ON THE ROLE OF BIDIRECTIONAL INTRAVERBAL NAMING ON EQUIVALENCE CLASS FORMATION

Heidi Skorge Olaff[1], Emilie Bjørklund Jørgensen[1], Erik Arntzen[1]

[1] Oslo Metropolitan University

The purpose of this study was to establish tacts and intraverbals to evaluate whether emergent matching-to-sample (MTS) and intraverbal responding occurred, in accordance with the emergence of intraverbal bidirectional naming (I-BiN) and the formation of equivalence classes. This study replicates and extends the work of Ma et al. (2016) by employing a one-to-many (OTM) training structure, instead of a linear series (LS) training structure. In addition, the study included sorting post-tests and a social validity survey. Eight adult participants underwent MTS pre-tests, followed by tact training of experimental stimuli and testing for emergent listener responses. Then, intraverbal training (A'B'/A'C') was introduced before conducting MTS post-tests and I-BiN post-tests, as well as sorting post-tests. A post-experimental interview and a social validity survey completed the study. The main finding indicated that training of baseline intraverbal relations using a OTM training structure, likely facilitated the establishment of new conditional relations in accordance with the emergence of I-BiN towards the experimental stimuli and the formation of equivalence classes. All participants reported using verbal mediation strategies, such as tacts and intraverbals established during training and testing, as well as self-generated tacts and intraverbals.

Keywords: intraverbal bidirectional naming, matching-to-sample, one-to-many, stimulus equivalence, verbal mediation

Equivalence class formation involves physically disparate stimuli that are interchangeable for one other within a class. That is, a few conditional discriminations are trained directly, which results in the emergence of novel relations. Matching-to-sample (MTS) is a type of conditional discrimination procedure that is commonly used to train and test conditional relations, such as stimulus equivalence (cf. Green & Saunders, 1998, p. 232). Conditional discrimination training involves training an individual to respond differently to stimuli, depending on how stimuli are presented. For example, using a linear series (LS) training structure when training to-be-formed three 3-member classes, clicking on sample stimulus A1 reveals, at least, three comparison stimuli, B1, B2, and B3. Then, the individual is trained to select B1 and not B2 or B3. In the presence of A2, the individual is trained to select B2 and not B1 or B3. In the presence of A3, the individual is trained to select B3 and not B1 or B2. Further, in the presence of sample stimulus B1, the individual is taught to select comparison stimulus C1, and not C2 or C3. In the same way, in the presence of B2 as the sample stimulus, C2 is the correct comparison stimulus and not C1 or C3. In the presence of B3 as the sample stimulus, C3 is the correct comparison stimulus and not C1 or C2. The defining properties of stimulus equivalence are reflexivity (A is related to A, B is related to B, and C is related to C), symmetry (B is related to A and C is related B), transitivity (A is related to C) and equivalence (C is related to A; Sidman & Tailby, 1982).

Stimulus equivalence and its relation to verbal behavior have been discussed within behavior analysis. For example, Sidman (2000) argued that responding in accordance with stimulus equivalence is a product of reinforcement in a four-term contingency, rather than the influence of verbal behavior. This argument is supported in experiments where the occurrence of verbal behavior is less likely. For example, equivalence class formation was achieved in participants with limited verbal repertoires (Carr et al., 2000) and in participants

Author Note: Emilie Bjørklund Jørgensen is affiliated at Gamlebyen Primary School.

Correspondence Address: Department of Psychology, Harald Schjelderups hus, Forskningsveien 3A, N-0851 Oslo, Norway.

Email: holaff@oslomet.no

Email: erik.arntzen@equivalence.net

who were exposed to a condition that required rapid responding (Tomanari et al., 2006).

However, Horne and Lowe (1996) suggested that verbal behavior and in particular the naming theory (speaker-listener relations) could account for a diversity of emergent behavior, including stimulus equivalence. They argued that a participant with naming skills who is trained during a MTS procedure will, possibly, name the sample stimulus (speaker behavior), which may evoke a selection response (listener behavior) of the corresponding comparison stimulus. Later, Miguel (2016, 2018) used the term “bidirectional naming” (BiN) to describe the integration of both listener and speaker behavior. He argued that the absence of either speaker or listener behavior can cause participants to fail to solve MTS or sorting tasks. To solve such tasks, participants must either react to their own speaker behavior or engage in an unobservable behavior (e.g., covert echoic). This process is established from an early age through repeated exposures, and one of the characteristics that can contribute to the creation of a BiN repertoire is tact relations.

Participants can, probably, successfully form equivalence classes using various bidirectional naming or mediating responses, and one of these strategies is intraverbal bidirectional naming (I-BiN). I-BiN involves training tacts and intraverbal relations to establish novel or emergent stimulus relations (Miguel, 2016, 2018). Thus, I-BiN means that tacts set the occasion for intraverbals, leading to novel intraverbals emerging without the relations being directly trained or reinforced. The I-BiN relations among auditory stimuli may facilitate grouping corresponding visual stimuli into classes and, thus, forming equivalence classes.

Recently, Schlinger, and Blakely (2024) argued that vocal-mediating responses (e.g., echoic, tacts, and intraverbals) present in the moment of reinforcement are directly taught during training of baseline relations and are essential for correct selection responses during stimulus equivalence tests. Several experiments have explored the role of vocal-mediating behavior during complex tasks (e.g., MTS tasks, sequencing tasks, and listener tasks) and whether this behavior impacts solving the relevant tasks (e.g., Clough et al., 2016; Jennings & Miguel, 2017; Ma et al., 2016; Miller et al., 2021; Petursdottir et al., 2019; Vie & Arntzen, 2019; Zaring-Hinkle et al., 2016).

Several studies (e.g., Chastain et al., 2022; Jennings & Miguel, 2017; Pérez-González & Oltra, 2023; Petursdottir et al., 2015; Petursdottir et al., 2019; Santos et al., 2015) support tacts and intraverbals as possible verbal-mediating responses and the findings suggest that the presence I-BiN may have a facilitative role on the formation of bidirectional relations and equivalence classes. In this line of research, the participants were subjected to tact training, listener post-test, and training of intraverbal relations. Then, posttests of I-BiN (i.e., baseline, symmetry, and transitivity/equivalence relations) and MTS tests (i.e., baseline-like, transitivity/equivalence-like relations) were completed. In these experiments, MTS baseline-like relations were not directly trained, merely tested along with the properties of equivalence relations.

Ma et al. (2016) investigated, in three experiments using a LS training structure, whether intraverbal training would be sufficient to form equivalence classes, and whether emergent intraverbal behavior would occur simultaneously with MTS tests selection responses. Tact training was implemented (AA’/BB’/CC’) followed by tests for listener responding (A’A/B’B/C’C; c.f. common BiN). Subsequently, intraverbal training (A’B’/B’C’) was initiated. Then, participants were exposed to MTS post-tests (visual stimuli; baseline-like, symmetry-like, transitivity-like and equivalence-like relations; AB, BC, BA, CA, AC and CA relations) and I-BiN post-tests (auditory stimuli; baseline, symmetry, and transitivity and equivalence relations; A’B’, B’C’, B’A’, C’B’, A’C’ and C’A’ relations). The results showed that emergent intraverbal relations (cf. I-BiN) and MTS responses correlated with the formation of equivalence classes.

The present study aimed to replicate and extend Experiment 2 in Ma et al. (2016) by using a one-to-many (OTM) training structure, rather than a LS training structure during intraverbal training (i.e., baseline relations), I-BiN post-tests (baseline, symmetry, transitivity, and equivalence relations) and MTS tests (i.e., baseline-like, symmetry-like, transitivity-like and equivalence-like relations). Although, some studies have replicated or extended Ma et al. (e.g., Chastain et al., 2022; Jennings & Miguel, 2017; Petursdottir et al., 2019), there is a gap in this line of research as all of them included participants from United States only and, they were using either a LS (Jennings & Miguel, 2017)

Figure 1. The Experimental Stimuli.

