Revista Mexicana de Análisis de la Conducta
ISSN: 0185-4534
ISSN: 2007-0802
editor@rmac-mx.org
Sociedad Mexicana de Análisis de la Conducta
México

Ayres-Pereira, Vanessa; Arntzen, Erik
Emergence of large equivalence classes as a function of training structures
Revista Mexicana de Análisis de la Conducta, vol. 45, núm. 1, 2019, pp. 20-47
Sociedad Mexicana de Análisis de la Conducta
Distrito Federal, México

DOI: https://doi.org/10.5514/rmac.v45.i1.70864

Revista Mexicana de Análisis de la Conducta • Mexican Journal of Behavior Analysis
2019 | Núm. 1 • junio | No. 1 • June | Vol. 45, 20 - 47
http://dx.doi.org/10.5514/rmac.v45.i1.70864

EMERGENCE OF LARGE EQUIVALENCE CLASSES AS A FUNCTION OF TRAINING STRUCTURES

EMERGENCIA DE GRANDES CLASES DE EQUIVALENCIA COMO FUNCIÓN DE ESTRUCTURAS DE ENTRENAMIENTO

Vanessa Ayres-Pereira and Erik Arntzen
Oslo Metropolitan University, Norway

Abstract

This experiment compared the outcomes of two training structures on the emergence of three 7-member equivalence classes. Seventeen adults were exposed to the Many-to-One (MTO) and another 17 to the One-to-Many (OTM) training structure. The MTO group trained the baseline relations BA, CA, DA, EA, FA,

Vanessa Ayres-Pereira, Erik Arntzen, Oslo Metropolitan University, Department of Behavioral Science.

This research was funded by Oslo Metropolitan University. The Authors declare that there is no conflict of interest. Informed consent was obtained from all individual participants. All procedures were in accordance with the ethical standards and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The Brazilian National Council on Human Research (CONEP; Protocol CAAE 59735416.2.0000.5504) approved the procedures and provided special authorization so the data collection, although coordinated by a foreign university, could be implemented in Brazil.

The authors express sincere gratitude to Dr. Deisy G. de Souza and Dr. Julio C. C. de Rose for granting the facilities at the Federal University of São Carlos, Brazil, where the study was conducted.

E-mail: vanessa.ayresp@gmail.com or erik.arntzen@equivalence.net

and GA, and the OTM group trained AB, AC, AD, AE, AF, and AG. After mastering the baseline, a test evaluated the maintenance of baseline and the emergence of symmetry and equivalence relations, under a simultaneous protocol. As a result, fifteen out of seventeen (88%) participants in both groups demonstrated stimulus equivalence. There was no significant difference between groups in the average number of training trials required to learn the baseline relations, nor in accuracy in emergent test trials. The MTO group was characterized by faster response speed in baseline training and test trials. Every participant who failed in the MTO group had persistent errors before four, five, or six out of the 18 sample stimuli during the training, while participants who failed in the OTM group had varied baseline acquisition patterns.

Keywords: Training structures, Many-to-One, One-to-Many, class size, stimulus equivalence, human participants

Resumen

Este experimento comparó los resultados de dos estructuras de entrenamiento en el surgimiento de tres clases de equivalencia de 7 miembros. Diecisiete adultos fueron expuestos a la estructura de entrenamiento Many-to-One (MTO) y otros 17 a la estructura de entrenamiento One-to-Many (OTM). El grupo de MTO entrenó las relaciones de línea base BA, CA, DA, EA, FA y GA, y el grupo de OTM entrenó AB, AC, AD, AE, AF y AG. Después del entrenamiento se evaluó el mantenimiento de la línea de base y la aparición de relaciones de simetría y equivalencia, bajo un protocolo simultáneo. Como resultado, quince de los diecisiete (88%) participantes en ambos grupos demostraron equivalencia de estímulo. No hubo diferencias significativas entre los grupos en el número promedio de ensayos requeridos para aprender las relaciones de línea base, ni en la precisión en los ensayos de pruebas emergentes. El grupo MTO se caracterizó por una velocidad de respuesta más rápida en el entrenamiento y en los ensayos de prueba. Todos los participantes que fallaron en el grupo MTO tuvieron errores persistentes antes de cuatro, cinco o seis de los 18 estímulos de muestra durante la capacitación, mientras que los participantes que fallaron en el grupo OTM tuvieron diferentes patrones de adquisición de la línea base.

