



Effects of Limited Hold on Equivalence Class Formation

Felix Högnason¹  · Erik Arntzen¹

Accepted: 22 June 2021 / Published online: 20 July 2021
© The Author(s) 2021

Abstract

In an attempt to limit the opportunity to engage in mediating behavior, two groups of adult participants received preliminary training in identity matching with limited hold levels (LH) for responding of 0.7 s for the sample and 1.2 s for the comparisons. The two groups were subsequently trained to form three 5-member classes, using the same LH levels, where the A, B, D, and E stimuli were abstract stimuli, and the C stimuli were meaningful pictures. In two tests for emergent relations, the LH for Group Short was unchanged, whereas 5 s were added to the LH for the comparisons for Group Long. None of the participants in Group Short responded in accordance with stimulus equivalence in either of the two tests. In Group Long, one participant responded in accordance with stimulus equivalence in the first test, and an additional eight participants formed equivalence classes in the second test.

Keywords stimulus equivalence · limited hold · meaningful stimuli · college students

Within behavior analysis, the term *stimulus equivalence* is employed when a group of stimuli, initially with no common defining characteristic, are related in such a way that they become interchangeable with each other (Green & Saunders, 1998; Sidman, 1994). Stimulus equivalence is defined by the properties of reflexivity, symmetry, and transitivity (Sidman & Tailby, 1982). When participants have been trained on six conditional discriminations and been tested for the emergence of three 3-member classes (e.g., $A \rightarrow B \rightarrow C$), reflexivity is demonstrated when each stimulus is related to itself; comparison A is selected when A is a sample; comparison B is selected when B is a sample, and so forth. Symmetry is assumed to emerge if comparison A is selected when B is a sample, and comparison B is selected when C is a sample. Transitivity is demonstrated if comparison C is selected when A is a sample. Finally, an equivalence test, or a combined symmetry and transitivity test, can be implemented by testing if comparison A is selected when C is a sample (Sidman & Tailby, 1982). Three training structures have been used to establish conditional discriminations: one-to-many (OTM), many-to-one

(MTO), and linear series (LS; Arntzen, 2012; Saunders et al., 1993). The LS training structure ($A \rightarrow B \rightarrow C$) is known to result in poor outcomes on subsequent tests for equivalence class formation and is thus ideal for investigating variables affecting the test outcome (Arntzen, 2012).

Additional measures, such as reaction time and stimulus sorting, have been suggested in research on emergent relations (Dymond & Rehfeldt, 2001). Time, as a variable, can be used either as a dependent variable in addition to accuracy measures (e.g., measuring reaction time), or as an independent variable (e.g., when responses are restricted under limited hold [LH] contingencies). Reaction time is usually defined as the time between the onset of a stimulus and a response to it. It has been suggested that speed (response/s), defined as the inverse of reaction time (s/response), provides a more valid representation of performance than reaction time (Whelan, 2008).

Although behavior associated with slower speed—covert naming or problem-solving behavior (e.g., Bentall et al., 1999; Donahoe & Palmer, 2004; Holth & Arntzen, 2000; Palmer, 1991)—could affect the formation of class consistent responses, there is some controversy within behavior analysis regarding the role of such covert behavior and its effect on the formation of the equivalence classes (e.g., Horne & Lowe, 1996; Sidman, 1994). It has been suggested, however, that one of the ways of reducing the likelihood of mediation, or problem-solving behavior, is to significantly limit the time the participants are given to respond to the stimuli (Holth & Arntzen, 2000; Spencer & Chase, 1996; Wulfert & Hayes,

✉ Felix Högnason
felixh@simnet.is

✉ Erik Arntzen
erik.arntzen@equivalence.net

¹ Department of Behavioral Science, Oslo Metropolitan University, PO Box 4, St. Olavs Plass, 0130 Oslo, Norway

1988). Only a few studies have used time as an independent variable in the field of stimulus equivalence research (e.g., Arntzen & Haugland, 2012; Arntzen & Liland, 2019; Holth & Arntzen, 2000; Imam, 2001, 2003; Tomanari et al., 2006), and only some of them have used a time limit for both the sample and the comparison stimuli.

Tomanari et al. (2006) speculated whether equivalence-class consistent performances emerged if participants had minimal time to engage in "subvocal or mediating behavior" (p. 349). Hence, the authors arranged an experiment in which five adult participants underwent conditional-discrimination training without time restrictions, followed by the introduction of LH contingencies. LH was titrated down to asymptotic levels of 0.4 to 0.5 s for the sample and 1.2 to 1.3 s for the comparisons. Three out of five participants responded in accordance with stimulus equivalence in a test for emergent relations. Tomanari et al. concluded that although equivalence class formation was not shown for all participants, the results for some of the participants indicated that the longer time interval typically used in stimulus equivalence procedures is not a necessary condition for equivalence class formation. The subvocal or mediating behavior could have played a role in the formation of equivalence classes because the conditional-discrimination training in the first phase did not comprise any time restriction. The high number of training trials also indicates that the experiment, being conducted over several weeks, could have affected the participants' performance in several ways, e.g., by increasing the probability of mediating behavior throughout the training and test period.

Based on the findings of Tomanari et al. (2006), Arntzen and Haugland (2012) investigated whether derived relations would emerge, given rapid response contingencies, with a shorter learning history than found in the Tomanari et al. study. Five adult participants attempted to form three 3-member classes with no time restrictions when the conditional-discrimination training was initiated. In the maintenance phase, the LH was titrated down to 1.2 s, but only for the comparison stimuli. On the test for emergent relations, one out of five participants responded in accordance with stimulus equivalence. The absence of limited hold contingencies in the conditional-discrimination training in the first phase, and the arrangement of implementing time restrictions only for the comparisons in the remainder of the experimental phases, could have provided the participants the opportunity to employ covert naming or other problem-solving strategies.