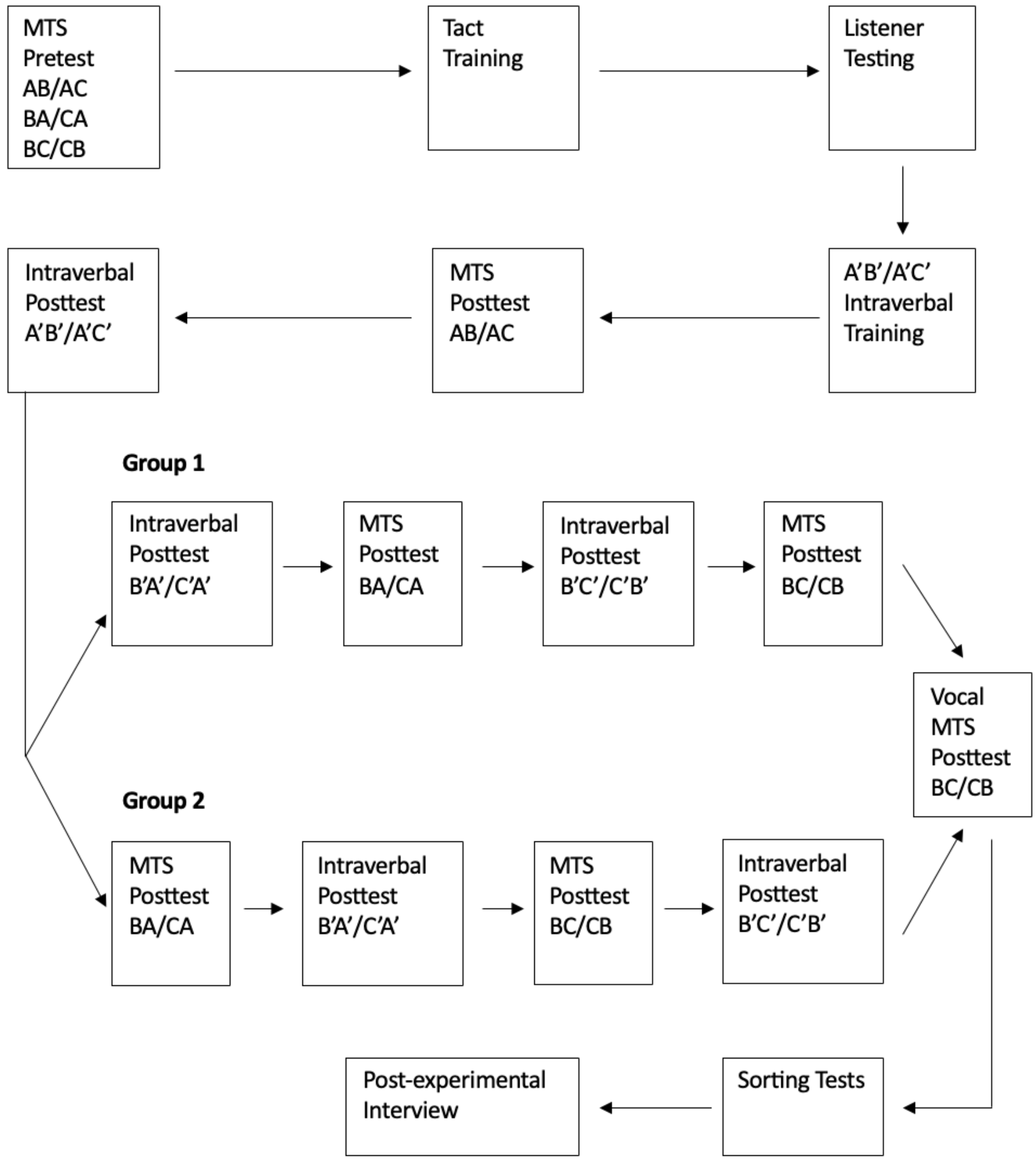
Note. A, B, and C are members of the experimenter-defined classes, and numbers designate the classes.

or many-to-one (MTO) training structure (Chastain et al., 2022). Despite this, several studies have found OTM to be the most effective in the context of training conditional relations to form responding in accordance with stimulus equivalence (e.g., Arntzen, 2004; Arntzen & Hansen, 2011), but use of this training structure is sparse regarding intraverbal training to form stimulus equivalence classes. To date, researchers have not investigated a OTM training structure when assessing for the emergence of I-BiN and stimulus equivalence. If verbal mediation is an important variable in producing equivalence classes, it should be important to arrange procedures using other training structures than LS. Another extension was the use of sorting post-tests (i.e., whether participants can "group" the stimuli into experimenter-defined classes) to confirm the results from the MTS post-tests, as an alternative mean to measure yields in stimulus equivalence research (e.g., Arntzen et al., 2017).

A final extension was the inclusion of a social validity questionnaire. Basic research is supposed to provide and influence changes in the real world and should be linked to the applied field to facilitate an understanding of basic principles (Malkin et al., 2024). Because the present study involves human participants learning general facts, we considered social validity assessment relevant, seeking to produce meaningful behavior change that produces essential outcomes (Huntington et al., 2023). Including social validity to evaluate practical outcomes that the research obtains for improving people's life, could potentially minimize the barriers of application of basic research (Malkin et al., 2024).

The purpose of this study was four-fold: (1) to establish tacts and intraverbals to investigate the prevalence of emergent intraverbals (I-

Figure 2. Order of Training and Testing Conditions.



Note. P1, P2, P3, and P4 were exposed to intraverbal post-tests before MTS post-tests. P5, P6, P7, and P8 were exposed to MTS post-tests before intraverbal post-tests. P is an abbreviation for Participant.

BiN)L, verbal-mediating responses as possible processes that may influence responding in accordance with stimulus equivalence, (2) to investigate whether changing to a OTM structure affects the formation of equivalence classes, (3) to evaluate whether the MTS-results were confirmed during the sorting post-tests and (4) to extend I-BiN studies by a social validity survey to evaluate how the participants experienced the procedures in the present experiment.

METHOD

Participants

The participants comprised eight adults (P1–P8), two men and six women, between the ages of 21–

Table 1. Intraverbal Relations During Training and Tests.

Relations	Antecedent Stimuli		Correct Responses
Baseline Relations Trained			
A -> B Fruit -> Plant	The plant for (A1) is…	(A1) Passion Fruit	(B1) Cactus
	The plant for (A2) is…	(A2) Lychee	(B2) J K
	The plant for (A3) is…	(A3) Watermelon	(B3) Raffia Palm
A -> C Fruit -> Animal	The animal for (A1) is…	(A1) Passion Fruit	(C1) Bush Dog
	The animal for (A2) is…	(A2) Lychee	(C2) Monkey Eagle
	The animal for (A3) is…	(A3) Watermelon	(C3) Okapi
Symmetry Test Trials			
B -> A Plant -> Fruit	The fruit for (B1) is…	(B1) Cactus	(A1) Passion Fruit
	The fruit for (B2) is…	(B2) J K	(A2) Lychee
	The fruit for (B3) is…	(B3) Raffia Palm	(A3) Watermelon
C -> A Animal -> Fruit	The fruit for (C1) is…	(C1) Bush Dog	(A1) Passion Fruit
	The fruit for (C2) is…	(C2) Monkey Eagle	(A2) Lychee
	The fruit for (C3) is…	(C3) Okapi	(A3) Watermelon
Equivalence Test Trials			
B -> C Plant -> Animal	The animal for (B1) is…	(B1) Cactus	(C1) Bush Dog
	The animal for (B2) is…	(B2) J K	(C2) Monkey Eagle
	The animal for (B3) is…	(B3) Raffia Palm	(C3) Okapi
C -> B Animal -> Plant	The plant for (C1) is…	(C1) Bush Dog	(B1) Cactus
	The plant for (C2) is…	(C2) Monkey Eagle	(B2) J K
	The plant for (C2) is…	(C3) Okapi	(B3) Raffia Palm

Note. J K = Japanese Knotweed.

26. These participants were recruited from the social network of the second author (the experimenter). Before the experiment commenced, the participants were asked whether they had previous experience with stimulus equivalence experiments. None of them had such experience. The participants were informed that the study aimed to establish answers to general-fact questions and that they would be fully debriefed after finishing the experiment. The participants signed an informed consent that was obtained before the experiment was initiated. All participants agreed to attend a 2.5-hour session with a 5-min break between conditions.

Setting and Materials

Experimental sessions were conducted in the laboratory for stimulus control studies at the university. The laboratory cubicle, 1.5 m2, contained a chair, a table, and a computer with a 17-inch screen, 36 cm wide and 23 cm high.

