Talk-Aloud Protocols during Conditional Discrimination Training and Equivalence Class Formation

Aleksander Vie¹ · Erik Arntzen¹

Published online: 2 May 2017

© Association for Behavior Analysis International 2017

Abstract The primary focus of the present experiment was to study participants' self-talk during conditional discrimination training and test for the emergence of conditional relations consistent with equivalence. Eighteen adult participants were exposed to either a Many-to-One (MTO) or a One-to-Many (OTM) training structure arranged as a 6-s delayed matching-to-sample (DMTS) procedure. During training and testing, participants were asked to talk aloud. Each trial was divided into three time-windows for recording vocal responses. The first time-window was when the sample was presented, the second time-window was the delay following the offset of the sample stimulus until the onset of the comparisons, and the final time-window was when the comparison array was presented. Overall, participants in the OTM condition talked more than participants in the MTO condition. The participants exposed to MTO increased talking during the third time-window from last part of training to test more than the participants in the OTM condition. For 16 of the 18 participants, talking aloud was less than 55% in the second time-window (during the delay). Conversely, talking in the first and third time-windows were higher than 90% in both groups. Thus, it seems that talking is affected by the shift from training to testing to a greater degree when exposed to MTO than OTM. Furthermore, the data indicate that the participants did not need to talk aloud during the delay to respond correctly on DMTS trials.

Keywords Adults · Conditional discrimination training · Delayed matching-to-sample · Stimulus equivalence · Talk aloud protocol

✉ Aleksander Vie
alek@stimuluscontrol.com

✉ Erik Arntzen
erik.arntzen@equivalence.net

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

Responding in accordance with stimulus equivalence is defined by three properties: reflexivity, symmetry, and transitivity (Sidman & Tailby, 1982). The most commonly used procedure to train the prerequisites for equivalence class formation is a simultaneous matching-to-sample procedure (SMTS) (e.g., Arntzen, 2004; Arntzen, Halstadro, Bjerke, & Halstadro, 2010; Imam, 2006; Tomanari, Sidman, Rubio, & Dube, 2006). Another procedure is delayed matching-to-sample (DMTS), in which the sample stimulus disappears before the comparison stimuli are presented (e.g., Arntzen, 2006; Arntzen, Galaen, & Halvorsen, 2007; Eilifsen & Arntzen, 2011; Lian & Arntzen, 2013; Saunders, Chaney, & Marquis, 2005; Vaidya & Smith, 2006). As an example of the DMTS procedure, Arntzen, (2006) conducted several experiments to explore the effects of increasing and decreasing delays on stimulus equivalence. In experiments 1, 2, and 3, results indicated higher yields of equivalence responding when participants were exposed to increasing delays. The authors found that starting with longer delays increased the probability of equivalence responding. Arntzen, (2006) argued that the high yield of equivalence responding could be a result of participants self-echoically rehearsing the stimuli during the delay. Therefore, in experiment 4, Arntzen, (2006) introduced distracting tasks during the delay in the test for emergent relations. Results showed that none of the participants responded in accordance with stimulus equivalence. Arntzen & Vie, (2013) replicated experiment 4 of Arntzen, (2006) with an ABA- and BAB-design and found high yields of emergent relations in the conditions without distracters (A) and low yields in the conditions with distracters (B). Both Arntzen, (2006) and Arntzen & Vie, (2013) suggested that some type of mediating behavior, like a rehearsal of names, could be critical for facilitating equivalence class formation and that the distracters presented in the B conditions could have hindered such rehearsal.

Several studies have suggested that verbal behavior may facilitate responding in accordance with stimulus equivalence (e.g., Mandell & Sheen, 1994; Randell & Remington, 2006). Both common naming and intraverbal or bidirectional naming has been proposed as a possible mechanism responsible for equivalence class formation (Horne & Lowe, 1996; Miguel, 2016). Hence, a common name, or a common response, to each stimulus within a class of stimuli may lead to a positive equivalence outcome (e.g., Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Horne, Lowe, & Harris, 2007; Horne, Lowe, & Randle, 2004; Lowe, Horne, Harris, & Randle, 2002). In intraverbal naming, however, each stimulus in a class is given different names, and a verbal rule is created that links behavior and stimulus together which could form equivalence classes (Horne & Lowe, 1996). Recent studies have focused on teaching participants to tact each of the stimuli with different names by training intraverbals between these names. For example, Carp & Petursdottir, (2015) arranged for conditional discrimination training in six typically developing children (5–7 years old). The children were trained to form three 3-member classes after they had received category-name training and tact training. The children then received equivalence and intraverbal tests. The main findings were that for three of the children, there was a co-occurrence of intraverbal behaviors and positive equivalence outcomes.

Petursdottir, Carp, Peterson, & Lepper, (2015) arranged for tact training and intraverbal training of baseline relations in 10 typically developing children. Five of the children passed the visual-visual MTS test. However, not all the children showed tact behavior and intraverbal behavior when they responded correctly on the MTS task.

Santos, Ma, & Miguel, (2015) replicated the study by Petursdottir et al., (2015) with adults. The children were trained to tact the experimentally defined classes, and then to relate their names intraverbally (i.e., “A goes with B”). Finally, the children were exposed to symmetry trials as bidirectional intraverbals (i.e., “B goes with A”). Their results showed that participants responded correctly to the matching task, and also produced the bidirectional intraverbals correctly. Lastly, Ma, Miguel, & Jennings, (2016) found that undergraduate students formed equivalence classes after training to tact the individual stimuli used in the experiment and intraverbally to relate the A stimuli to their experimentally defined corresponding B stimuli and the B stimuli to their experimentally defined corresponding C stimuli.

