control that is different from what was defined by the experimenter (e.g., McIlvane, Serna, Dube, & Stromer, 2000) or the participants' behavior is 'random'. This is only speculation and should be investigated further by a closer analysis of the participants' individual responding.

Usually the LS training structure gives a low outcome of stimulus equivalence test when using the simultaneous protocol (e.g., Arntzen et al., 2010; Arntzen & Hansen, 2011; Fields et al., 1997).

It can be difficult to explain the high equivalence outcome on the test in the present study, but it might be because of the low number of participants in each group and a result of coincidence. Whether or not and how a required response would have an effect on the performance on the test is difficult to say, since all of the participants reached the mastery criterion in the training and therefore necessarily had learned the conditional discriminations before the test, regardless of what group they were in. Further research should also focus on MTO and OTM, because in MTO, many samples are trained to one comparison, while in OTM, one sample is trained to many comparisons. Such differences could influence the role of an observing response.

In addition, the higher outcome on the equivalence tests for participants in Group 4, needs commented on. The only condition that did not enable the participant to observe the sample stimulus alone on the screen was the group with simultaneous matching without a required response (Group 3). In the other groups the sample was present for a variable amount of seconds, and with or without a response to the sample, the comparisons appeared on the screen. With regards to Group 4, could the presence of the sample stimulus alone on the screen facilitate some sort of observing behavior that differed topographically from the observing behavior defined in this experiment? Some sort of observing behavior must have been emitted other than responding to the sample stimulus, as the participants in the two groups without a required response to sample did learn the discrimination and therefore in fact observed the sample stimulus. Further research with eye-tracking equipement, as mentioned earlier could give some useful answers.

The categorization data are quite interesting. The data show that the participants did correct categorization of the stimuli cards even if they did not show formation of equivalence classes during the testing that is in accordance with other studies (Fields, Arntzen, Nartey, & Eilifsen, in press). The categorization of the stimuli cards is an easier task because the participant can scan all the stimuli at once, while in tests based on an MTS format, one trial with one sample and four comparisons, as in the present study, is presented at a time. Further research should focus on how useful such categorization tasks are and how they can be used.

There are some limitations with the present study. First, we assigned the participants to different groups as they arrived at the lab. We wanted to ensure that we had about the same number of participants in each group at any given time. However, it is difficult to see how such an assignment system could have influenced the results. Second, we had only 5 participants in each group. In further research we should increase the number of participants in each group, or rather, arrange it as a within-subject design, to eliminate any coincidences that may might arise in data with only five participants in each group. Third, because of software restrictions it was not possible to arrange the MTS task so that it included a presentation of a sample stimulus, NRRS, 2-s SDur and SMTS. Such a condition might be easier to compare to Group 1 in the present experiment. Further research should include such a condition. In addition, replications of the experiment should be conducted with other populations, for example, people with autism and developmental disabilities, children, and persons with dementia. Research with other populations would give us a broader view of observing behavior and might have an application value.

In sum, the results from the present study showed that a required response to sample established the conditional discriminations in fewer trials than when a response to sample was not required, and suggest that the required response may increase observing behavior. The results also showed that a required response to sample did not affect responding according to the emergent relations in the test in any substantial way.

Data show that there are systematic differences in reaction time, depending on the trial type in the test, and that incorrect responses are not randomly selected. Further research on conditional discrimination and observing behavior should also include eye-tracking. Unmasking the variables that underlie observing behavior and conditional discrimination training is of great importance in understanding human behavior.

References

Arntzen, E. (2006). Delayed matching to sample and stimulus equivalence: 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., Galaen, T., & Halvorsen, L. R. (2007). Different retention intervals in delayed matching-to-sample: Effects of responding in accord with equivalence. *European Journal of Behavior Analysis, 8*, 177–191. Retrieved from http://www.ejoba.org/

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., & Hansen, S. (2011). Training structures and the formation of Equivalence Classes. *European Journal of Behavior Analysis, 12, 483–503.* Retrieved from http://www.ejoba.org/

Baron, A., & Menich, S. R. (1985). Reaction-times of younger and older men: Effects of compound samples and a pre-choice signal on delayed matching-to-sample performances. *Journal of the Experimental Analysis of Behavior, 44*, 1–14. doi: 10.1901/jeab.1985.44-1

Carlin, L. A., Wirth, O., & Chase, P. N. (1998). Effects of sample response requirements on matching-to-sample performance with humans. *Experimental Analysis of Human Behavior Bulletin, 16*, 2–5. Retrieved from http://www.eahb.org/NewSitePages/BulletinHomepage.htm

Cumming, W. W., & Berryman, R. (1965). The complex discriminated operant: Studies of matching-to-sample and related problems. In D. I. Mostofsky (Ed.), *Stimulus generalization* (pp. 284–330). Stanford, CA: Stanford University Press.