Palabras clave: estructuras de entrenamiento, Many-to-One, One-to-Many, tamaño de la clase, equivalencia de estímulos, participantes humanos

Experiments on stimulus equivalence consist of training sets of arbitrary conditional discriminations interrelated by a common stimulus (e.g., B in AB and BC training) and testing for the emergence of several novel conditional discriminations. When consistent relations emerge holding the properties of symmetry (e.g., BA and CB), transitivity (e.g., AC), combined symmetry and transitivity (also called equivalence; e.g., CA), and reflexivity (e.g., AA, BB, CC), they are said to meet the formal criteria for equivalence class formation (Sidman & Tailby, 1982).

The above-mentioned "common stimulus" was termed a node by Fields, Verhave, and Fath (1984). Variations in the position of the node—as sample and/or comparison—over the training of baseline conditional discriminations was designated the training structure, a parameter of great importance in equivalence-oriented procedures due to its impact on the yields of class formation. According to K. J. Saunders, Saunders, Williams, and Spradlin (1993), there are three basic types of training structures: Sample-as-Node, or One-to-Many (OTM; e.g., AB and AC); Comparison-as-Node, or Many-to-One (MTO; e.g., AC and BC); and Linear Series (LS; e.g., AB and BC). Several experiments have consistently demonstrated that the MTO and OTM are more effective than the LS, but presented great variability following the first two (e.g., Arntzen & Holth, 1997; Fields, Hobbie-Reeve, Adams, & Reeve, 1999).

The variability in test outcomes as a function of training structures has been given a number of interpretations (e.g., Fields & Moss, 2007; Zentall, Wasserman, & Urcuioli, 2014). The most prominent and well-known account is the Discriminative Analysis (DiAn; R. R. Saunders & Green, 1999). The main assumption of DiAn is that the establishment of simple discriminations underlies the training of conditional discriminations and is requisite for positive results in emergent relations. Consider, for example, the training of four simultaneous conditional discriminations for the emergence of two 3-member classes (Class 1: A1, B1, C1; Class 2: A2, B2, C2). In the OTM arrangement, the concurrent training of AB and AC (i.e., A1/B1B2, A1/C1C2, A2/B1B2, and A2/C1C2) would require simultaneous discriminations between the comparisons B (B1 ≠ B2) and C (C1 ≠ C2), between the samples and comparisons A-B and A-C (A1 ≠ B1, A1 ≠ B2, A1 ≠ B1, A1 ≠ B2, and A1 ≠ C1, A1 ≠ C2, A2 ≠ C1, A2 ≠ C2), and successive discriminations between the samples A (A1 ≠ A2). Therefore, the training of AB and AC would not require the B-C discriminations, necessary to respond in accordance with both, symmetry (here, B1/A1A2, C1/A1A2, B2/A1A2, C2/A1A2) and equivalence (B1/C1C2, C1/B1B2, B2/C1C2, C2/B1B2). Similarly, the LS (e.g., training AB and BC; A1/

B1B2, A2/B1B2, B1/C1C2, B2/C1C2) would not establish simple discriminations between A and C.

In contrast, in the MTO arrangement, the concurrent training of BA and CA (B1/A1A2, C1/A1A2, B2/A1A2, and C2/A1A2) would require simultaneous discriminations between the comparisons A, between the samples and comparisons B-A and C-A, and successive discriminations between the samples B, C, and B-C. Therefore, the MTO would provide the training of all simple discriminations required for positive results in the test and should be the most effective training structure. Another prediction of the DiAn is that the MTO would produce more errors over training than the other structures for demanding an increased number of successive discriminations between the various samples. Successive discriminations tend to be more difficult than simultaneous ones (Brady & Saunders, 1991).