Although in the aforementioned studies, except Carp & Petursdottir, (2015), the effects of verbal behavior training have been measured on MTS performance, most stimulus equivalence research is conducted without explicitly arranging such training. In the studies that have exclusively employed MTS, verbal reports have been assessed by post-session questionnaires (e.g., Arntzen, 2006; Lane & Critchfield, 1996; Sidman, Willson-Morris, & Kirk, 1986). Arntzen, (2006) used post-session questionnaires to probe whether participants used the same name to each stimulus within each class of stimuli or different names to each stimulus within each class. Sidman et al., (1986) also used post-session questionnaires to explore the role of naming on the formation of equivalence classes. Lane & Critchfield, (1996) obtained verbal reports after every trial in the test for equivalence. One limitation of post hoc verbal reports is that it is not known if the verbal behavior collected after the experiment corresponds to what happened during the experiment. Furthermore, by just asking if the participants gave names to the stimuli may lead the participants to create names for the stimuli (Arntzen, 2006).

A fruitful direction of collecting verbal reports is to use concurrent verbal reports using the protocol analysis methodology or talk-aloud protocol (TAP) (Ericsson & Simon, 1993). Ericsson and Simon argued that thinking and verbal reports correspond closest when participants are asked to verbalize at the same time as solving tasks.

Within stimulus equivalence research, relatively few experiments have employed talk-aloud protocols (Garcia & Rehfeldt, 2008; Rehfeldt, Dixon, Hayes, & Steele, 1998; Wulfert, Dougher, & Greenway, 1991). The focus of these studies has been on the content of participant speech, and this content has been correlated with the outcome of tests for responding in accordance with stimulus equivalence. For example, Wulfert et al. asked if it was possible to use TAP to explore and identify variables that could explain the individual differences in equivalence yields. The participants were trained to talk aloud before the experiment started, and verbal responses during the experiment session were transcribed. The participants were instructed to talk aloud during the entire experiment and then given math tasks to assess comprehension by talking aloud. If the participants just answered the math task, they were then given a new math problem and asked to solve the math task aloud once again. When the participants solved the math task by talking about how they solved the math task, the experimental session started. The verbal responses were categorized into four categories. The first category was relational responding, which described the relation between stimuli (“the finger goes together with the pig”), to the nodal stimulus (“both stimuli went with the pig”), or to a common name for the class (“both belong to the pig”). The second category was common physical features (“both stimuli have 30-degree angles”). The third category

was stimulus compounds, which referred to that participants talked about visually integrating the stimuli (“together the stimuli form an a”). The fourth category referred to all responses that were not classified in the previous categories. The main finding was that a high yield of equivalence outcome was highly correlated with participants talking about relations among the stimuli (the first category).

In another experiment by Garcia & Rehfeldt, (2008, experiment 2), participants were pretrained to give common names (e.g., VEX to one class, and ZAP to another class) or common simple fixed ratio (FR) responses to each stimulus (e.g., FR1 to one of the classes and FR3 to one of the other classes) before an MTS procedure was initiated. Also, the participants had to talk aloud during both MTS training and testing. The main findings were that the participants talked either about the description of relations between stimuli based on the names or the numerical values or single names/numerical values in almost all cases. Finally, Rehfeldt et al., (1998) trained four participants the necessary baseline relations to test for equivalence class formation when they were required to talk aloud during training and testing. The results showed that during the test for stimulus equivalence, two of the four participants talked about descriptions of sample and comparison stimulus relations in 98% of the trials and the remaining two in half of the trials (47% and 50%). However, they found no correlation between high accuracy on the matching performance and talking aloud about the sample and comparison stimulus relations.

In sum, the three studies described above showed that the participants often talked aloud about the relations between the sample and the experimenter-defined correct comparison both during the establishment of conditional discriminations and during testing for responding in accordance with stimulus equivalence. In the present study, we wanted to examine a different dimension of verbal behavior than what has been previously studied. Specifically, we wanted to investigate whether participants would talk in different time-windows of DMTS trials when instructed to talk aloud during training and testing for responding in accordance with stimulus equivalence. Hence, we had three recording periods: (a) the first 18 training trials, (b) the last 18 training trials, and (c) the test trials. The recordings were employed in three time-windows: (1) when the sample was presented, (2) the delay from the offset of the sample to the onset of the comparisons, and (3) when the comparison array was presented. These three recording periods were included in order to assess any differences in the amount of talking aloud as a function of the MTO or OTM condition. In the MTO condition, there is always only one comparison stimulus that is correct to each sample during training (e.g., the A or B stimuli are presented as samples, and the array of C stimuli is always presented as comparisons), while in the OTM condition, there are several correct comparisons to each sample during training (e.g., the A stimuli are always presented as samples, and either the comparison array of B stimuli or C stimuli is presented as comparisons). Therefore, one prediction is that the talking in the comparison window could be less frequent for the MTO compared to the OTM condition at the end of training since it is easier to predict the correct comparison in the MTO condition. If the prediction holds true, then there should be no such difference during testing, since each sample in the MTO condition now is related to several comparison stimuli as in the OTM condition (e.g., in the MTO condition, if the A stimulus is presented as the sample, then the correct comparison can be the C stimulus, as in a baseline trial, or the B stimulus, as in an equivalence trial). Finally, we wanted to find out if there was a correspondence

between the verbal and nonverbal behavior which would replicate the findings from previous studies conducted with talk-aloud procedures. A correspondence between verbal and nonverbal behavior in DMTS is important for the understanding of variables that influence equivalence class formation when testing is arranged with delays.