To reduce the possibility of some sort of mediating behavior facilitating correct responding when not using time constraints during the establishment of baseline relations, Arntzen and Liland (2019) studied the feasibility of using preliminary rapid response training to establish responding with very low LH values that could be used from the initial conditional-discrimination training. Five participants used a 0-s DMTS procedure and an OTM training structure with a concurrent

presentation in an attempt to form three 3-member classes. In Phase 1, using identity matching with colors, the LH levels were titrated down from the initial time of 2 s to an asymptotic level for the sample and the comparison stimuli. The LH values were determined on the grounds of the asymptotic level, plus 0.2 s, to ensure enduring performance. In Phase 2, the participants underwent conditional-discrimination training, with a limit of 720 training trials, each using their fixed individual LH levels obtained in the previous phase. The levels ranged from 0.4 to 0.7 s for the sample and 0.8 to 1.1 s for the comparisons. None of the participants were able to reach the training criterion for this phase. In the third phase, the LH for the comparison was titrated up to a mastery level for the conditional-discrimination training, which eventually ranged from 1.6 to 9.8 s. On reaching the training criterion of 90%, the participants were tested for emergent relations with the LH levels set to 2.5 s for the comparisons only. One participant responded in accordance with stimulus equivalence in the first test, and an additional three participants in the second test. In this study, the very narrow time window in the second phase prevented the participants from reaching the training criterion for that phase. As a result, the LH levels were titrated upwards, giving the participants the time to engage in subvocal mediating behavior, which could have affected their performance in the subsequent test-

As can be seen from the above-mentioned studies, the attempts to limit the participants' response time, e.g., to reduce their opportunities to engage in subvocal mediating behavior, have not been altogether sufficient. Some studies did not use time restriction in the initial phase of the conditional-discrimination training. When used, the LH levels were so strict that the participants could not achieve the training criterion and advance to a subsequent test. Moreover, some studies applied LH levels only for the comparison's stimuli and not for the sample. The purpose of the present study is, therefore, to prevent as far as possible the participants' opportunity to engage in mediating behavior beyond what may be expected within the LH settings and, at the same time, allow the participants to reach the criterion for the conditional-discrimination training. The question is whether the LH levels applied to achieve the training criterion will also be sufficient to form experimenter-defined classes

A track of research on the role of meaningful stimuli has employed an LS training structure training on 12 conditional discriminations and testing for emergent relations (e.g., Arntzen & Mensah, 2020; Arntzen et al., 2015a; Fields et al., 2012). These experiments have included two reference groups, all abstract stimuli and C stimuli as meaningful stimuli and A, B, D, and E stimuli as abstract shapes. The main findings were that the inclusion of meaningful stimuli had increased the outcomes substantially compared to classes with only abstract stimuli. In the present experiment, increasing the probability of emergent relations by using meaningful pictures

as C stimuli enables us to explore more precisely the effect of time restriction on equivalence class formation.

The main purpose of the present experiment is to investigate how rapid response contingencies affect conditional-discrimination training and equivalence class formation when employing an LS training structure with C stimuli as meaningful stimuli. First, will very short, limited-hold contingencies for responding, 0.7 s and 1.2 s for the sample and comparisons, respectively, in training and subsequent tests, reduce equivalence class formation even with meaningful stimuli? Second, do these LH levels in training prevent the participants from forming equivalence classes if they are given prolonged time to respond in the tests for emergent relations?

Apart from reaction time, stimulus sorting is another additional measure that has been suggested in research on emergent relations (Arntzen et al., 2017; Arntzen et al., 2015b; Dymond & Rehfeldt, 2001; Sigurðardóttir et al., 2012; Smeets et al., 2000). A postclass formation test is used in the present study to explore further the predicting outcome of the sorting task.

Gradual or delayed emergence of equivalence classes describes the increased responding in accordance with experimenter-defined classes with repeated testing (Sidman, 1994; Sidman et al., 1985). According to Sidman (1994), some stimuli may be members of several equivalence classes despite adequate training on conditional discrimination. However, the consistency in the sample-comparison relation established in the conditional-discrimination training will eventually lead to a delayed class consistent performance, even in the absence of programmed consequences. Therefore, two tests for emergent relations are conducted to access the delayed emergence of equivalence classes in the present experiment.

Method

Participants

Thirty-seven adults were recruited from an undergraduate course in psychology at the University of Iceland and via personal contacts. There were 26 females and 11 males, aged between 18 and 36 years ($M = 22$, $SD = 2.8$). None of them had participated in stimulus equivalence research or were familiar with the procedure. The participants were given an informed consent form to read, which contained general information about the experiment.¹ In the consent form, they were informed that the experiment could last for 2 days or more, that it would take 2 hr each day, that their anonymity would be assured, and that they were free to withdraw from the experiment at any time without penalty. All the

participants were given 1,500 ISK (approximately \$11.50) for each hour they attended the experiment. In addition, a 5% course credit was given to the undergraduates. When the participants had completed their participation, they were debriefed, thanked, and paid.

Apparatus, Setting, and Stimuli

Hardware

An HP ProBook, 15 in. portable computer running Windows 7 Enterprise system, and an LG Flatron T1710, 17 in. touch screen was used in the experiment.

Software

Custom-made MTS software was used to conduct conditional-discrimination training and test emergent relations, present the stimuli, and record the data.

Setting

The sessions were conducted in a lab at the university campus, measuring approximately 2×5 m, divided by a portable wall. The participants sat in a chair at a table, facing a blank wall, with the touch screen in front of them and the computer out of reach. The experimenter sat on the other side of the wall divider.

Stimuli

Fifteen stimuli, 12 abstract shapes, and 3 meaningful picture stimuli, approximately 3 cm x 3 cm, were used for the conditional-discrimination training and in the tests for emergent relations (see Figure 1). An additional three colored stimuli, green, red, and yellow, were used in the preliminary training and the retraining phases of rapid responding. These stimuli were approximately 3 cm x 6 cm on the screen.

The sample stimulus was displayed in the center of the monitor, and the three comparison stimuli were presented randomly in a circle, approximately 10 cm from the center of the screen.

Experimental Design

A between-group design was used in this study. Participants were quasi-randomly assigned to one of two groups. The conditions in the preliminary training and the conditional-discrimination training were the same for both groups, as was the postclass formation sorting task. However, the limited hold levels in the tests for emergent relations were different for each group. For Group Short, the LH levels used in the test were the same as the LH levels used in the conditional-

¹ All information given to the participants was written in Icelandic.