In this study, three 3-member classes were used: fruits (A), plants (B), and animals (C; see Figure 1). The images were retrieved from Word Creative Commons. For MTS testing, an HP ProBook 440 running Microsoft Windows 10 Education was used to conduct the experimental sessions. During MTS tests, we used an MTS software, developed by the third author. A trial started with the sample stimulus displayed in the middle of the screen, and a mouse-click on

Table 2. Mastery and Test Criteria for Each Experimental Condition.

Condition	Targets per Block	Trials per Block	Number of Attempts	Mastery/Test criterion (%)
MTS Pretest AB/AC	6	18	1	≤50
MTS Pretest BA/CA	6	18	1	≤50
MTS Pretest BC/CB	6	18	1	≤50
Tact Training	9	27	N/A	1 block at 100
Listener Testing	9	27	2	1 block at 100
A'B'/A'C' Intraverbal Training	6	18	N/A	1 block at 100
AB/AC MTS Posttest	6	18	2	1 block at ≥94
A'B'/A'C' Intraverbal Posttest	6	18	2	1 block at ≥94
B'A'/C'A' Intraverbal Posttest	6	18	2	1 block at ≥94
BA/CA MTS Posttest	6	18	2	1 block at ≥94
B'C'/C'B' Intraverbal Posttest	6	18	2	1 block at ≥94
BC/CB MTS Posttest	6	18	2	1 block at ≥94
BC/CB MTS Vocal Posttest	6	18	1	1 block at ≥94
Post Sorting Tests	9	1	3	100

Note. Targets per block mean number of different stimuli or relations trained or tested, while number of trials per block indicate how many times the targets are repeated in the same block.

the stimulus was followed by the presentation of three comparison stimuli at three of the four corners of the computer screen (simultaneous MTS). The location of the comparison stimuli varied across the experiment. The size of each stimulus was ca. 9.4 cm x 3.4 cm. We used a MacBook Air 2018 model (macOS Ventura version 13.3.1) for tact training and testing auditory-visual MTS (i.e., listener responses), where the relevant stimuli were presented in PowerPoint presentations. The images during tact training and the listener post-test were 14.5 cm x 10 cm, exhibited in PowerPoint presentations. During sorting post-tests, we used laminated images, sized 4 cm x 3.5 cm. Intraverbal training and tests of the emergence of I-BiN were conducted without the presence of visual stimuli (see Table 1).

Video recordings were used for trial blocks involving vocal responses (tacts and intraverbals) during training and testing, including listener post-tests, to measure reliability and procedural integrity. All vocal responses during tact training, intraverbal training, I-BiN post-tests, vocal MTS post-tests of BC/CB relations and selection responses during listener post-tests were scored by human observers, by pencil and prepared scoring sheets.

Experimental Design

In this study, we used a pre-posttest design. The design was set up to compare two groups of participants, four participants in each group. The groups were exposed to two different experimental sequences (Figure 2) to control for potential sequence effects.

Dependent Variables and Response Measurement

The primary dependent variable was the number of correct stimulus selections in the tests arranged in a MTS format. Selection responses were recorded automatically by the experimental software. The participants used the computer mouse to click on the correct comparison stimuli within 5 s. The test criterion for the MTS pre-test was set at 50% or less correct responses. During MTS post-tests, the test criterion was 17 out of 18 correct responses (94%) or above in two blocks (see Table 2 for details).

The secondary dependent variable was the number of emergent vocal responses during I-BiN post-tests (no visual stimuli were present). A correct intraverbal was scored if the participant vocalized the response (e.g., B1') corresponding with the presented auditory sample stimulus (e.g., A1'). When the participant vocally responded correctly to an intraverbal antecedent within 5 s, the experimenter marked a plus in the relevant data recording sheet. If an incorrect vocal response occurred during intraverbal training of baseline relations and I-BiN tests (e.g., no response, the response occurred after 5 s had elapsed or the intraverbal stated did not correspond with the auditory antecedent), the experimenter marked a minus. Participants received a maximum of two blocks of I-BiN post-tests (symmetry and equivalence relations) with a test criterion of at least 17 out of 18 (94%) correct responses or higher within one block (see Table 2).

The third dependent variable was correct listener responses. During the listener post-test, a correct response was scored if the participants selected the corresponding stimulus to the auditory conditional stimulus. For example, in the presence of A′, the correct response was to select A within 5 s. Alternatively, if auditory conditional stimulus and the selected stimulus did not correspond, a minus was scored. The test criterion for the listener post-test was 27 out of 27 (100%) correct responses in at least one of two blocks. If the listener test criterion was not met, the participant was subjected to additional blocks of tact training.

Other dependent variables were vocalizations during the vocal MTS post-test and the number of correct stimuli sorted during sorting post-tests. During vocal MTS post-tests, the experimenter scored the number of vocalizations and whether the participants used (1) established tacts, (2) learned or emergent intraverbals, or eventually, whether (3) self-generated tacts or (4) intraverbals were used (i.e., tacts and intraverbals made up by the participants). The vocal reports during post-experimental interviews (see Appendix A) were transcribed into the same categories as the vocal MTS post-test. Finally, the number of stimuli grouped correctly in the sorting post-tests was photographed with a mobile camera for further analysis.

To describe the results of the sorting test, a three-string system consisting of clusters of numbers was used, each cluster indicating the sorting of the experimenter-defined classes (see, for example, Arntzen et al., 2017). The numbers in each cluster represented how many stimuli from each class were placed correctly (e.g., a result of "300 030 003" signifies that all stimuli in each class were grouped correctly, as each class consisted of three members). The sorting tests were conducted using a tabletop format, and thus, images of participants' sorting were captured with mobile cameras to assess reliability.

As in previous studies (e.g., Ma et al., 2016; Petursdottir et al., 2019), the dependent variable was trials to criterion during tact and intraverbal training. A correct tact trial was recorded (a plus was noted in the data sheet) if the participant, within 5 s, vocally uttered the name of the stimulus (e.g., A1A1′) presented on the computer screen. An incorrect response was scored (by marking a minus in the datasheet) if the participant vocally gave other names or if the response occurred after 5 s had elapsed. The tact mastery criterion was 27 out of 27 (100%) correct responses in one block (see Table 2) for continuation to the listener post-test.

During intraverbal training, a correct vocal response was recorded if the vocal response (e.g., B1′) occurred in the presence of the corresponding intraverbal antecedent (e.g., A1′; an auditory antecedent) within 5 s. However, a minus was scored if an incorrect intraverbal was emitted or the response was given 5 s after the presentation of the antecedent. During intraverbal training, the mastery criterion was 18 out of 18 (100%) correct responses in one block.

Interobserver Agreement and Procedural Fidelity

Interobserver agreement (IOA) and procedural fidelity (PF) were assessed for 50% of the trial blocks during conditions that were not automatically recorded by the MTS software (i.e., tact training, listener post-tests, intraverbal training and I-BiN post-tests) by a second observer. The mean IOA for the training conditions across participants was 97.5%, range 93.4–99.2%, and for the test conditions a mean of 97.3% was obtained, range 91.2–100%.

PF was collected to ensure trials were implemented correctly within manually completed conditions, including: (a) the appropriate use of prompts (e.g., 0-second and 5-s delayed prompts), and (b) the delivery of consequences (e.g., using praise and the prompting procedure during training, and no programmed consequences during testing). The mean PF for training conditions across participants was 95.4%, range 93.9–97.2%, and for test conditions a mean of 98.7%, range 97.2–99.6%.

General Procedure

An overview of the experimental sequences across Groups 1 and 2 is provided in Figure 2. The order of the conditions for each participant was similar as in Ma et al. (2016), except for the pre-training and a review that were omitted in the present experiment (cf. Jennings & Miguel, 2017). All mastery and test criterions are described above, under Dependent Variables and Response Measurement.

Participants were exposed to only one block of the MTS pre-test of all relation types (AB, AC, BA, CA, BC and CB relations) to reduce experience and familiarization with the experimental stimuli before the training conditions commenced. Providing one MTS pre-test deviated from Ma et al. (2016) that exposed one participant in a tier to two MTS pre-test blocks.