Method

Participants

Eighteen students, five males and 13 females, ranging from 18 to 55 years of age, with a mean age of 28, participated in the study. Participants were recruited from two different universities in Oslo. They had no previous experience with either stimulus equivalence experimental preparations or the stimuli used in the experiment. When the participants entered the experimental setting, they were asked to read and sign an informed consent form, which oriented them to the background of the project and the experimental situation. After finishing the experimental session, participants were fully debriefed and thanked for their participation. They did not receive any extra credit or money for serving as participants.

Design

In the current experiment, two AB designs were used. In the first AB design, an MTO training structure was used to train baseline relations in the A-phase. The B-phase consisted of a test for responding in accordance with stimulus equivalence. The second AB design was arranged the same as the first AB design, except that the baseline relations were trained with an OTM training structure. Both AB designs included eight replications. The MTO replications were done first, and then all of the OTM replications. In addition, the MTO and the OTM participants were compared as two separate groups.

Setting and Apparatus

The experiment was conducted in an office cubicle containing a table and a chair. The cubicle measured 4.83 m². The matching-to-sample tasks were presented on a laptop computer with a 17-inch screen. The participants clicked on the stimuli with a computer mouse. A custom-made matching-to-sample software was used. Data were gathered automatically by the software and included the number of training trials, sample stimulus, selected comparison stimulus, and whether the selected comparison choice was in accordance with the experimenter-defined classes or not. The program also recorded the reaction time for responses to the sample and comparison and summarized the different test trial types. CamStudio 2.7 (<http://camstudio.org>) was used to record the screen and the sound from the onboard laptop computer microphone. MorphVOX Jr. (<http://screamingbee.com/Product/MorphVOXJunior.aspx>) was used to mask the voice and to secure the anonymity of the participants. The experiment was conducted in one session, and the experimental sessions lasted from 50 min to 2 h and 14 min, with a median time of 1 h and 18 min.

Stimuli

The stimulus set included nine abstract shapes and were the same stimuli that were used by Arntzen & Vie, (2013). The stimuli were black on a white background as shown in Fig. 1. The stimuli were presented on rectangular zones on the screen with the stimuli

Fig. 1 Training structures and stimuli used in the experiment. The *numbers* indicate the experimenter-defined classes and the *letters* indicate the different class members. The *letter-number* combination is only for illustration purposes and not visible for the participants. The left-most *rectangles* show the possible sample stimuli during training of baseline relations. The *dotted lines* indicate that each of the samples can be presented as sample. The *solid lines with arrows* point to the possible comparison arrays for each training structure. The bottom of the figure indicates the different time-windows which trials were separated. Time-window 1 was defined as when the sample was at the screen, time-window 2 was when the screen was blank for 6 s, and time-window 3 was when the comparison array was present

averaging in size of 3.5×2.5 cm. The rectangular zones were sensitive to participants' mouse clicking. The click-sensitive areas were 10.5×3.8 cm. The sample was presented to the center of the screen, and the comparisons were presented in 3 of the 4 corners of the screen. The position of the presentation of comparisons was randomized by the software, without any constraints. Furthermore, the stimulus classes and their members were the same across participants. The comparisons were organized in a distance of 17 cm diagonally from the sample stimulus. The mouse cursor was set to a fixed position, 3 cm above the left side of the click-sensitive area of the sample before the sample stimulus was presented.

Procedure

Both training and testing were arranged as a DMTS procedure. A click on the sample was followed by the offset of the sample and an onset of the comparisons after 6 s. The participants were instructed to talk aloud in three parts of the experiment: in the first and last part of training and throughout the testing.

Instructions The TAP instructions in the present experiment were the same as in Wulfert et al., (1991) and several other TAP experiments within stimulus equivalence research (e.g., Garcia & Rehfeldt, 2008). Before the experimental session started, the participants were asked to solve a math task to see if they talked aloud in the correct manner. The task was “What is 127 plus 35?” If the participants just stated the answer “162,” the experimenter demonstrated how the participants should solve another problem such as 123 plus 42: “123 plus 12 makes 135, plus 30 makes 165.” The participants were then given a similar math problem and were instructed to talk aloud. This demonstration was repeated until the participants had solved two consecutive math problems as described above. They were then instructed to read the instruction on the computer screen. The instructions were displayed in Norwegian on the computer screen, and read as follows:

A stimulus will appear in the middle of the screen. You should click on it with the left mouse button. Three other stimuli will appear. Choose one of these with the use of the left mouse button. If you choose the stimulus we have defined as correct words like “correct”, “great”, etc. will be presented on the screen. If you press an incorrect stimulus, “wrong” will be presented on the screen. At the bottom of the screen, the number of correct responses will be counted. During some stages of the experiment, the computer will not give feedback about whether the choices are correct or wrong. However, based on what you have learned, you can get all the tasks correct. We are interested in understanding how people solve problems. You should from the beginning of the experiment talk out loud what you are thinking. What you are saying will be recorded. Your voice will be distorted to keep you anonymous. It is important that you do not say your name or something else that can identify you as a participant. Do your best to get everything right. Good luck!