Fig. 1 Stimuli Used in the Conditional-Discrimination Training and Subsequent Tests



discrimination training, i.e., 0.7 s for the sample and 1.2 s for the comparisons. For Group Long, the LH levels in the test were 0.7 s for the sample and 6.2 s for the comparisons, which added 5,000 ms to the original LH for the comparisons.

Dependent Measures

Responses to sample and comparison stimuli were recorded, as well as the reaction time for the comparison response within the LH. The reaction time was recorded from the presentation of the comparison stimuli until a stimulus was touched. The response speed is the inverse of the reaction time to respond to the comparison stimuli, calculated as the mean of the median speed, for the last five training trials for the baseline relations (BSL-TR), and the first five and the last five test trials for the baseline probes (BSP), symmetry relations (SYM), 1-node (1-N), 2-node (2-N), and 3-node (3-N) relations. Responding in accordance with equivalence was defined as 90% accuracy or more for all relations with a maximum of one error for each trial type.

Procedure

An overview of the procedure is given in Table 1. The length of the experiment for each participant depended on how rapidly and correctly each of them responded. Daily sessions were also limited to 2 hr each day to reduce fatigue. Twelve of the participants conducted the experiment in 1 day, and 23 did so in 2 subsequent days (except participant 15377, who had a 46-hr break between the two sessions). The experimental sessions extended over 3 subsequent days for the remaining two participants. The test for emergent relations was administered immediately after the training session.

Information to the Participants

During recruitment, the participants were told that the research was in the field of experimental behavioral analysis. Before the experimental sessions commenced, the participants were required to read through and sign a consent form that informed them about their anonymity and their right to terminate their

Table 1 An Overview of the Experimental Phases

Experimental Phases	Trial Types	Trials per Block
1. Identity Matching with Limited Hold (LH)	Three Colored Stimuli	30
2. Acquisition of Baseline Relations with LH	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	15
Serialized Trials (presented randomly)		30, 45, 60
Mixed Trials (presented randomly)		
2.1. Maintenance of baseline relations		60
Mixed Trials (presented randomly)		
PC: 75%–50%–25%–0%		
3. Test Block 1		180
Baseline Relations	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3, C1D1, C2D2, C3D3, D1E1, D2E2, D3E3	
Symmetry Relations	B1A1, B2A2, B3A3, D1C1, D2C2, D3C3, E1D1, E2D2, E3D3	
1-Node Relations	A1C1, A2C2, A3C3, C1A1, C2A2, C3A3, B1D1, B2D2, B3D3, D1B1, D2B2, D3B3, C1E1, C2E2, C3E3, E1C1, E2C2, E3C3	
2-Node Relations	A1D1, A2D2, A3D3, D1A1, D2A2, D3A3, B1E1, B2E2, B3E3, E1B1, E2B2, E3B3	
3-Node Relations	A1E1, A2E2, A3E3, E1A1, E2A2, E3A3	
4. Test Block 2		
5. Post Card-Sorting Task	Same as the Initial Test	180

Note. In Phases 1 and 2, the LH levels for both groups were 0.7 s for the sample and 1.2 s for the comparisons. In Phases 3 and 4, the LH levels were 0.7 s for the sample and 1.2 s for the comparisons, and 0.7 s for the sample and 6.2 s for the comparisons, for Group Short and Group Long, respectively. In Phase 2, the programmed consequences (PC) were initially 100%, and then reduced stepwise to 0% (see Phase 2 in the Method section). Phases 3 and 4 were administered without programmed consequences.

participation at any time with no questions asked. They were told that the experiment could last 2 days or more, depending on how rapidly and correctly they responded. Finally, the participants were given a brief demonstration of how the touch screen responded

When the participants sat down in front of the touch screen, the following instructions were displayed on the screen:

A stimulus will appear in the middle of the screen. Click on it by pressing the touch screen. Three other stimuli will then appear on the screen. Choose one of these by pressing it in the same way. 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 or press it too late, the word *Wrong* will appear on the screen.

During some stages of the experiment, the computer will not tell you if your choices are right or wrong. However, based on what you have learned, you can complete all the tasks correctly. Please do your best to get everything right. Good Luck!

Each trial started with a presentation of a sample stimulus in the middle of the screen. When the participant touched the

sample, three comparison stimuli appeared in a circular layout, approximately 10 cm from the middle of the screen. If the participant chose the experimenter-defined, class-consistent stimulus, a word such as "Good," "Correct," and "Excellent" was displayed in the middle of the screen for 1 s. Choosing an incorrect comparison stimulus or responding outside the time limit was followed by the word "Wrong" displayed on the screen. An ITI was set to 1.4 s. During that time, the white screen did not respond to touch.

Phase 1. Preliminary Training

First, the participants underwent rapid response training using identity matching with colors (green, red, and yellow). The stimuli were presented in a matching-to-sample format with a 0-s delay. The LH level for responding was 0.7 s for the sample and 1.2 s for the comparison. Each training block consisted of 30 trials, with a mastery criterion of 90% for each block. At the beginning of each new day, a rapid response retraining phase was implemented. This phase was similar to the preliminary training phase, except that the criterion was 80% for each block.

Phase 2. Training on Conditional Relations

When the criterion of 90% in the preliminary phase was reached, the participants underwent conditional-discrimination training, using an LS training structure and a simultaneous protocol, with 15 stimuli comprising 12 abstract and 3 meaningful picture stimuli. The trial types were introduced serialized, and the LH level for responding was 0.7 s for the sample and 1.2 s for the comparison. The participants started training on AB trials in a block of 15 trials and when the criterion of 90% was met, the participants trained on a block of BC trials in the same fashion. When the criterion for the BC trial type was met, training was conducted on a block of 30 randomly mixed AB and BC trials until a criterion of 90% was accomplished. Next, the participants trained on a block of 15 CD trials and a subsequent block of 45 mixed AB, BC, and CD trials. When the criterion for the mixed block of AB, BC, and CD trials was met, a block of 15 DE trials was introduced, and when the criterion for that block was met, the participants trained on a block of 60 mixed AB, BC, CD, and DE trials until the criterion was met. The programmed consequences were initially 100%. On reaching the criterion of 90% accuracy or more for the last mixed block and a maximum of one error for each trial type, the programmed consequences were faded to 75%, 50%, 25%, and 0%. If the participants failed to reach the mastery criterion, the block was repeated until the criterion was met. For the AB relations, the participants were trained to choose the B1, B2, or B3 comparison stimulus when sample stimulus A1, A2, or A3 was presented, respectively. For the BC relations, the choice of the C1, C2, or

C3 comparison was reinforced when the sample stimulus B1, B2, or B3 was presented, etc. The trial types for the baseline relations are presented in Table 1.