Dinsmoor, J. A. (1985). The role of observing and attention in establishing stimulus-control. *Journal of the Experimental Analysis of Behavior, 43*, 365–381. doi: 10.1901/jeab.1985.43-365

Dixon, M. R., Rehfeldt, R. A., Zlomke, K. R., & Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *The Psychological Record, 56*, 83–103. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

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

Dube, W. V., Balsamo, L. M., Fowler, T. R., Dickson, C. A., Lombard, K. M., & Tomanari, G. Y. (2006). Observing behavior topography in delayed matching to multiple samples. *The Psychological Record, 56*, 233–244. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

Dube, W. V., Dickson, C. A., Balsamo, L. M., O'Donnell, K. L., Tomanari, G. Y., Farren, K. M. et al. (2010). Observering behavior and atypically restricted stimulus control. *Journal of the Experimental Analysis of Behavior, 94*, 297–313. doi: 10.1901/jeab.2010.94-297

Dube, W. V., & McIlvane, W. J. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis, 32*, 25–33. doi: 10.1901/

jaba.1999.32-25

Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures and derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8–12. Retrieved from http://www.eahb.org/NewSitePages/BulletinHomepage.htm

Eckerman, D. A., Lanson, R. N., & Cumming, W. W. (1968). Acquisition and maintenance of matching without a required observing response. *Journal of the Experimental Analysis of Behavior, 11*, 435–441. doi: 10.1901/jeab.1968.11-435

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis, 10*, 187–202. Retrieved from http://www.ejoba.org/

Fields, L., Arntzen, E., Nartey, R., & Eilifsen, C. (in press). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior*.

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., & Hobbie, S. A. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects 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

Foster, T. M., Temple, W., Mackenzie, C., Demello, L. R., & Poling, A. (1995). Delayed matching-to-sample performance of hens: Effects of sample duration and response requirements during the sample. *Journal of the Experimental Analysis of Behavior, 64*, 19–31. doi: 10.1901/jeab.1987.48-317

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, NY: Plenum Press.

Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record, 50*, 305–338. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

Lian, T., & Arntzen, E. (2010, May). *Matching-to-sample tasks and observing responses.* Poster presented at the annual meeting of the Association for Behavior Analysis, San Antonio, TX.

Maki, W. S., Gillund, G., Hauge, G., & Siders, W. A. (1977). Matching to sample after extinction of observing responses. *Journal of Experimental Psychology: Animal Behavior Processes, 3*, 285–296. doi: 10.1037//0097-7403.3.3.285

McIlvane, W. J., Dube, W. V., & Callahan, T. D. (1996). Attention: A behavior analytical perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 97–117). Baltimore, MD: Brookes.

McIlvane, W. J., Serna, R. W., Dube, W. V., & Stromer, R. (2000). Stimulus control topography coherence and stimulus equivalence: reconciling test outcomes with theory. In J. C. Leslie & D. E. Blackman (Eds.), Experimental and Applied Analysis of Human Behavior (pp. 85–110). Reno, NV: Context Press.

Palmer, D. C. (2010). Behavior under the microscope: Increasing the resolution of our experimental procedures. *The Behavior Analyst, 33*, 37–45. Retrieved from http://www.abainternational.org/TBA.asp

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record, 43*, 725–744. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html

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

Spencer, T. J., & Chase, P. N. (1996). Speed analysis of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643–659. doi: 10.1901/jeab.1996.65-643

Spetch, M. L., & Treit, D. (1986). Does effort play a role in the effect of response requirements on delayed matching to sample? *Journal of the Experimental Analysis of Behavior, 45*, 19–31. doi: 10.1901/jeab.1986.45-19

Wyckoff, L. B., Jr. (1952). The role of observing responses in discrimination learning: Part 1. *Psychological Review, 59*, 431–442. doi: 10.1037/h0053932