The DiAn conclusions were based on a set of empirical observations (e.g., R. R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986), however, the evidence was recognized as limited by the authors, who referred to the account as a "hypothesis" and recommended additional experimentation. R. R. Saunders and Green (1999) noted that investigations involving increased class sizes would be more appropriate to support the DiAn, because experiments targeting the emergence of small classes (e.g., 3-member) could fail to demonstrate the MTO superiority due to the requirement of a few simple discriminations.

A set of experiments with increased class sizes confirmed the MTO superiority on the emergence of two equivalence classes with five (Arntzen & Nikolaisen, 2011; Fields et al., 1999; R. R. Saunders, Drake, & Spradlin, 1999; R. R. Saunders et al., 1988; Smeets & Barnes-Holmes, 2005; Spradlin & Saunders, 1986) and seven members (Fields et al., 1999). However, a body of literature has emerged offering contradictory findings on the emergence of three equivalence classes with four, five, and six members, when the OTM produced similar or slightly higher yields of class formation than the MTO (Arntzen, Grondahl, & Eilifsen, 2010; Arntzen & Hansen, 2011; Arntzen & Vaidya, 2008; Sadeghi & Arntzen, 2018).

The findings have also been inconsistent regarding the number of errors over training. Some experiments have reported that the MTO generated more errors and demanded more trials to criterion than the OTM (e.g., Arntzen et al., 2010; Arntzen & Hansen, 2011; Arntzen & Nikolaisen, 2011), while others reported no significant differences (e.g., Hove, 2003; R. R. Saunders, Chaney, & Marquis, 2005; Smeets & Barnes-Holmes, 2005).

To our knowledge, much of the research on training structures has focused on the emergence of only three 3-member classes (e.g., Fiorentini et al., 2013; Hove, 2003; Plazas & Villamil, 2016) and no experiment has verified the predictions of the DiAn on the emergence of three classes with seven members. Experiments on the emergence of three classes are seen as relevant because investigations on the emergence of two classes are often associated with the use of two-choice procedures, a confounder for failure in the test (Carrigan & Sidman, 1992). Therefore, the present study set out to expand this literature by assessing the effects of the MTO and OTM training structures on the emergence of three 7-member equivalence classes. Our population (normal adults) and procedural parameters are comparable to those used by Fields et al. (1999) when investigating the emergence of two 7-member classes. In Fields et al. (1999), the MTO was more effective than the OTM and only a small number of participants (less than 50% in total) formed classes.

Method

Participants

Participants were thirty-four typically functioning adults. They were divided in two groups, 17 were randomly assigned to the MTO group (seven males and 10 females; with ages ranging from 18 to 32 years old, $M = 25.6$, $SD = 4.2$), and the other 17 were assigned to the OTM group (eight males and nine females; with ages ranging from 19 to 32 years old, $M = 24.7$, $SD = 4.3$). Six other adults participated in the experiment, but did not finish the procedure for varying reasons. Their data were excluded from the analysis. Of the six, five were exposed to the MTO training structure (one asked to quit the experiment, and four did not finish the task within a four hours session and declared to be unable to attend another meeting). Another participant was exposed to the OTM training structure but complained about fatigue and was dismissed.

Participants were recruited through personal contacts and advertisements on Universities' social media pages. They declared to have no knowledge of stimulus equivalence. Every participant provided informed consent before starting the experiment. They were allowed to quit at any time, without negative consequences. After finishing the session, participants were fully debriefed (i.e., they were informed about the experiment's purpose, had access to their own data, and received an introductory article about stimulus equivalence).