The participants were allowed to take a 1-min break every 36th trial to reduce fatigue. Irrespective of whether they did so, the experimenter instructed the participants to take a longer break every 30 min or so, or around the 360th trial. The maximum work each day was 2 hr or approximately 1,440 trials. The daily session then included three longer 5-min breaks and 36 shorter 1-min breaks.

Phase 3. Test Block 1

After reaching the criterion in the last part of Phase 2, the participants were given a block of randomly mixed test trials with no programmed consequences. The test block comprised a total of 180 trials, i.e., 36 baseline probes, 36 symmetry trials, 54 one-node trials, 36 two-node trials, and 18 three-node trials (see Table 1). Each relation was tested three times in random order. The LH levels for Group Short were 0.7 s for the sample and 1.2 s for the comparisons, and the LH levels for Group Long were 0.7 s for the sample and 6.2 s for the comparisons. The test criterion was 90% or more correct responses for all relations, with a maximum of one error for each trial type.

Phase 4. Test Block 2

The test was repeated immediately after Phase 3, with the object of assessing the delayed emergence of equivalence classes.

Phase 5. Post-Card Formation Sorting Task

In this phase, the experimenter randomly placed a set of laminated cards containing the 15 stimuli used in the conditional-discrimination training and subsequent tests on a table in front of the participants. Then the experimenter instructed the participants to arrange the cards in the way they felt most appropriate. If the participants asked questions regarding the task, the experimenter only repeated the instruction.

Results

Acquisition and Maintenance of Baseline Relations

The range of the number of trials to acquire and maintain baseline relations was 945–3,885 trials for Group Short, ($M = 2,029$, and $Mdn = 1,958$) and 855–4,785 trials for Group Long ($M = 1,808$, and $Mdn = 1,770$). The difference between the groups regarding the number of training trials was not significant, $F(1, 35) = 0.675$, $p = .417$.

Immediate and Delayed Class Formation

Overall, for both groups, 9 out of 37 participants responded in accordance with stimulus equivalence in the two MTS test blocks. Eleven participants responded in accordance with the criterion for at least one of the four emergent relations needed to form an equivalence class. None of the participants in Group Short responded in accordance with stimulus equivalence in either of the two test blocks. One participant in Group Long formed equivalence classes in Test Block 1, with the maintenance of the classes in Test Block 2. In addition, eight participants showed delayed emergence of equivalence classes in Test Block 2 (see Fig. 2).

Test Block 1 Only one participant in Group Short had intact baseline relations, which means a breakdown of the baseline relations in the test for 94% of the participants in that group (average responding to baseline relations was reduced from 93% in the last training block to 71% in Test Block 1). None of the participants in Group Short met the criterion for any of the emergent relations in this test block. In Group Long, apart from the one participant who responded in accordance with stimulus equivalence, nine participants had intact baseline relations, whereas eight participants did not, and seven participants met the criterion for at least one of the emergent relations.

Test Block 2 Two participants in Group Short had intact baseline relations, and three participants responded in accordance with symmetry relations. In Group Long, apart from the nine

who responded in accordance with stimulus equivalence, eight participants had intact baseline relations, and seven participants responded in accordance with at least one of the emergent relations. For Test Block 2, Fisher's Exact Test showed significant differences between the two groups for all relations, $p = .000$.

An overview of the results for both groups in Test Block 1 and Test Block 2 is shown in Fig. 3.

Response Speed

Figure 4 shows the speed of correct responses within the time limit for responding for both groups, in Test Block 1 and Test Block 2. For Group Short, the speed is highest for the BSL-TR relation (1.30 per s), and lowest for the first five test trials for the 3-N relations in the first test (0.96 per s; see Fig. 4, the upper panel on the left). The typical pattern of higher speed for the baseline probes and the symmetry trials, compared to the 1-, 2-, and 3-node trials, is not apparent.

For Group Long, the mean of the median speed for the last five training trials was 1.31 per second. The typical pattern of higher speed for the baseline probes and the symmetry relations, compared to the 1-, 2-, and 3-node relations in Test Block 1, was significant, $p < .05$ (see Fig. 4, the lower panel on the left). In addition, in Test Block 2, the speed for the BSP and the SYM relations was significantly higher compared to the 2-N and the 3-N relations, $p < .01$ (see Figure 4, the lower panel on the right).

The LH values for the two groups determined a lower speed limit of 0.83 for Group Short, and 0.16 for Group Long. The range of speed for both groups in both tests are

Number of Participants who Formed Equivalence Classes in Test Block 1 and Test Block 2

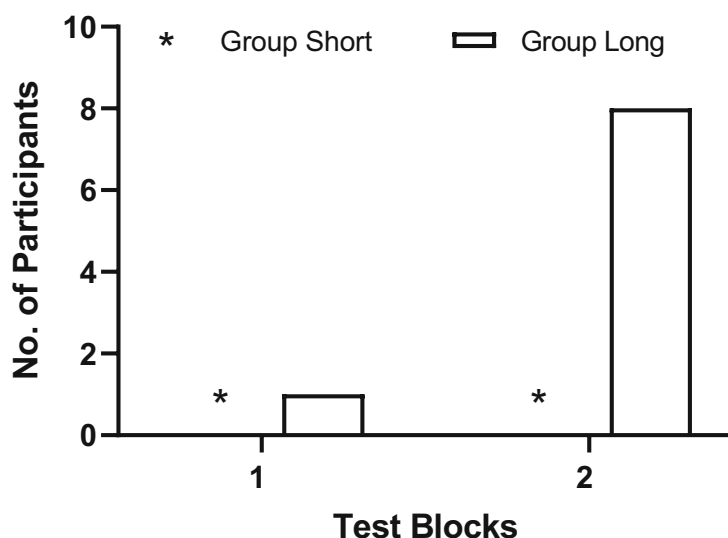


Fig. 2 Number of Participants Who Formed Equivalence Classes in Test Block 1 and Test Block 2. *Note.* None of the participants in Group Short (*) formed equivalence classes in either of the tests