An additional distinction from Ma et al. (2016) was the use of a OTM structure (instead of LS) during intraverbal training and testing, and MTS tests. Another variable that deviated from Ma et al., during intraverbal training, I-BiN tests and MTS tests, was that we allowed 5 s to respond to comparison stimuli (cf. Petursdottir et al., 2019) rather than 10 s. During the I-BiN post-tests and the MTS post-tests, the participants were given two attempts, as in Ma et al. (2016).

Vocal MTS post-tests and sorting post-tests were given as a final condition for all participants, of which the latter was an extension from Ma et al. (2016). A post-experimental interview and a social validity survey completed the present experiment.

The MTS arrangement in the pre-test consisted of AB/AC (labeled as baseline relations in Ma et al., 2016), BA/CA relations (labeled as symmetry relations in Ma et al.), and BC/CB relations (labeled as transitivity relations by Ma et al.)—18 trials for each trial type—a total of 54 trials, while the MTS post-tests were split up into relation types (an 18-trial block for each relation type). Thus, each relation type was tested separately. For both groups, post-tests of MTS baseline-like and intraverbal baseline relations were set up in an equal sequence: First MTS baseline-like relations were tested and then, intraverbal baseline relations (see Figure 2). Next, the two groups underwent different sequences of post-tests.

Both groups were exposed to the following similar experimental sequences: MTS pre- test, tact training, listener post-test, A′B′/A′C′ intraverbal training, AB/AC post-test of baseline-like relations, and A′B′/A′C′ post-test of baseline relations (Figure 2). Then, the sequences of MTS and I-BiN post-tests were set up to control for sequence effects and thus, counterbalanced between the two groups. Group 1 (P1–P4) underwent the following sequence of post-tests: B′A′/C′A′ post-test of symmetry relations (I-BiN test), BA/CA post-test of symmetry-like relations (MTS test), B′C′/C′B′ post-test of equivalence relations (I-BiN test) and BC/CB post-test of equivalence-like relations (MTS test). Group 2 (P5–P8) experienced post-tests in a reversed order, the symmetry-like and equivalence-like MTS post-tests were carried out before the I-BiN post-tests (e.g., BA/CA MTS post-test before B′A′/C′A′ I-BiN post-test; Figure 2). After MTS and I-BiN post-tests, both groups underwent a BC/CB vocal post-test of equivalence-like relations and then three sorting post-tests.

During all test conditions, no programmed consequences were delivered (MTS pre- and post-tests, listener post-tests, I-BiN post-tests, vocal MTS post-tests and sorting post-tests). In contrast, during the training conditions, the experimenter gave praise (e.g., “Excellent”) for all correct tact and intraverbal responses. If an incorrect response occurred during training, a prompting procedure was implemented, as described below.

During the training conditions (i.e., tact and intraverbal training) a 0-s prompt delay was used in the first block. Contingent on the first prompted responses, positive consequences were provided by the experimenter in the form of praise (e.g., “Good”). During tact training, the experimenter presented the 0-s prompt delay procedure on the first nine trials in the first 27-trial block. Likewise in the first block of intraverbal training, the experimenter modeled a correct intraverbal (e.g., in the presence of A1′; “The plant for passion fruit is…”, B1′; “cactus”) for the participants to repeat (e.g., B1′; “cactus”), using a 0-s prompt delay on the first six trials of the 18-trial block. Prompts were faded to a 5-s constant time delay in subsequent trials. In the remaining intraverbal training blocks, the experimenter presented the same intraverbals in a fill-in-correct-answer format. For example, given the vocal antecedent, “The plant for passion fruit is...” (A1′), the experimenter gave the participants 5 s to complete the sentence vocally by saying, “cactus” (B1′).

In the subsequent training blocks, participants had 5 s to respond. If the participant responded incorrectly during training, the experimenter would say, "Try again," then repeat the trial and immediately prompt the correct answer (cf. 0-s prompt delay). The participants were then asked to repeat the experimenter, and programmed consequences were delivered (i.e., praise or the correction procedure). If an incorrect or no response was exhibited, the experimenter conveyed a vocal prompt (saying the correct name of the stimulus) according to a constant 5-s prompt delay and then, repeated the same trial one more time.

Procedure

MTS Pre- and Post-Tests

Before all MTS tests were initiated, the following instructions were given by the experimenter:

In this phase, you will use a computer program. Thoroughly read the instruction on your computer screen aloud to ensure your comprehension of its content. Afterwards, in your own words, repeat the instruction aloud. Then, press "I accept" and start.

On the computer screen, the following instruction was presented:

A stimulus will appear in the middle of the screen. You should click on it with the computer mouse. Three other stimuli will appear. Select one of these by clicking with the mouse. No feedback will be given on whether your choices are correct or incorrect. Do your best to get as many correct as possible. Good luck! Click Start to begin the experiment.

All MTS tests consisted of 54 trials; each relation tested 18 times (Table 2). Each trial of simultaneous MTS-tests started with participants clicking on the sample stimulus which revealed three comparisons. For AB/AC relations, A stimuli served as sample stimuli while B and C stimuli served as comparison stimuli. For BA/CA relations, B and C stimuli served as sample stimuli, while A stimuli served as comparison stimuli. For BC/CB relations, B stimuli served as sample stimuli and C stimuli as comparison stimuli in half of the trials, while in the other half, C stimuli served as sample stimuli and B stimuli served as comparisons.

Tact Training

Tact training was initiated by the experimenter who gave the following instructions:

In this phase, I will teach you the names of the images. One by one image will appear in the middle of the screen. I will say the name of the picture and want you to repeat the name you hear. After I ask you to name all of the images, you will get five seconds to name them independently. I will provide feedback and assistance if necessary. Can you repeat the instruction in your own words?

Tact training consisted of 27 trials per block. For a correct tact response to be recorded (e.g., A3A3'; "watermelon" in the presence of A3), the participants had to emit the correct tact, vocally, without prompts within 5 s.

Listener Post-Test

The listener post-test commenced when the experimenter gave the following instructions:

During this phase, three images will be displayed on the screen on each slide. At the same time, an audio file will automatically play that gives the name of one of the images. Point to the picture that best matches the name you heard from the computer. You have five seconds to point to one of the images, and your first response is recorded as your response. I won't give any feedback along the way. Can you repeat the instruction for me?

During the listener post-test, all nine experimental stimuli were presented three times for one test block, comprised of 27 trials. Three comparison stimuli were presented as images on a computer screen per trial. The participants were given 5 s to emit the correct response—pointing with their finger on the correct stimulus. The computer dictated the name of all target stimuli (e.g., A3'; "watermelon"). Moving to the next trial was controlled by the experimenter by counting 5 s, silently, after the presentation of the auditory stimulus. Correct responses were recorded if the participants pointed to the comparison stimulus that corresponded to the auditory stimulus (e.g., A3'A3; in the presence of the auditory conditional stimulus "watermelon," the participant pointed to the image of watermelon––discriminative stimulus; SD) and did not point the two s-delta (SΔ) comparison stimuli. All participants underwent two blocks of the listener post-test. The mastery criterion was met if the participants obtained at least 100% correct responses in at least one of the two listener test blocks. If the participants did not pass listener post-tests, tact training was re-administered until the mastery criterion was met. Then, the two listener post-tests were repeated.

Intraverbal Training

Intraverbal training was initiated by the following instructions presented by the experimenter:

In this phase, I will say a series of phrases that you will repeat. For each sentence you have repeated, I will give you the same sentence as a "fill-in-what's-

missing-task", and you will complete that sentence. You have five seconds to complete the sentences, and your first response will be recorded. I will give you feedback and help you along the way. Eventually, I'll just say, "fill-in-what's-missing," the sentences that you're going to finish. Can you repeat the instruction for me?

All intraverbals trained and tested are presented in Table 1. Intraverbal training consisted of 18 trials per block. Training blocks consisted of a mix of A′B′ and A′C′ baseline relations, where each relation was presented three times. No visual stimuli were presented during this condition. A trial consisted of a vocal SD to which the participants responded vocally, according to the definition of intraverbals (Skinner, 1957, pp. 71–77). Each stimulus was related across stimulus classes to each other (e.g., A1′ to B1,′ and A1′ to C1′). That is, when the experimenter presented a vocal SD (e.g., A1′ for…), the participants were to respond vocally (e.g., B1′).