Following the presentation of the first 18 training trials, an instruction was presented on the screen with the following message “You can now stop talking aloud.” The

participants were asked to stop talking aloud after 18 trials to reduce the amount of data to be scored. When the programmed consequences were thinned to 0% programmed consequences, a new instruction was presented on the screen with the following message “You are now going to start to think out loud again”.

Training In training, programmed consequences were presented on the screen for 500 ms followed by an intertrial interval of 500 ms. The participants were either exposed to an MTO training structure or an OTM training structure with all the baseline relations presented in one block. In these blocks, baseline relations were presented three times, which resulted in blocks consisting of 18 trials. The trials in each block were presented in a randomized order. When the participant responded 100% correctly within a training block, the probability of programmed consequences was thinned to 75%, then 50, 25, and 0%. If the participants did not respond 100% correctly, then a block with the same probability of programmed consequences was presented again, with the trials in a randomized order.

In the MTO training structure, the AC trials and BC trials were mixed in blocks of 18 trials, A1/C1C2C3, A2/C1C2C3, A3/C1C2C3, B1/C1C2C3, B2/C1C2C3 and B3/C1C2C3. In this notation, trials are illustrated by first referring to the sample, then using a slash mark to separate the sample from the comparison array with the underline illustrating the experimenter-defined correct comparison. Each trial type was presented three times within a block.

In the OTM training structure, AB trials and AC trials were mixed in blocks with 18 trials (A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, A1/C1C2C3, A2/C1C2C3 and A3/C1C2C3). Each trial type was presented three times.

Testing The test block for responding in accordance with stimulus equivalence was initiated after the thinning of the programmed consequences. No programmed consequences were presented under testing. The test block consisted of baseline trials, symmetry trials, and equivalence trials, with a total of 54 trials in the test block. Furthermore, the criterion for responding in accordance with stimulus equivalence was set to 90% of baseline, symmetry, and equivalence trials.

In the MTO training structure, the test for symmetry consisted of C1/A1A2A3, C2/A1A2A3, and C3/A1A2A3, C1/B1B2B3, B2/B1B2B3, B3/B1B2B3 trials, and the test for equivalence consisted of A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/A1A2A3, B2/A1A2A3, B3/A1A2A3 trials.

In the OTM training structure, the symmetry test consisted of B1/A1A2A3, B2/A1A2A3, B3/A1A2A3, C1/A1A2A3, C2/A1A2A3, and C3/A1A2A3 trials, and equivalence test consisted of B1/C1C2C3, B2/C1C2C3, B3/C1C2C3, C1/B1B2B3, C2/B1B2B3, and C3/B1B2B3 trials.

Scoring and Interobserver Agreement of the Verbal Reports

When scoring the verbal reports, the first 18 trials of the experiment, the 18 last trials in the phase with 0% programmed consequences, and the whole test for responding in accordance with stimulus equivalence were scored. Specifically, each trial was separated into three time-windows, and it was scored whether

participants talked in the time-windows or not. Time-window 1 was when the sample stimulus was present. Time-window 2 was during the delay. Time-window 3 was when the comparison array was present. If a participant started to talk within one time-window and continued to the next time-window, the talking was considered as a part of the first time-window. For example, if the participants said, “the car goes together with” when the sample was present, and the word “car” followed directly during the delay, then “car” was scored as being a part of the sample time-window. If the participant paused before completing a sentence, while the second or third time-window started, then the last part of the sentence was scored as being a part of the delay time-window or comparison time-window. If the participant talked about the previous trial in a new trial, talking was not considered to be a part of the new trial.

The number of time-windows with talking within each training block was divided by the total number of time-windows within the same block and multiplied by 100. Furthermore, we analyzed the percentage of time-window with talking for time-windows 1, 2, and 3 for the 18 first training trials, the last 18 training trials, and the test trials.

Interobserver agreement was assessed by obtaining 36 trials from three randomly selected participants. The trials were from the last part of training and the test. The first author and another trained experimenter scored the verbal reports individually. The agreement was calculated by dividing the number of agreed occurrences of talking in the time-windows by the total number of occurrences of talking multiplied by 100. The agreement scores ranged from 90.7 to 95.3%, with most of the disagreements occurring during the delay time-window.

Results

Results showed that all participants responded in accordance with stimulus equivalence (see Table 1). Furthermore, the average number of baseline training trials was 288.2 and 304 for MTO and OTM, respectively.

Talk Aloud Results for MTO

As shown in Fig. 2, for the MTO condition, there was a median increase of 55.55% (range, 27.78 to 83.33%) in talking during the sample time-window when comparing the first 18 training trials and the last 18 training trials. When comparing the last 18 training trials and the test, a further increase of 11.11% of time-windows with talking was observed (range, 83.33 to 94.44%).

For the delay time-window, the participants showed a median increase in time-windows with talking of 27.77% (range, 5.56 to 33.33%) from the first 18 training trials to the last 18 training trials. When comparing the last 18 training trials to the test, the number of time-windows with talking decreased by 9.26% (range, 33.33 to 24.07%).

For the comparison array window, the talking decreased from the first 18 training trials to last training 18 trials by a median of 27.78% (range, 55.56 to 27.78%) and increased by 64.81% in the test (range, 27.78 to 92.59%).