Performance for Both Groups in Test Block 1 and Test Block 2

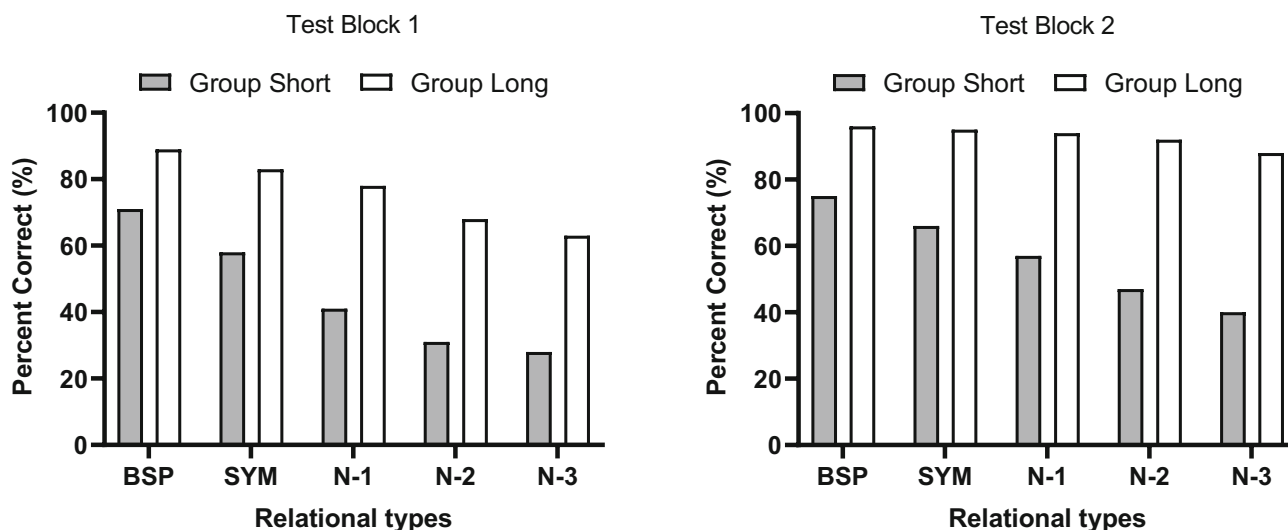


Fig. 3 Performance for Both Groups in Test Block 1 and Test Block 2. *Note.* The percentages of correct responses for each relational type in Test Block 1 (left) and Test Block 2 (right) for Group Short (gray bars) and Group Long (white bars). Baseline probes (BSP), symmetry probes

(SYM), and the 1-, 2-, and 3-node probes (1-N, 2-N, and 3-N). There were significant differences between the two groups in Test Block 2 for all relations, $p = .000$

shown in Figure 5. In the first test, the speed for Group Short ranged from 1.0 to 1.15, ($M = 1.07$, $SE = 0.04$), and the speed for Group Long ranged from 0.32 to 0.89 ($M = 0.62$, $SD = 0.16$). The difference between the groups was significant, $M = 0.45$, $SE = 0.04$, $t(20.490) = 11.949$, $p = .000$. In the second test, the speed for Group Short ranged from 1.01 to 1.21 ($M = 1.09$, $SD = 0.06$), and the speed for Group Long, ranged from 0.35 to 0.92 ($M = 0.67$, $SD = 0.17$). The difference between the groups was also significant here, $M = 0.42$, $SE = 0.04$, $t(21.871) = 10.189$, $p = .000$.

Post-Class Formation Sorting Task

In the postclass formation sorting task, all participants in Group Long and all but two participants in Group Short arranged the laminated cards in accordance with the experimenter-defined classes.

Discussion

As in previous studies using time restrictions in conditional-discrimination training, the results in the present study show a considerably higher number of training trials compared to studies without time restrictions. However, despite an extended amount of training, the results for Group Short clearly show that the restricted time participants have to respond to stimuli in the MST test eliminates the anticipated enhancing effect

of the meaningful pictures on participant performance. These results support the view that, when not given time, e.g., to employ mediating or problem-solving strategies, participants' ability to form equivalence classes is severely reduced or even absent. The results for Group Long and from the postclass formation sorting task further indicate that conditional-discrimination training with time restrictions is sufficient to promote responses in accordance with stimulus equivalence on subsequent tests as long as the time restrictions in the tests are less stringent than in training.

This result is inconsistent with the findings of Tomanari et al. (2006) and Arntzen and Haugland (2012), where the studies showed a more positive outcome for some of the participants trained under limited hold conditions. However, the result is similar to that of Arntzen and Liland (2019), where LH contingencies were implemented from the very beginning of conditional-discrimination training, unlike Tomanari et al. and Arntzen and Haugland, where time restrictions were only implemented after the participants had reached the mastery criterion during the conditional-discrimination training.

Number of Training Trials

The number of trials to acquire and maintain baseline relations is noticeably higher compared to studies not using time restrictions (e.g., Arntzen, Nartey, & Fields, 2015; Fields et al., 2012). In the Fields et al. (2012) study, for example, the median number of trials to acquire baseline relations for the