Intraverbal Post-Tests

Before the intraverbal post-tests (I-BiN post-tests) were initiated, the following instructions were given by the experimenter:

This phase will consist of phrases similar to those before. I will only give "fill-in-what's-missing-tasks" for you to complete. You have five seconds to respond. I won't give feedback along the way. Can you repeat the instruction for me?

This condition was performed similarly to intraverbal training of baseline relations. I-BiN post-tests consisted of 18 trials in which included A′B′/A′C′ intraverbals—each of the six target relations presented three times in a random order. The relations tested were (a) A′B′/A′C′ intraverbals (baseline relations), (b) B′A′/ C′A′ intraverbals (symmetry intraverbal relations), and (c) B′C′/C′B′ intraverbals (equivalence intraverbal relations) across stimulus classes (see Table 1).

Vocal MTS Post-Test

Before the vocal MTS post-test, the following instructions were given by the experimenter:

In this phase, you will use the same computer program as before. Read aloud the instruction on your computer screen and make sure you have understood what it says. This time, you should also say out loud how you solved the tasks and what you chose. Afterward, in your own words, you repeat the instruction to me. Then press "I accept" and start the exercise.

This condition consisted of an additional 18-trial block of a BC/CB MTS post-test where participants were asked to talk aloud during the MTS task. This condition aimed to bring the participants' self-talk to an observable level for measurement.

Sorting Post-Tests

After MTS and intraverbal post-tests were completed, three sorting post-tests were conducted. The three sorting post-tests were carried one after another without breaks in between. A sorting pre-test was omitted to limit exposure to the experimental stimuli. All nine experimental stimuli were presented as laminated pictures in a block. The experimenter shuffled the pictures before every test block and spread the pictures on the table in front of the participants. Participants were asked to, "group the images."

Post-Experimental Interview

Immediately following the sorting post-tests, the participants were asked by the experimenter (see Appendix A) how they solved the MTS tasks and what they said to themselves during the tasks.

Social Validity Survey

After finishing the experimental conditions, participants were asked to complete a social validity survey. The survey consisted of five statements, and participants were instructed to indicate their level of agreement with each of them (see Appendix B). They did so by selecting a response on a 1–6 Likert scale, ranging from "strongly disagree" (1) to "strongly agree" (6).

RESULTS

MTS Pre-Tests

During MTS pre-tests, all participants (P1–P8) scored 50% or below on each relation type, as demonstrated in Figures 3 and 4. Baseline-like AB/AC relations obtained a range of 11–50%, symmetry-like BA/CA relations a range of 16–38%, and equivalence-like BC/CB relations a range of 0−38%.

Tact Training and Listener Post-Test

Trials to criterion varied across participants—range 54–648 trials, as shown in Table 3. The participants required a mean of 226 tact trials to achieve the mastery criterion. All participants met the test criterion on the subsequent listener post-test. Only P3, P4, and P5 did not pass the first listener post-test block. However, they achieved 100% correct responses during the second block.

Intraverbal Training

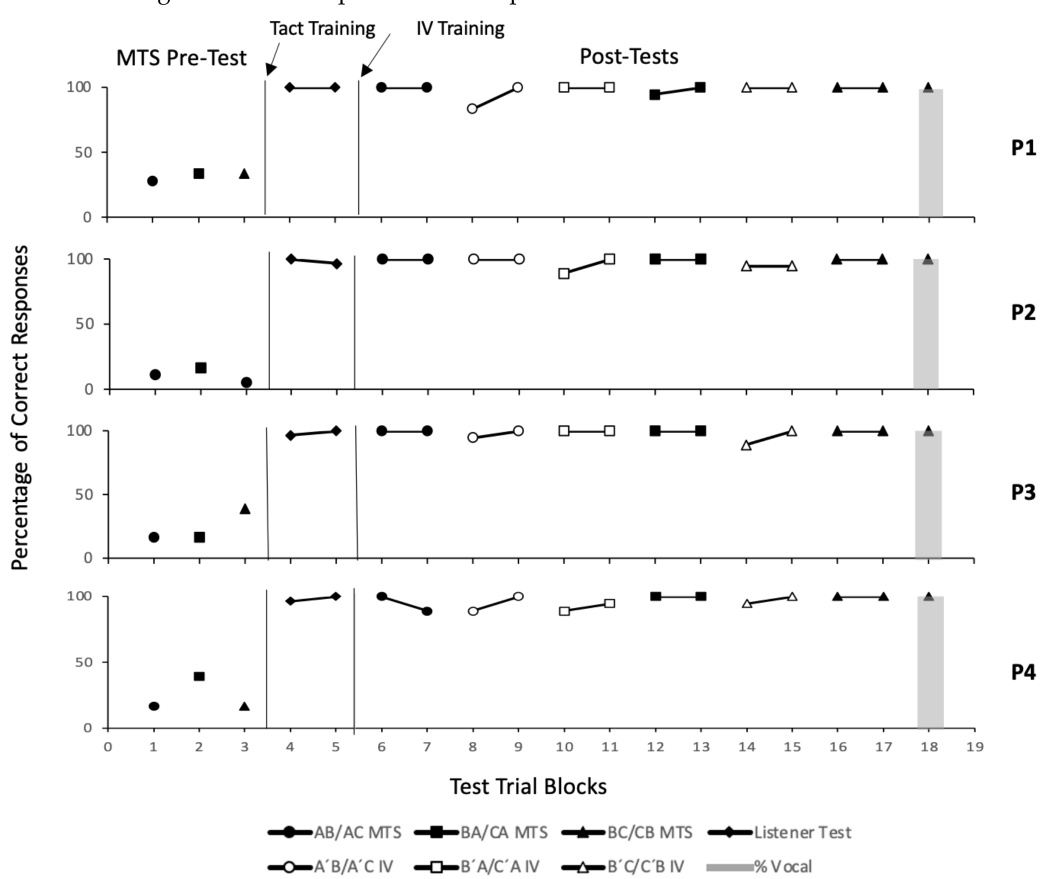
During intraverbal training (A'B'/A'C' relations), the participants reached the criterion of 100% mastery in one block after 54−180 trials (Table 3).

Table 3. Number of Trials to Reach Mastery Criterion for Tact and Intraverbal Training.

P#	Tact Training	Intraverbal Training
P1	297	72
P2	81	72
P3	162	144
P4	189	90
P5	648	126
P6	54	72
P7	162	180
P8	216	54

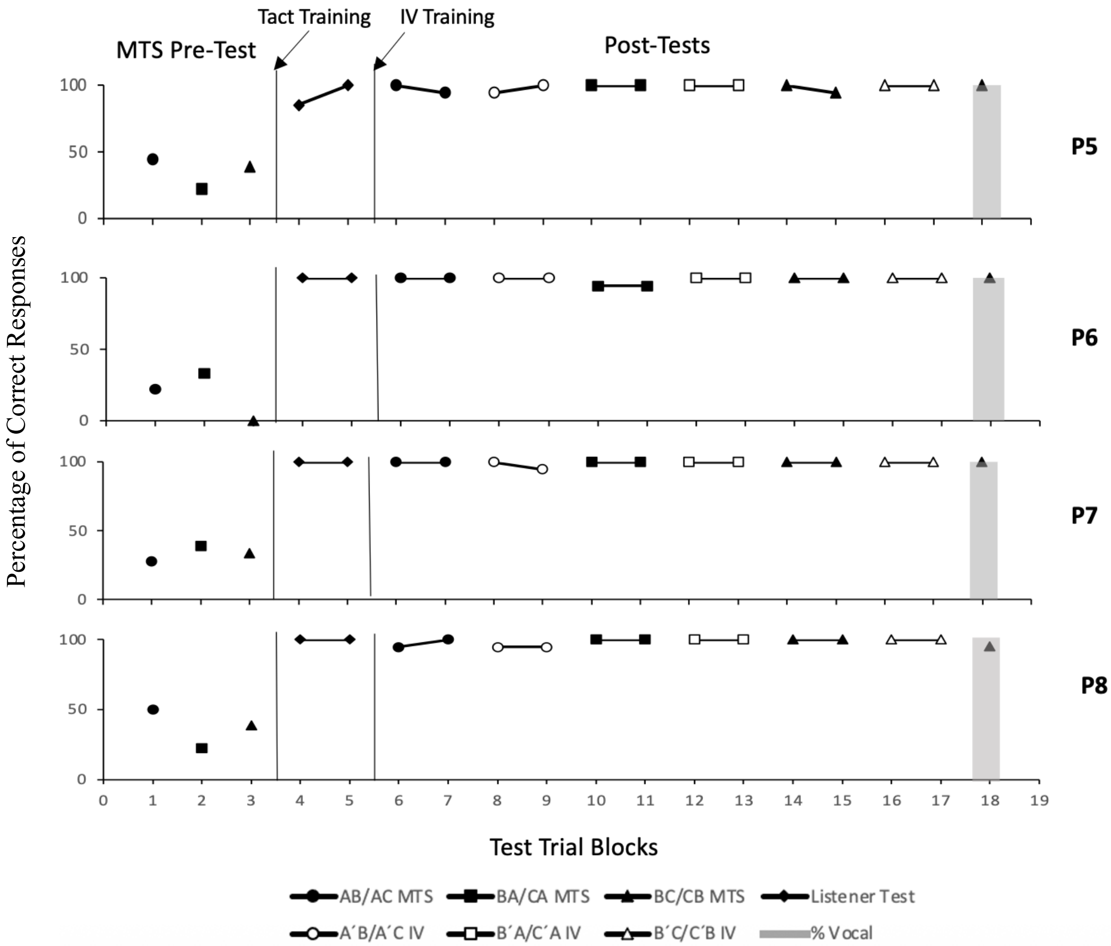
Note. P# = participant number. Tact training consisted of 27-trial blocks and intraverbal training of 18-trial blocks.