Table 1 Number of correct trials on the equivalence tests

P no.	TBR	Test			ECF
		BR	SYM	EQ	
MTO training structure					
3165	234	18	18	17	Y
3166	180	18	18	18	Y
3167	216	18	17	18	Y
3169	198	18	18	18	Y
3171	306	18	18	18	Y
3172	324	18	18	18	Y
3173	506	18	18	18	Y
3174	360	18	18	18	Y
3176	270	18	18	18	Y
OTM training structure					
3180	198	18	18	17	Y
3181	324	18	17	18	Y
3182	198	18	18	18	Y
3183	198	18	18	18	Y
3185	224	18	18	18	Y
3187	324	18	18	18	Y
3188	540	18	18	18	Y
3191	378	18	18	18	Y
3217	342	18	18	18	Y

P no. Participant number, *TBR* training of baseline relations, *BR* baseline relations, *SYM* symmetry, *EQ* equivalence, *ECF* equivalence class formation

Talk Aloud Results for OTM

For the OTM condition, the talking in all three time-windows increased from the first 18 to the last 18 training trials and continued at the about same level in the test (see Fig. 3). There was a median of 33% increase (range, 66 to 100%) in talking in the sample time-window when comparing the first 18 training trials and the last 18 training trials. While comparing talking in the last 18 training trials with the test, there was a median of 2.78% decrease (range, 97.22 to 100%).

For the delay time-window, talking increased by a median of 16.66% (range, 16.67 to 33.33%) from the first 18 training trials to the 18 last training trials. Then, from the last 18 training trials compared to the test, there was a 2.67% (range, 33.33 to 36%) increase in talking.

From the first 18 training trials to the last 18 training trials, in the comparison array time-window, there was an increase in talking of 27.78% (range, 61.11 to 100%). When

Fig. 2 The median percentage of the number of time-windows with talking in the first 18 training trials, the last 18 training trials, and the test trials for all participants in the MTO condition. The *error bars* show the median absolute deviation

comparing the last 18 training trials to the test, there was a 3% decrease in talking (range, 97 to 100%).

When all delay time-windows are combined, all participants talked less than 61% (range, 15 to 61%) of the total number of delay time-windows.

Correspondence between Verbal Behavior and Matching Performance

All participants who talked in more than 50% of the total number of time-windows in the MTO condition generally showed a pattern of saying two names when the sample (one about the sample name and one about the comparison) was present at the end of training. Figure 4 shows the correspondence between verbal behavior and the matching

Fig. 3 The median percentage of the number of time-windows with talking in the first 18 training trials, the last 18 training trials, and the test trials for all participants in the OTM condition. The *error bars* show the median absolute deviation

performance for the MTO and OTM conditions. Hence, the participants named the comparison stimulus that they selected in 94.44% of the cases in the MTO condition. For example, in the first five trials in the last training block (the block with 0% probability of programmed consequences), P3165 said “This, I think resembles a cup and then I think that the cup goes together with Y which resembles a W or a Y” (A1C1). Such a vocal response could be described as stimuli that go together. At the end of training, it was the same pattern but the length of the vocal responses was shorter (e.g., “cup to the Y symbol”) (A1C1).

In the test for responding in accordance with stimulus equivalence, the MTO participants, who talked in more than 50% of the total amount of time-windows, generally said one name when the sample was present and another name when the comparison array was present. P3165 showed a pattern in which he did not talk when the sample was present for the first five test trials. However, when the comparison array was presented, the participant said, for example, “Y, then I click on the one which looks like a cup because they both belonged together before” (C1A1). In the last five test trials, P3165 talked during all the five trials when the sample was present on the screen and remained silent until the comparison array was presented. For example, in the equivalence trial A1B1, the participant said “cup” in the presence of the sample (A1) and “Together with the L because they belonged together earlier” in the presence of the comparison (B1) before clicking on B1.

There were also in correspondence in the OTM condition between what the participants were saying and what they were doing (see the right-most bar in Fig. 4 for the OTM condition). The results show that participants said the name of the comparison that they selected in 98.14% of the trials. During the end of training and the test, all participants who talked in more than 50% of time-windows showed in general the pattern of saying one name when the sample was present and another name when the comparison array was present.

As a representative example, during all the first five trials in the last training block (the block with 0% probability of programmed consequences), P3181 talked when the sample and the comparison array were present. For example, saying, “This resembles a spaceship and is associated with” when the sample (A1) was present, and then saying “A backward

Fig. 4 The percentage of correspondence between verbal behavior and nonverbal behavior (talking about and selecting the comparison stimulus) during testing. Participants who talked in less than 50% of the time-windows are not included. The *error bars* show the median absolute deviation

L with a S” (C1) before clicking on C1. The same pattern was also shown in the last five training trials in the same block, with P3181 talking when the sample and comparison array were present. For example, saying “The spaceship” when the sample (A1) was present, and then saying “and apple” in the presence of the comparison (C1) and then clicking on C1. Also, in the first five test trials, the participant talked when the sample was present and when the comparisons were present. For example, saying the apple when the sample (B1) was present in the first test trial, and saying “Rotated, okay, belongs to the symbol which also is associated with the spaceship I think” in the presence of the correct comparison (C1) and clicking on C1. Similarly, in the last five test trials, P3181 talked when the sample was present and when comparison array was present. For example, said “The hole” when the sample (B3) was present, and “The H” in the presence of the correct comparison (C3) and then clicked on C3.