Response Speed for both Groups in Test Block 1 and Test Block 2

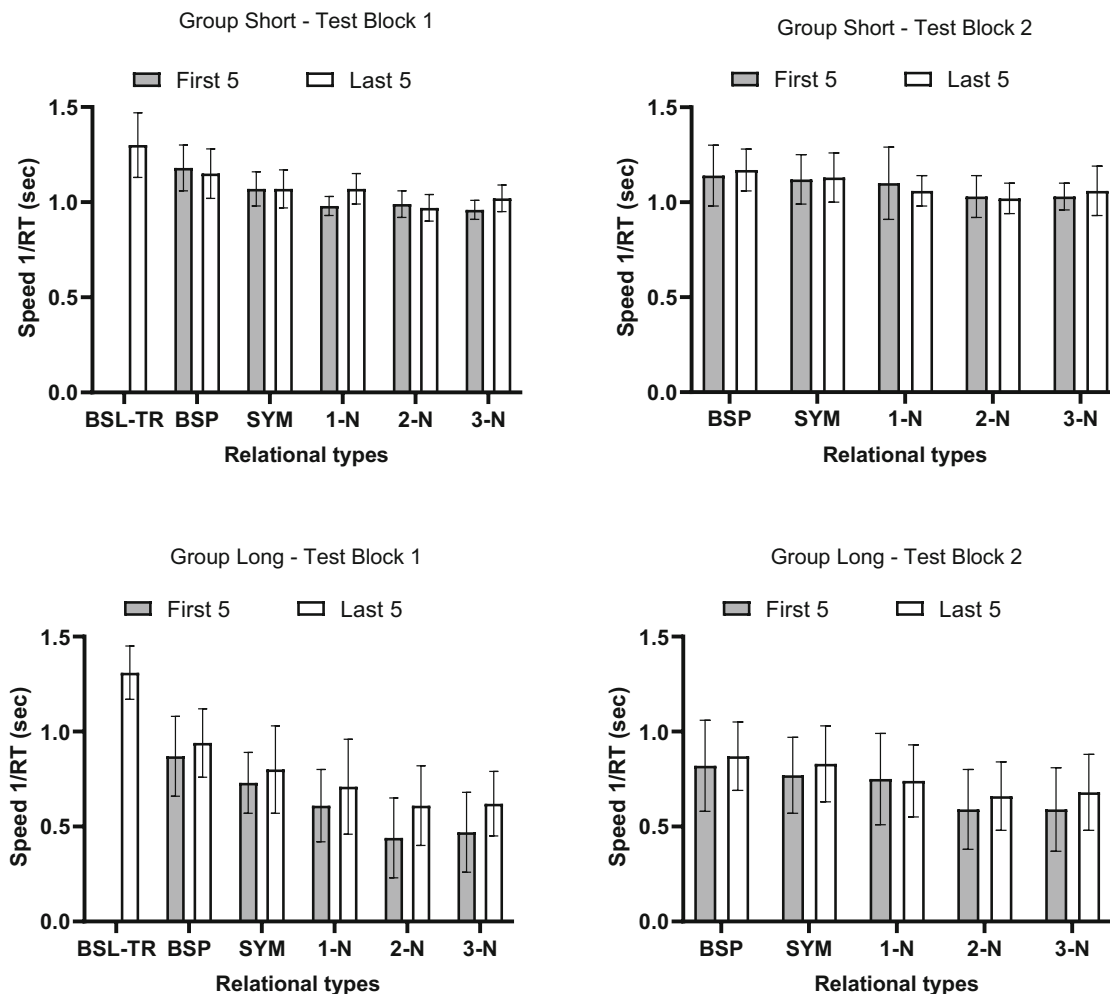


Fig. 4 Response Speed for both Groups in Test Block 1 and Test Block 2. *Note.* Response speed (inverse of the reaction time) is calculated as the mean median speed for the last five training trials (BSL-TR) and the first five and last five trials during testing for baseline probes (BSP), symmetry

(SYM), 1-node (1-N), 2-node (2-N), and 3-node (3-N) relations. Error bars show the standard deviation of the mean. The upper panels show the speed for Group Short, and the lower panels show the speed for Group Long. Test Block 1 is on the left, and Test Block 2 is on the right.

groups was approximately 350 trials, whereas the median number for the groups in the present study was 1,864 trials. In the conditional-discrimination training phase, before the thinning of the programmed consequences, all incorrect responses and responses outside the time limits resulted in the programmed feedback, "wrong." Thus, the participants were unable to distinguish between the responses that were outside the time limits and the responses that were not in accordance with the experimenter-defined relations. This limitation could have made conditional discrimination more difficult for the participants. However, it is clear that the time restriction in training impeded the acquisition of baseline relations (e.g., Arntzen & Liland, 2019). Further research will have to clarify in what way time restrictions during training will affect the acquisition of emergent relations in tests without time limits.

Delayed Emergence

The delayed, or gradual emergence of equivalence classes, refers to an increase in the number of correct responding with repeated testing for equivalence relations (Arntzen & Mensah, 2020; Sidman, 1994). Arntzen and Mensah (2020) defined delayed emergence as responding correctly below 90% in Test Block 1 and at least 90% in a Test Block 2. In three experiments involving two reference groups, an abstract group using all abstract stimuli and a picture group using meaningful pictures as part of the stimuli set, the participants were trained on 12 baseline conditional-discriminations and tested for the formation of three 5-member equivalence classes in two subsequent tests. Whereas 40% of the participants in the three experiments responded in accordance with stimulus

Speed Range

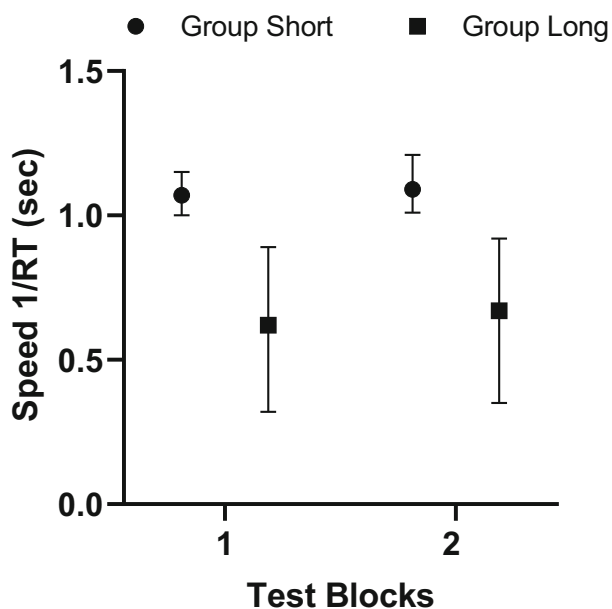


Fig. 5 Speed Range. *Note.* The figure shows the range of speed for both groups in Test Block 1 and Test Block 2. The difference in the median speed between the groups was significant in both tests, $p = .000$