Figure 3. The Percentage of Correct Responses for Group 1.



Note. The filled data points indicate selection responses to visual stimuli, either MTS responses or listener responses, while the open datapoints represent intraverbal relations. The y-axis shows the percentage of correct responses and the x-axis test trial blocks. For both groups, MTS pre-tests were completed before tact and intraverbal training. IV is an abbreviation for intraverbal. P = Participant.

Figure 4. The Percentage of Correct Responses for Group 2.



Note. The filled data points indicate selection responses to visual stimuli, either MTS responses or listener responses, while the open datapoints represent intraverbal relations. The y-axis shows the percentage of correct responses and the x-axis test trial blocks. For both groups, MTS pre-tests were completed before tact and intraverbal training. IV is an abbreviation for intraverbal. P = Participant.

MTS Post-Tests

Regardless of different post-test sequence (Group 1 received I-BiN post-tests first while Group 2 received MTS post-tests first), all participants passed the first MTS post-test block (94–100% correct responses on each trial type). In contrast, seven of eight participants met the criterion for the MTS post-tests during the second block, as shown in Figures 3 and 4.

Intraverbal Post-Tests

During the first I-BiN post-test block of symmetry and equivalence relations, three out of four participants in Group 1 (P1–P4) did not pass the test (94% correct responses; 17 out of 18 correct responses; Figure 3). During the second I-BiN post-test block (symmetry and equivalence), Group 1 participants (P1–P4) emitted 100% correct responses, except for P2 (94% correct responses on B′C′/C′B′ equivalence relations) and P4 (94% correct responses on B′A′/C′A′ symmetry relations). However, both met the test criterion.

In contrast, Group 2 participants (P5–P8) responded according to the criterion on all relation types, during both the first and the second block of I-BiN post-tests, ranging from 94–100% correct responses (see Figure 4).

Vocal MTS Post-Test

None of the participants vocalized during the MTS post-tests until explicitly instructed to do so. The vocal MTS post-test revealed that seven out of eight participants vocalized on all 18 trials (100%). P4 vocalized on 17 out of 18 trials. They either tacted the sample or the comparison stimuli. Two participants (P6 and P8) consistently used the established tacts throughout the block. The remaining participants (P1–P5 and P7) used few established tacts and rather, used self-generated tacts. For example, for the stimulus Japanese Knotweed (in Norwegian "parkslirekne"), one of the participants responded "parksliblomst" (Norwegian for flower is blomst).

P1, P3, and P6 were the participants who most accurately used both the established and emergent intraverbals, alternatively, self-generated intraverbals. For example, P3 vocalized, "The animal for cacti is bush dog." The other participants mainly relied on self-generated intraverbals. For instance, P1 said, "Bush dog gets its food from cacti," and P5 stated, "Raffia palm belongs to okapi."

P2, P4, P5, and P6 frequently tacted stimuli belonging to the same experimenter-defined class across multiple trials, others provided more elaborations. For example, P2 listed "Okapi, watermelon, and raffia palm," while P7 offered detailed descriptions, such as "When I see cacti, I think of passion fruit, and passion fruit belongs with bush dog, so I choose bush dog."

Sorting Post-Tests

Regardless of group, the sorting post-tests show that all participants grouped stimuli into experimenter-defined classes at least once (300, 030, 003), as illustrated in Table 4. Two participants from Group 1 (P1 and P3) and two from Group 2 (P6 and P7) grouped the stimuli correctly on all three administrations of the sorting tests. The remaining participants (P2, P4, P5, and P8) grouped all stimuli correctly in two of the three sorting tests. In the second sorting test, P2, P4, and P8 grouped the stimuli according to categories (A1/A2/A3, B1/B2/B3, and C1/C2/C3), rather than according to classes, whereas P5 sorted the stimuli according to categories in the third sorting test.

Post-Experimental Interview

In the post-experimental interview (see questions in Appendix A), all participants reported that during the MTS pre-test they looked for similarities or patterns between the stimuli to match them correctly, such as physical similarities and size (e.g., P2, P3, P4 and P5). Other participants (P1, P6, P7 and P8) reported grouping the stimuli in terms of geographical belongingness, such as "which animal eats which fruit, and what area the fruit, plant or animal resides in."

During the MTS post-tests, all participants reported using intraverbals as a strategy to respond correctly, either they were previously established tacts and/or intraverbals, emergent intraverbals, or self-generated tacts and intraverbals. They also reported that repetition of the acquired "fill-in-the-blank-statements" was the most common strategy used during the subsequent vocal MTS post-test and intraverbal post-tests. In addition, P3 stated that she found it easier to respond to the MTS tests as the stimuli were visual and therefore present, compared with responding during the intraverbal post-tests. Further, P8 reported that she used the established intraverbals to create a pattern of her own, to link stimuli together. During the MTS post-tests, P4 reported that she gradually omitted the established intraverbals and instead used her own statements to remember which stimuli belonged together and used this strategy throughout the post-tests.

Social Validity Survey

The survey consisted of five questions, each of which contributed to a calculated mean score from the participants' answers (see Appendix B). The mean scores for all statements ranged from 4.4 to 4.9 out of a total score of 6. Question (Q)1 received an average score of 4.4, with a range of 2–6. Q2 had an average score of 4.6, range 3–6,

Table 4. The Results of Post-Sorting Tests.

P#	**Sorting Tests**		
	1	2	3
		Group 1	
P1	**300 030 003**	**300 030 003**	**300 030 003**
P2	**300 030 003**	111 111 111	**300 030 003**
P3	**300 030 003**	**300 030 003**	**300 030 003**
P4	**300 030 003**	111 111 111	**300 030 003**
		Group 2	
P5	**300 030 003**	**300 030 003**	111 111 111
P6	**300 030 003**	**300 030 003**	**300 030 003**
P7	**300 030 003**	**300 030 003**	**300 030 003**
P8	**300 030 003**	111 111 111	**300 030 003**

Note. P# = participant number. The numbers show participant-defined categories for each participant. Bold text indicates that the participant-defined categories were sorted in accordance with experimenter-defined categories.

while Q3 achieved a mean score of 4.7, also with a range of 3–6. Finally, Q4 recorded a mean score of 4.9, with a range of 3–6, and Q5 also obtained a mean score of 4.9, within the same range of 3–6.

DISCUSSION

Main Findings

The purpose of the present experiment was to replicate and extend Ma et al. (2016) by employing a OTM structure rather than a LS structure during training. The processes investigated were whether training tacts and baseline intraverbal relations, affected the emergence of I-BiN towards the experimental stimuli and MTS-selection, consistent with stimulus equivalence. Finally, we studied whether the sorting post-tests confirmed the results of the MTS post-tests.