Discussion

The purpose of the present experiment was to study participants’ talking aloud when establishing conditional discriminations and during testing. A DMTS 6-s procedure was used both for training and testing. The recordings (for the first 18 and the last 18 training trials, and the test trials) were employed within three different time-windows (during the presentation of the sample, during the delay, and during the presentation of the comparison array) when the participants were instructed to talk aloud. The main findings were that participants talked more in the first and third time-windows compared to the second time-window. Furthermore, they talked in more comparison time-windows after the MTO training arrangement, when comparing from the last part of training and the test. The participant in the OTM condition did not show the same increase since the OTM participant talked the same amount in both the last part of training and test. Finally, all participants demonstrated a high correspondence between verbal and nonverbal behavior.

Trials to Criterion and Equivalence Class Formation

The results showed no systematic difference of number of trials to criterion between the two training structures, with the average difference only being 16 more trials in the OTM condition compared to the MTO condition, which is in accordance with other studies that have compared MTO and OTM (e.g., Fields, Hobbie-Reeve, Adams, & Reeve, 1999). Other studies have shown a difference in the number of training trials depending on training structures (e.g., Arntzen, Grondahl, & Eilifsen, 2010). Furthermore, all participants formed equivalence classes. This finding is supported by studies which have focused on equivalence class formation as a function of different training structures (e.g., Arntzen & Hansen, 2011).

Talking in Different Time-Windows

The findings from the present experiment show that talking is more related to the time-windows when the sample and the comparison array are present than during the delay. In the present study, the percentage of talking during the delay varied from 16.67 to 36% for all the recording periods. For example, Horne & Lowe, (1996) suggested that repeating the sample may preserve the sample during the delay to function as a selection response when the

comparison array is presented. The present experiment does not confirm that it is necessary to self-echo aloud the sample or the correct comparison during the delay.

Overall, the percentages of talking were higher for the participants in the OTM condition. Seven of the nine participants in the MTO condition showed a substantial increase in talking from training to testing when the comparison array was presented. One interpretation can be based on the joint control account (e.g., Lowenkron, 1998), where it has been argued that an introduction of a more complex task may lead to the need to tact the comparison stimuli. Specifically, the auditory stimuli which are produced by tacting the comparison stimuli are, according to the joint control account, needed to jointly control the correct comparison choice in combination with the response evoked by the sample (Miguel, Petursdottir, Carr, & Michael, 2008).

In MTO, because of how the stimuli are being related, there is only one correct comparison to the sample stimuli as shown in Fig. 1. Hence, at the end of training, it is possible to predict the correct comparison before the comparisons are presented. For example, in MTO the A and B stimuli are presented as samples, and the C stimuli are presented as comparisons during training. However, in the test, there are multiple possible correct comparison stimuli to each sample. For example, if the A stimulus is presented as the sample, the B stimuli or the C stimuli can be presented as the comparisons during testing. In OTM, there are several possible correct comparisons to each sample in both training and testing. Thus, in training, the A stimuli are always presented as samples, and the B and C stimuli are presented as comparisons, which is also the case during testing. Therefore, at the end of the training, there is a possibility that the participants in the MTO talked aloud about the correct comparison when the sample was presented while in the OTM, the participants only talked about the sample when the sample was presented since there were two possible correct comparisons. For example, P3165 in the MTO said the name of the sample and the correct comparison when the sample was present (e.g., saying “Y to circle”) in one of the 18 last training trials. During testing, the participant said aloud the name of the sample stimulus when the sample was present in more than 90% of the cases, and the name of the correct comparison stimulus when the comparison array was present. The same pattern was also shown by the participants exposed to MTO who talked in more than 50% of the total amount of time-windows. On the other hand, the participants exposed to OTM said only one name when the sample was present, and another name when the comparison array was present. For example, P3181 saying “The fist” when the sample was present, and saying “Hits the h” when the comparison array was present. The participants who talked more than 50% of the time-windows showed the same pattern during testing. Findings from the present experiment show that the OTM participants talked 100 and 97% when the comparison array was present in training and test, respectively.

Giving Individual Names

In the current study, the pattern of talking shows that all participants, except P3165 (MTO) and P3181 (OTM), who talked in more than 50% of the time-windows gave individual names to each stimulus. Although the present experiment was not explicitly arranged for participants to name the stimuli individually, the results are in accordance with recent research which has suggested that participants may intraverbally relate

sample and comparisons which facilitated responding in accordance with stimulus equivalence (e.g., Carp & Petursdottir, 2015; Ma et al., 2016; Petursdottir et al., 2015; Santos et al., 2015). The data from experiments with DMTS have shown that longer delays may facilitate responding in accordance with stimulus equivalence (e.g., Amtzen, 2006). There might be a reason to believe that DMTS increases the probability of participants giving names to the abstract stimuli. Furthermore, we did not observe any evidence of participants giving the same name to all the stimuli within a class.

An interesting finding is that P3165 (MTO) and P3191 (OTM) seems to be giving the same name to two members of a class. P3165 said “circle” in the presence of A3 and C3, and P3191 said “angle” in the presence of B1 and C1. One interpretation can be that the participants gave the stimuli names based on physical similarities of the stimuli. In the case of P3165 and the A3 and C3 stimuli, there are components of the stimuli which may resemble circles, the whole A3 stimulus minus the line trough in the middle, and the top left side of C3. Thus, the equivalence trials can be interpreted as the second category from Wulfert et al., (1991)—common physical features. The same type of interpretation applies to P3191. Hence, the B1 and C1 stimuli in the OTM condition. It is important to note that both P3165 and P3191 gave all the other stimuli individual names.