equivalence in both test blocks, 18.8% of the participants who did not show immediate equivalence class formation in Test Block 1 responded in accordance with stimulus equivalence in Test Block 2. Other studies in the same line of research, as mentioned above (e.g., Arntzen, Narthey, & Fields, 2015; Arntzen & Narthey, 2018), have reported similar findings. In the present study, two test blocks of 180 trials each were conducted in a similar manner as in the Arntzen and Mensah (2020) study; Test Block 1 measured the immediate performance, and Test Block 2 measured the delayed emergence. This high number of test trials allowed us to further investigate the development of delayed emergent relations (see Figure 6). The results for the PIC-groups in Experiment 1 and 2 in the Arntzen and Mensah (2020) study, where the C stimuli were meaningful pictures, and the A, B, D, and E stimuli were abstract shapes show that 66.7% and 60% of the participants scored 90% or higher in the first test, and 80% and 93.3% of the participants scored 90% or above in the second test, in Experiment 1 and 2, respectively. Comparing these results with the results from the present study, it is clear that the time restrictions for the groups in the present study affected the participants' responding unfavorably. The different time restrictions in the tests for the different groups also clearly affected the participants. None of the participants in Group Short reached an overall score of 90% in any of the four test halves (in 360 trials), although the participants in this group made persistent progress up to a certain point and reached a peak in the third test half (see Fig. 6). In Group Long, one participant had an overall score of 90% or more in the first test half, 11 in the second test

half, 13 in the third test half (first half of the second test), and 17 in the fourth test half. Although the time restrictions under training and testing prevented the participants in the present study from responding in accordance with stimulus equivalence, easing the time restriction in the test for some of the participants (see Fig. 5) increased the probability of responding in accordance with delayed emergent relations.

Speed Pattern

The particular speed pattern, usually seen in tests for stimulus equivalence in studies not using time restrictions, shows higher speed for the baseline probes compared to the emergent relations and higher speed for the symmetry relations compared to the transitivity and equivalence relations (e.g., Arntzen & Lian, 2010; Bentall et al., 1993; Dymond & Rehfeldt, 2000; Imam, 2001; Spencer & Chase, 1996). Furthermore, higher speed is observed for trial types presented later in the test blocks compared to the early test trials (e.g., Arntzen et al., 2007; Donahoe & Palmer, 2004). It is apparent that time restrictions in the present study interfere with the configuration of such a pattern. As seen in Figure 4, upper panel, the response speed for participants in Group Short only partly complies with the pattern mentioned above, and the difference between the trial types is negligible. However, this pattern is clear for Group Long (see Figure 4, lower panel), where the time to respond to the comparisons is increased.

Sorting Performance

The results from the present study support the findings from previous studies on the correspondence of performance on MTS and sorting tests (e.g., Arntzen et al., 2017; Arntzen, Norbom, & Fields, 2015; Sigurðardóttir et al., 2012; Smeets et al., 2000). All but two participants sorted their cards in accordance with experimenter-defined classes, and only nine participants responded in accordance with stimulus equivalence in one of the two MTS tests. The two participants who failed to sort their cards correctly were part of Group Short, where the LH restrictions in training and tests were the same. It is important to emphasize that the participants sorted their cards without any time limits. Further studies could demonstrate that time limitations on the card-sorting task could reflect the performance on the MTS tests to some degree. Despite the rigid time frame in the conditional-discrimination training, the moderate results for Group Short in the card-sorting task indicate that the training provided the participants with the prerequisite to sort the stimuli in accordance with experimenter-defined classes, and the main impediment to responding in accordance with stimulus equivalence in the MST tests for many of the participants was the limited time they had at their disposal in the tests.

The Percent of Correct Responses in Each Test Half for Both Groups in Test Block 1 and Test Block 2

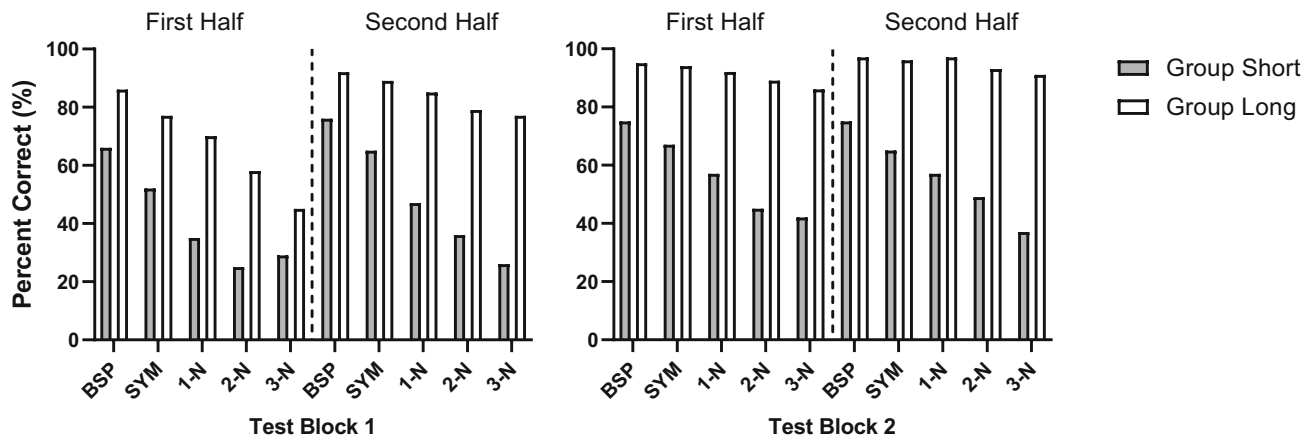


Fig. 6 The Percent of Correct Responses in Each Test Half for Both Groups in Test Block 1 and Test Block 2

Limitations

Several studies from our lab (e.g., Arntzen & Mensah, 2020; Arntzen, Nartey, & Fields, 2015; Arntzen & Nartey, 2018; Fields et al., 2012; Nartey et al., 2014), using similar training and testing procedures under normal time conditions, have reported on the enhancing effect of using meaningful pictures, in a set of otherwise abstract stimuli, on equivalence class formation. However, the absence of a control group in the present study, not using time restrictions, should be regarded as a limitation.

Another limitation concerns the order of the MTS and the sorting task. Future studies should counterbalance the order to rule out history effects as a possible threat to the study's internal validity.