The findings suggest that training tacts (AA′, BB′, CC′) and baseline intraverbal relations (A′B′, A′C′) led to the emergence of bidirectional relations between the relevant intraverbals (I-BiN) that potentially impacted the formation of equivalence classes during MTS post-tests. Thus, the present experiment supports previous studies (e.g., Chastain et al., 2022; Jennings & Miguel, 2017; Ma et al., 2016; Petursdottir et al., 2019; Santos et al., 2015).

In the present study, during I-BiN post-tests, a minor difference was observed in scores between the groups. In Group 1 participants, who were exposed to I-BiN post-tests first, three out of four participants responded below the criterion in the presence of at least one relation type in the first test block of the I-BiN post-tests. In comparison, Group 2, who underwent MTS post-tests first, participants passed the test on both the first and the second I-BiN post-test block. This difference in correct responses could possibly be a result of a sequence effect of first being exposed to MTS post-tests before I-BiN post-tests that familiarized Group 2 participants with the relations between the visual stimuli, that in turn influenced responding during I-BiN post-tests. However, this minor difference in I-BiN test performance between the groups, did not impact the achievement of the criterion on MTS post-tests. Seven out of eight participants (P4, in Group 1, obtained 88.9% correct responses) passed all MTS post-tests in the second test block (94% correct responses or higher), suggesting emergence of responding in accordance with equivalence.

Notably, in the present experiment and likewise in Ma et al. (2016), MTS baseline-like conditional discriminations were never trained. Therefore, one could argue that the emergence of symmetry-like and equivalence-like relations in the MTS test, in the present experiment, does not accurately describe these types of relations according to a formal definition of stimulus equivalence. However, these MTS relations correspond with the intraverbal relations trained (baseline relations) and tested (symmetry and equivalence relations). Regardless of high yields during MTS post-tests, these results could also be described as a product of other types of responses or the sequence of training conditions (e.g., conditioned seeing and presenting tact training before intraverbal training; Petursdottir et al., 2019), rather than the emission of tacts and intraverbals during the MTS tasks (see Training Sequence below).

The conclusion that I-BiN was likely responsible for high yields during MTS post-tests were supported by the vocal responses during the vocal MTS post-test and post-experimental interview, where using trained and self-generated tacts and intraverbals were evident.

Training and Testing Structure

Most previous studies on I-BiN to form equivalence classes, used a LS structure (e.g., Jennings & Miguel, 2017) or MTO (Chastain et al., 2022). This study may be the first that demonstrates similar findings with a OTM structure. Training intraverbal baseline relations using a OTM structure was effective for the emergence of novel intraverbals, such as symmetry and equivalence relations, as well as novel selection responses during MTS post-tests of symmetry-like and equivalence-like relations––consistent with stimulus equivalence.

Zaring-Hinkle et al. (2016) compared LS and OTM to facilitate the emergence of novel intraverbals. The results suggest that novel intraverbal relations are more likely to emerge when baseline intraverbal relations are trained with a OTM structure, rather than LS. However, during baseline intraverbal training, more trials to criterion were used during the OTM condition, compared with the LS condition. Although, more trials to criterion, OTM resulted in more novel intraverbals. During the MTS tests, the criterion was only met for three out of eight participants using LS, while all participants passed the test during the OTM condition. The results are consistent with previous research that LS is less effective in establishing equivalence classes than OTM (See Arntzen, 2012, for an overview) and thus, sheds light on the importance of the results of the present experiment. Since the present study did not directly compare training structures, it is not possible to conclude whether OTM was more beneficial, than LS or MTO, in training intraverbals to facilitate I-BiN towards the experimental stimuli and MTS performances.

Training Sequence

In the present experiment, tact training was implemented before intraverbal training. Petursdottir et al. (2019) examined how two different training sequences affected the emergence of conditional discriminations in college students, which shed light on the findings in the present study. In Petursdottir et al., one group (the TI-group) learned tacts (AA', BB') before intraverbals (A'B'), while the other group (the IT-group) learned intraverbals first. After training, both groups underwent a MTS test (AB and BA relations). The TI-group had significantly shorter reaction times than the IT-group, while the number of correct responses did not differ significantly during MTS tests. Subsequent tests showed that MTS accuracy correlated with intraverbal retention in the IT-group but not in the TI-group. The authors suggested an impact of conditioned seeing, because the TI participants obtained immediate stimulus control during the MTS tests which did not necessitate verbal mediation. This indicates that TI participants, potentially, relied less on intraverbal retention, while IT participants, probably, used verbal mediation skills (established intraverbals) to form equivalence classes. Thus, the results of the present experiment, could be explained by the conditioned seeing because relations between visual stimuli may have been established directly during intraverbal training (i.e., the participants were "seeing" the images while hearing their names). The conditioned seeing account suggests that the participants who were exposed to tact training before intraverbal training are not necessarily dependent on retaining the intraverbal relations between the tacts. The interpretation of Petursdottir et al. suggests that the recency of the intraverbal training, rather than the training sequence (training tacts before intraverbals) resulted in shorter reaction times and improved performances during MTS tests. In contrast, tacts were unaffected by recency for the TI-participants and in the present study, but rather the MTS post-test performances were possibly influenced by the recency of intraverbal training and conditioned seeing. The result from Petursdottir et al. is supported in a recent study by Chastain et al. (2022; Experiment 3). Thus, the effects of training sequences and possibly conditioned seeing warrant further investigations.

Correspondence Between MTS Post-Tests and Sorting Post Tests

All participants responded according with experimenter-defined classes on the MTS post-tests. Across the three sorting post-tests, four out of eight participants sorted the stimuli into experimenter-defined classes. The fact that participants who grouped stimuli correctly on all three tests came from different groups indicates that the order of the MTS post-tests did not influence sorting accuracy. The certainty that not all participants grouped the stimuli correctly on all three tests, may suggest a weaker degree of stimulus control. Despite variation in the results of the sorting post-tests compared to the MTS post-tests, the number of correct responses during the MTS post-tests confirmed the formation of equivalence classes. The results of the sorting post-tests moderately support the correspondence of performance on MTS and sorting tests (e.g., Arntzen et al., 2021; Arntzen et al., 2017; Arntzen et al., 2015). Thus, it is important to emphasize that the participants who did not sort the stimuli according to the experimenter-defined classes (P2, P4, P5, and P8) in one of the three sorting post-tests, sorted the stimuli according to categories and not classes.

Limitations and Future Studies

There are several limitations of the present experiment. The first limitation is that following baseline intraverbal training (A'B'/A'C'), the participants were exposed to a MTS post-test including baseline-like relations (AB/AC), and then the intraverbal post-test of baseline relations. The intraverbal training could have influenced the MTS post-test and the exposure to these test trials could have influenced the results of the intraverbal post-tests. Future experiments should control for this effect of order.

A second limitation is the possibility of a contextual cue as part of the intraverbal trials (e.g., Petursdottir et al., 2019). As in previous research, the present experiment presented a contextual stimulus, “for,” in the intraverbal trials (e.g., “The plant for passion fruit is…”). The presence of this contextual cue could, potentially, facilitate contextual control in the following test trials during MTS post-tests. Future studies should experimentally manipulate the control by such contextual cues. Alternatively, it is also likely that the stimulus control established during the intraverbal training across stimulus classes could have been facilitated by autoclitic frames (e.g., Jennings et al., 2023; Skinner, 1957), such as, “The plant for…is….” Skinner (1957) described that when autoclitic frames are established, they are readily used to encounter novel stimuli. Jennings et al. (2023) suggested if the participants had a history of responding to such autoclitic frames, it could be likely that these frames impacted acquisition of the intraverbal statements.

A third limitation of the present study is that some of the stimuli were already familiar to the participants (e.g., watermelon, cactus, and passion fruit). The results showed that some of the participants provided correct tact responses to relevant stimuli in the first tact training block. The fact that several tacts were already established might have influenced the formation of intraverbals—linking the various stimulus relations vocally together and contributing to the successful grouping of stimuli into experimenter-defined classes. Future experiments should include a pre-tact probe before training conditions commence to ensure the novelty of the experimental stimuli (e.g., Jennings & Miguel, 2017).