Correspondence between Verbal and Nonverbal Behavior

The results in Fig. 4 suggest that there is a correspondence between verbal and nonverbal behavior. Thus, MTO and OTM participants said a specific name when selecting a particular comparison stimulus in 94.44 and 98.14% of the trials, respectively. For example, P3181 said “spaceship” when selecting B1 even when B1 was the wrong choice, which indicated that P3181 said the name of the stimulus he or she was going to select. It is important to note that the results in Fig. 4 are based on correspondence. Another line of research which could tell us more about the functional relation between verbal behavior and the task-relevant behavior as matching performance. Hence, experiments with the use of the *silent dog* procedure which consists of three controls for evaluating if verbal behavior is controlling nonverbal behavior (see Hayes, White, & Bissett, 1998). The first control is to test if on-task performance is not interrupted by talking aloud. The second control is to test if disrupting ongoing talking alters on-task performance. And the third, control is to test if verbal reports from the first control alter other participants’ performance) have shown that there is a functional relation between the verbalization of participants and their nonverbal behavior (e.g., Alvero & Austin, 2006; Amtzen, Halstadro, & Halstadro, 2009; Cabello, Luciano, Gomez, & Barnes-Holmes, 2004).

Limitations and Further Studies

One issue is how the math task in the current experiment was presented may have affected how the participants talked during the experiment. It could be that making the preliminary training in the present experiment is quite different from the MTS task, and such a difference could be responsible for a relatively low number of vocal responses especially in the delay window. Hence, instead of being presented vocally, the math tasks can be presented like the DMTS tasks, in future research. The participants will be

instructed to talk aloud as the math task is presented on the screen, and when the participants start to talk, the experimenter can make the screen turn white, which would resemble the delay in a DMTS trial.

Future research could evaluate the use of a preliminary training procedure with different names to the different stimuli used or the same name to the stimuli within each class. The use of experimenter-defined names makes it possible to use the software to collect data automatically. Such automatic collection of data will reduce the time used by the experimenters in scoring data.

Conclusion

The present experiment extends the knowledge about participants' possible talking during DMTS arrangements. The implication of the findings in the current study that participants did not talk in most the delay time-windows may indicate that talking during the delay is not necessary to respond correctly on DMTS tasks. Hence, it is important to note that it is not possible to conclude that no behavior is occurring because it could be the case that behavior is unavailable to the experimenter. For example, there may be ongoing behavior that is not reported by the participants. Regarding the difference in amount of comparison time-windows with talking when comparing the MTO and OTM training, it may be that the different training structures leads to unlike ways of solve the DMTS task during training. Finally, the participants demonstrated high correspondence between verbal and nonverbal behavior as have been previously found in stimulus equivalence research with TAP.

Compliance with Ethical Standards

Ethics Statement 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 was obtained from all individual participants included in the study.

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Alvero, A. M., & Austin, J. (2006). An implementation of protocol analysis and the silent dog method in the area of behavioral safety. *The Analysis of Verbal Behavior*, 22, 61–79.
- Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record*, 54, 275–291.
- 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.
- 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. doi:10.1080/15021149.2007.11434281
- Arntzen, E., Grondahl, T., & Eilifsen, C. (2010b). 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.