Conclusion

The present study clearly demonstrates that restricted time limits in conditional-discrimination training and subsequent MTS tests hinder the participants from responding in accordance with stimulus equivalence, whereas moderate time restrictions enabled the formation of such classes for some participants. There are still some issues that need to be clarified. To what extent are the results of the present study joint effects of the LH parameters for the samples and the comparisons? At which point does an LH start to affect the participant's ability to respond in accordance with stimulus equivalence? Are more moderate time limits in the conditional-discrimination training, with fewer training trials, more effective regarding positive results on the MTS tests than a more rigid LH with a higher number of training trials? Further research should aim to verify these results more thoroughly and identify the role of some of the variables that allegedly play a part in these results,

e.g., the role of the LH in the conditional-discrimination training and the LH parameters related to the sample and the comparisons in the tests.

Funding Open access funding provided by OsloMet - Oslo Metropolitan University. This research was funded by Oslo Metropolitan University.

Declarations

Conflict of Interest The authors declare that there are no conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis*, 13(1), 123–135. <https://doi.org/10.1080/15021149.2012.11434412>.

- Amtzen, 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(2), 177–191. <https://doi.org/10.1080/15021149.2007.11434281>.
- Amtzen, E., Granmo, S., & Fields, L. (2017). The relation between sorting tests and matching-to-sample tests in the formation of equivalence classes. *The Psychological Record*, 67, 81–96. <https://doi.org/10.1007/s40732-016-0209-9>.
- Amtzen, E., & Haugland, S. (2012). Titration of limited hold to comparison in conditional discrimination training and stimulus equivalence testing. *The Psychological Record*, 62(2), 243–262. <https://doi.org/10.1007/BF03395800>.
- Amtzen, E., & Lian, T. (2010). Trained and derived relations with pictures versus abstract stimuli as nodes. *The Psychological Record*, 60(4), 659–678. <https://doi.org/10.1007/BF03395738>.
- Amtzen, E., & Liland, K. H. (2019). Effect of rapid responding on establishment of conditional discriminations and formation of equivalence classes. *The Psychological Record* 69(3), 367–378. <https://doi.org/10.1007/s40732-019-00345-6>.
- Amtzen, E., & Mensah, J. (2020). On the effectiveness of including meaningful pictures in the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 113(2), 305–321. <https://doi.org/10.1002/jeab.579>.
- Amtzen, E., Nartey, R., & K., & Fields, L. (2015a). Enhanced equivalence class formation by the delay and relational functions of meaningful stimuli. *Journal of the Experimental Analysis of Behavior*, 103(3), 524–541. <https://doi.org/10.1002/jeab.15>.
- Amtzen, E., Norbom, A., & Fields, L. (2015b). Sorting: An Alternative Measure of Class Formation? *The Psychological Record*, 65(4), 615–625. <https://doi.org/10.1007/s40732-015-0132-5>.
- Amtzen, E., & Nartey, R. K. (2018). Equivalence class formation as a function of preliminary training with pictorial stimuli. *Journal of the Experimental Analysis of Behavior*, 110(2), 275–291. <https://doi.org/10.1002/jeab.466>.
- Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *Quarterly Journal of Experimental Psychology Section B*, 46(2), 187–214. <https://doi.org/10.1080/14640749308401085>.
- Bentall, R. P., Jones, R. M., & Dickins, D. W. (1999). Errors and response latencies as a function of nodal distance in 5-member equivalence classes. *The Psychological Record*, 49(1), 93–115. <https://doi.org/10.1007/BF03395309>.
- Donahoe, J. W., & Palmer, D. C. (2004). *Learning and complex behavior*. Ledgetop Publishing.
- Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, 23(2), 239–254. <https://doi.org/10.1007/BF03392013>.
- Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin*, 19, 8–12 <http://www.eahb.org/1/wp-content/uploads/2014/03/Dymond-Rehfeldt.pdf>.
- Fields, L., Amtzen, E., Nartey, R. K., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior*, 97(2), 163–181. <https://doi.org/10.1901/Jeab.2012.97-163>.
- Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Peron (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). Plenum Press.
- Holth, P., & Amtzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record*, 50(2), 305–337. <https://doi.org/10.1007/BF03395358>.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65(1), 181–241. <https://doi.org/10.1901/jeab.1996.65-18>.
- Imam, A. A. (2001). Speed contingencies, number of stimulus presentations, and the nodality effect in equivalence class formation. *Journal of the Experimental Analysis of Behavior*, 76(3), 265–288. <https://doi.org/10.1901/jeab.2001.76-265>.
- Imam, A. A. (2003). Assessing transfer of response speed and nodality via conditional discriminations. *Experimental Analysis of Human Behavior Bulletin*, 21, 1–7 <https://www.eahb.org/bulletin2003>.
- Nartey, R. K., Amtzen, E., & Fields, L. (2014). Two discriminative functions of meaningful stimuli that enhance equivalence class formation. *The Psychological Record*, 64(4), 777–789. <https://doi.org/10.1007/s40732-014-0072-5>.
- Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior*. (pp. 261–279). Context Press.
- Saunders, K., Saunders, R., Williams, D., & Spradlin, J. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record*, 43(4), 725–744. <https://doi.org/10.1007/BF03395909>.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Authors Cooperative.
- Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, 43(1), 21–42. <https://doi.org/10.1901/jeab.1985.43-21>.
- 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(1), 5–22. <https://doi.org/10.1901/jeab.1982.37->.
- Sigurðardóttir, Z. G., Mackay, H. A., & Green, G. (2012). Stimulus equivalence, generalization, and contextual stimulus control in verbal classes. *Analysis of Verbal Behavior*, 28(1), 2–29. <https://doi.org/10.1007/BF033>.
- Smeets, P. M., Dymond, S., & Barnes-Holmes, D. (2000). Instructions, stimulus equivalence, and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record*, 50(2), 339–354. <https://doi.org/10.1007/BF03395359>.
- Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, 65(3), 643–659. <https://doi.org/10.1901/jeab.1996.65-64>.
- Tomanari, G. Y., Sidman, M., Rubio, A. R., & Dube, W. V. (2006). Equivalence classes with requirements for short response latencies. *Journal of the Experimental Analysis of Behavior*, 85(3), 349–369. <https://doi.org/10.1901/jeab.2006.107-04>.
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475–482. <https://doi.org/10.1007/BF03395630>.
- Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, 50(2), 125–144. <https://doi.org/10.1901/jeab.1988.50-125>.