Since previous experiments have primarily focused on children and university students as participants (e.g., Carp & Petursdottir, 2015; Jennings & Miguel, 2017; Santos et al., 2015), future studies should include a variety of participants. Further, investigations about the use of self-generated intraverbals are needed. Some of the participants, in previous studies, did not correctly produce tact or intraverbal responses during the vocal MTS post-test (e.g., AC and CA relations), instead relying on self-generated intraverbals. The use of self-generated intraverbals suggests that not all used the trained tacts or intraverbals during the vocal MTS post-tests. Finally, future research should investigate the long-term effects of training that may facilitate mediation responses (e.g., I-BiN) and assess whether the equivalence classes remain intact during follow-up tests.

Summary

The present study extends previous experiments (e.g., Carp & Petursdottir, 2015; Chastain et al., 2022; Jennings & Miguel, 2017; Ma et al., 2016; Petursdottir et al., 2019; Santos et al., 2015) by demonstrating the formation of equivalence classes, as a potential product of the establishment of tacts and intraverbal baseline relations using a OTM structure. In this experiment, as well as in previous studies, the training conditions may have led to the emergence of I-BiN towards the experimental stimuli and responding in accordance with stimulus equivalence during the MTS post-tests. Sorting post-tests where the participants grouped stimuli into experimenter-defined classes, moderately supported the results from the MTS post-tests. Additionally, the use of verbal mediation was supported by the results of the vocal MTS post-test and post-experimental interview from most of the participants.

REFERENCES

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *Psychological Record*, *54*(2), 275–291. https://doi.org/10.1007/BF03395474

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

Arntzen, E., Dechsling, A., & Fields, L. (2021). Expansion of arbitrary stimulus classes and function-transfer measured by sorting performances. *Journal of the Experimental Analysis of Behavior*, *115*(1), 326–339. https://doi.org/10.1002/jeab.665

Arntzen, 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*(1), 81–96. https://doi.org/10.1007/s40732-016-0209-9

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis*, *12*(2), 483–503. https://doi.org/10.1080/15021149.2011.11434397

Arntzen, E., Norbom, A., & Fields, L. (2015). Sorting: An alternative measure of class formation? The *Psychological Record*, *65*(4), 615–625. https://doi.org/10.1007/s40732-015-0132-5

Carp, C. L., & Petursdottir, A. I. (2015). Intraverbal naming and equivalence class formation in children. *Journal of the Experimental Analysis of Behavior*, *104*(3), 223–240. https://doi.org/10.1002/jeab.183

Carr, D., Wilkinson, K. M., Blackman, D., & McIlvane, W. J. (2000). Equivalence classes in individuals with minimal verbal repertoires. *Journal of Experimental Analysis of Behavior*, *74*(1), 101–114. https://doi.org/10.1901/jeab.2000.74-101

Chastain, A. N., Luoma, S. M., Love, S. E., & Miguel, C. F. (2022). The role of irrelevant, class-consistent, and class-inconsistent intraverbal training on the establishment of equivalence classes. *The Psychological Record*, *72*, 383–405. https://doi.org/10.1007/s40732-021-00492-9

Clough, C. W., Meyer, C. S., & Miguel, C. F. (2016). The effects of blocking and joint control training on sequencing visual stimuli. *The Analysis of Verbal Behavior*, *32*(2), 242–264. https://doi.org/10.1007/s40616-016-0067-1

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. L. M. Perone (Ed.), Handbook of Research Methods in Human Operant Behavior (pp. 229–262). Springer.

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), 185–241. https://doi.org/10.1901/jeab.1996.65-185

Huntington, R. N., Badgett, N. M., Rosenberg, N. E., Greeny, K., Bravo, A., Bristol, R. M., Byun, Y. H., & Park, M. S. (2023). Social validity in behavioral research: A selective review. *Perspectives on Behavior Science*, *46*(1), 201–215. https://doi.org/10.1007/s40614-022-00364-9

Jennings, A. M., & Miguel, C. F. (2017). Training intraverbal bidirectional naming to establish generalized equivalence class performances. *Journal of the Experimental Analysis of Behavior*, *108*(2), 269–289. https://doi.org/10.1002/jeab.277

Jennings, A. M., Vladescu, J. C., Miguel, C. F., Reeve, K. F., & Sidener, T. M. (2023). A translational evaluation of component skills for the establishment of multiply controlled intraverbals. *Journal of the Experimental Analysis of Behavior*, *119*(3), 513–528. https://doi.org/10.1002/jeab.837

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*(3), 409–426. https://doi.org/10.1002/jeab.203

Malkin, A., Jacobs, E. A., & Kretschmer, A. (2024). Behavior analysts' relationship to relating relations: A survey on perceptions, acceptability, knowledge, and capacity for derived stimulus relations research and practice. *Behavior Analysis in Practice*, *17*(1), 257–269. https://doi.org/10.1007/s40617-023-00834-9

Miguel, C. F. (2016). Common and intraverbal bidirectional naming. *The Analysis of Verbal Behavior*, *32*(2), 125–138. https://doi.org/10.1007/s40616-016-0066-2

Miguel, C. F. (2018). Problem-solving, bidirectional naming, and the development of verbal repertoires. *Behavior Analysis: Research and Practice*, *18*(1), 340–353. https://doi.org/10.1037/bar0000110

Miller, A. C., Cox, R. E., Swensson, R. M., Oliveira, J. S. C. D., & Petursdottir, A. I. (2021). Effects of blocking echoic responses on tact emergence following stimulus pairing. *European Journal of Behavior Analysis*, *22*(2), 213–225. https://doi.org/10.1080/15021149.2021.1896070

Pérez-González, L. A., & Oltra, J. (2023). Learning symmetrical relations facilitates emergence of intraverbals after reading a text: Effects on reading comprehension. *The Psychological Record*, *73*, 13–31. https://doi.org/10.1007/s40732-022-00532-y

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*(2), 332–348. https://doi.org/10.1002/jeab.136

Petursdottir, A. I., Cox, R. E., McKeon, C. A., & Mellor, J. R. (2019). Baseline training sequence affects speed of emergent conditional discriminations. *Journal of the Experimental Analysis of Behavior*, *112*(1), 60–73. https://doi.org/10.1002/jeab.539

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*(2), 162–182. https://doi.org/10.1007/s40616-015-0040-4

Schlinger, H. D., Jr., & Blakely, E. (2024). A mediational theory of equivalence relations and transformation of function. *Journal of Experimental Analysis of Behavior*, *122*(2), 207–223. https://doi.org/10.1002/jeab.4204

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*, 127–146. https://doi.org/10.1901/jeab.2000.74-127

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, *37*(1), 5–22. https://doi.org/10.1901/jeab.1982.37-5

Skinner, B. F. (1957). Verbal behavior. Appleton-Century-Crofts.

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

Vie, A., & Arntzen, E. (2019). Role of distractors in delayed matching-to-sample arrangements in tests for emergent relations. *International Journal of Psychology and Psychological Therapy*, *1*, 71–88.

Zaring-Hinkle, B., Carp, C. L., & Lepper, T. L. (2016). An evaluation of two stimulus equivalence training sequences on the emergence of novel intraverbals. *The Analysis of Verbal Behavior*, *32*(2), 171–193. https://doi.org/10.1007/s40616-016-0072-4

Appendix A

Post-Experimental Interview Questions

#	Questions
1	Can you explain how you solved the tasks on the computer program when we started?
2	How did you solve these tasks after learning the "fill in the blanks" sentences? Please describe the strategies you used.
3	Did you use any strategies to learn the "fill in the blanks" sentences? If so, please describe them.
4	What kind of strategies did you use? Did you use the sentences throughout the experiment? If any, please explain the strategies you used.

Appendix B

Social-Validity Survey

#	Statements	Mean response (range)
1	"I really liked this way of learning factual knowledge, and I am willing to use this method of learning knowledge again."	4.375 (2–6)
2	"I would reccomend this way of learning factual knowledge to others."	4.625 (3–6)
3	"This would be an acceptable training method for teaching young adults to acquire knowledge."	4.687 (3–6)
4	"This training method was not too intensive or intrusive."	4.875 (3–6)
5	"I found the procedures and the different phases to be appropriate and effective."	4.875 (3–6)

Note. The scores range from 1 (strongly disagree) to 6 (strongly agree).