- Arntzen, E., Halstadro, L. B., Bjerke, E., & Halstadro, M. (2010a). Training and testing music skills in a boy with autism using a matching-to-sample format. *Behavioral Interventions*, 25, 1–15. doi:[10.1002/Bin.301](https://doi.org/10.1002/Bin.301)
- Arntzen, E., Halstadro, L. B., & Halstadro, M. (2009). The “silent dog” method: A procedure for analyzing the impact of self-generated rules when teaching different computer chains in two boys with autism. *The Analysis of Verbal Behavior*, 25, 51–66.
- Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis*, 12, 483–503. doi:[10.1080/15021149.2011.11434397](https://doi.org/10.1080/15021149.2011.11434397)
- Arntzen, E., & Vie, A. (2013). The expression of equivalence classes influenced by distractors during DMTS test trials. *European Journal of Behavior Analysis*, 14, 151–164. doi:[10.1080/15021149.2013.11434453](https://doi.org/10.1080/15021149.2013.11434453)
- Cabello, F., Luciano, C., Gomez, I., & Barnes-Holmes, D. (2004). Human schedule performance, protocol control analysis, and the “silent dog” methodology. *The Psychological Record*, 54, 405–422.
- Carp, C. L., & Petursdottir, A. I. (2015). Intraverbal naming and equivalence class formation in children. *Journal of the Experimental Analysis of Behavior*, 104, 223–240. doi:[10.1002/jeab.183](https://doi.org/10.1002/jeab.183)
- Dugdale, N., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman & H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 115–138). Hove, England: Lawrence Erlbaum Associates.
- Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior*, 58, 123–133. doi:[10.1901/jeab.1992.58-123](https://doi.org/10.1901/jeab.1992.58-123)
- Eilifsen, C., & Arntzen, E. (2011). Single-subject withdrawal designs in delayed matching-to sample procedures. *European Journal of Behavior Analysis*, 12, 157–172. doi:[10.1080/15021149.2011.11434361](https://doi.org/10.1080/15021149.2011.11434361)
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (revised ed.)*. Cambridge, MA: The MIT Press.
- Fields, L., Hobbie-Reeve, S. A., Adams, B. J., & Reeve, K. F. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record*, 49, 703–724.
- Garcia, Y. A., & Rehfeldt, R. A. (2008). The effects of common names and common FR responses on the emergence of stimulus equivalence classes. *European Journal of Behavior Analysis*, 9, 99–120. doi:[10.1080/15021149.2008.11434298](https://doi.org/10.1080/15021149.2008.11434298)
- Hayes, S. C., White, D., & Bissett, R. T. (1998). Protocol analysis and the “silent dog” method of analyzing the impact of self-generated rules. *The Analysis of Verbal Behavior*, 15, 57–63.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 185–241. doi:[10.1901/jeab.1996.65-185](https://doi.org/10.1901/jeab.1996.65-185)
- Horne, P. J., Lowe, C. F., & Harris, F. D. A. (2007). Naming and categorization in young children: V. Manual sign training. *Journal of the Experimental Analysis of Behavior*, 87, 367–381. doi:[10.1901/jeab.2007.52-06](https://doi.org/10.1901/jeab.2007.52-06)
- Horne, P. J., Lowe, C. F., & Randle, V. R. (2004). Naming and categorization in young children: II. Listener behavior training. *Journal of the Experimental Analysis of Behavior*, 81, 267–288. doi:[10.1901/jeab.2004.81-267](https://doi.org/10.1901/jeab.2004.81-267)
- Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior*, 85, 107–124. doi:[10.1901/jeab.2006.58-04](https://doi.org/10.1901/jeab.2006.58-04)
- Lane, S. D., & Critchfield, T. S. (1996). Verbal self-reports of emergent relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior*, 65, 355–374. doi:[10.1901/jeab.1996.65-355](https://doi.org/10.1901/jeab.1996.65-355)
- Lian, T., & Arntzen, E. (2013). Delayed matching-to-sample and linear series training structures. *The Psychological Record*, 63, 545–561.
- Lowe, C. F., Horne, P. J., Harris, F. D. A., & Randle, V. R. L. (2002). Naming and categorization in young children: Vocal tact training. *Journal of the Experimental Analysis of Behavior*, 78, 527–549. doi:[10.1901/jeab.2004.81-267](https://doi.org/10.1901/jeab.2004.81-267)
- Lowenkron, B. (1998). Some logical functions of joint control. *Journal of Experimental Analysis of Behavior*, 69, 327–354. doi:[10.1901/jeab.1998.69-327](https://doi.org/10.1901/jeab.1998.69-327)
- Ma, M. L., Miguel, C. F., & Jennings, A. M. (2016). Training intraverbal naming to establish equivalence class performances. *Journal of the Experimental Analysis of Behavior*, 105, 409–426. doi:[10.1002/jeab.203](https://doi.org/10.1002/jeab.203)
- Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Processes*, 32, 29–46. doi:[10.1016/0376-6357\(94\)90025-6](https://doi.org/10.1016/0376-6357(94)90025-6)
- Miguel, C. F. (2016). Common and intraverbal bidirectional naming. *The Analysis of Verbal Behavior*, 32, 125–138
- Miguel, C. F., Petursdottir, A. I., Carr, J. E., & Michael, J. (2008). The role of naming in stimulus categorization by preschool children. *Journal of the Experimental Analysis of Behavior*, 89, 383–405. doi:[10.1901/jeab.2008-89-383](https://doi.org/10.1901/jeab.2008-89-383)

- Petursdottir, A. I., Carp, C. L., Peterson, S. P., & Lepper, T. L. (2015). Emergence of visual–visual conditional discriminations. *Journal of the Experimental Analysis of Behavior*, 103, 332–348. doi:[10.1002/jeab.136](https://doi.org/10.1002/jeab.136)
- Randell, T., & Remington, B. (2006). Equivalence relations, contextual control, and naming. *Journal of the Experimental Analysis of Behavior*, 86, 337–354. doi:[10.1901/jeab.2006.82-05](https://doi.org/10.1901/jeab.2006.82-05)
- Rehfeldt, R. A., Dixon, M. R., Hayes, L. J., & Steele, A. (1998). Stimulus equivalence and the blocking effect. *The Psychological Record*, 48, 647–664
- Santos, P. M., Ma, M. L., & Miguel, C. F. (2015). Training intraverbal naming to establish matching-to-sample performances. *The Analysis of Verbal Behavior*, 31, 162–182
- Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two, three-, and four-choice matching to sample by senior citizens. *The Psychological Record*, 55, 539–559
- 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](https://doi.org/10.1901/jeab.1982.37-5)
- Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: the role of naming. *Analysis and Intervention in Developmental Disabilities*, 6, 1–19. doi:[10.1016/0270-4684\(86\)90003-0](https://doi.org/10.1016/0270-4684(86)90003-0)
- Tomanari, G. Y., Sidman, M., Rubio, A. R., & Dube, W. V. (2006). Equivalence classes with requirements for short response latencies. *Journal of the Experimental Analysis of Behavior*, 85, 349–369. doi:[10.1901/jeab.2006.107-04](https://doi.org/10.1901/jeab.2006.107-04)
- Vaidya, M., & Smith, K. N. (2006). Delayed matching-to-sample training facilitates derived relational responding. *Experimental Analysis of Human Behaviour Bulletin*, 24, 9–16
- Wulfert, E., Dougher, M. J., & Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal behavior and equivalence class formation. *Journal of the Experimental Analysis of Behavior*, 56, 489–504. doi:[10.1901/jeab.1991.56-489](https://doi.org/10.1901/jeab.1991.56-489)