Equipment and Setting

The experimental sessions were conducted in one of two rooms, both of 1.5 x 3 meters, sound proof, and furnished with a table and a chair. One of the rooms had a window covered by blackout curtains and the other had no window. The experimental task was presented on an HP EliteBook 8760w computer running Windows 10 and a 17-in monitor. Custom-made software presented the stimuli and consequences, and recorded the data (i.e., stimuli and consequence presented per trial, stimulus selected by response, and reaction time to comparison). Responses were made by clicking on computer mouse.

The experimenter (the first author) followed each participant individually to the room. Firstly, she presented the consent form and asked the participant to read and, if s/he agreed, sign the document. Secondly, the experimenter presented the stimuli (see Figure 1) individually printed on cards and asked the participant to "Arrange the stimuli in groups as you feel like" (Pre-Class Formation Sorting Test). This test assessed whether the participant would categorize the stimuli according to the experimenter-defined equivalence classes prior to training (Arntzen, Norbom, & Fields, 2015). The data of participants who did so would have been discarded from the analysis, however none of them did so.

Next, the experimenter asked the participant to sit facing the computer and said, "There are detailed instructions for your task on the computer screen. Read the instruction carefully. Later, a box will pop up on the screen to advise you when it is time for a break". Then, the experimenter left the room. The written instructions presented on the screen were in Portuguese, but in equivalent translation they were as follows:

> "Once you start, a figure will appear in the middle of the screen. Click on this by using the computer mouse. Three other figures will then appear. Choose one of these using the computer mouse. If you choose one of the figures we have defined as correct, words like 'Very Good', 'Excellent', and so on will appear on the screen. If you press an incorrect figure, the word 'Wrong' will appear on the screen. During some stages of the experiment, the computer will not tell you if your choices are correct or incorrect. However, based on what you have learned, you can get all the tasks correct. Please, do your best to get everything right. Good luck!"

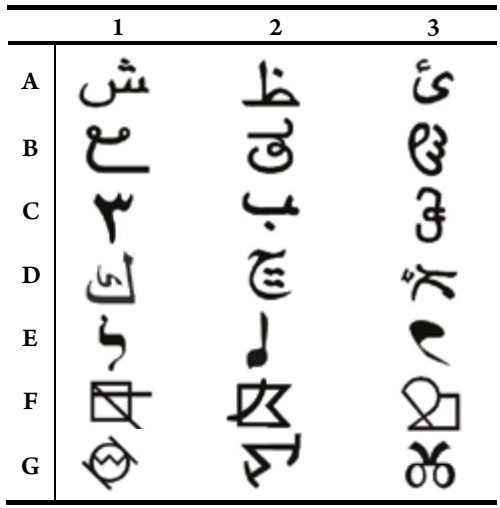

Figure 1. Each stimulus was labeled with a number and a letter, and was expected to function as member of an equivalence class (Class 1, Class 2, or Class 3) at the end of the procedure.

The session lasted a maximum of 4 hr, and had breaks of approximately 10 min after every 50 min. If the participant did not finish the task after four hours, the experimenter invited her/him to return to the experimental session no more than two days later in order to finish it.

Stimuli

The stimuli were Arabic and Hebrew letters, and abstract shapes (see Figure 1). These types of stimuli often are used in studies of stimulus equivalence and were expected to be meaningless for the participants (Portuguese speakers). Some letters were rotated and modified, to make them appear less similar to other stimuli potentially previously known. For analytic purposes, each of the 21 stimuli was labeled with a Roman letter (A, B, C, D, E, F, or G) and a number (1, 2, or 3). The procedure aimed to produce the emergence of three equivalence classes, with seven members each (Class 1: A1, B1, C1, D1, E1, F1, G1; Class 2: A2, B2, C2, D2, E2, F2, G2; and Class 3: A3, B3, C3, D3, E3, F3, G3).

Procedure

One group of participants was exposed to the OTM training structure, and the other to the MTO training structure. For both groups, the procedure included two steps: training of baseline relations followed by testing of baseline, symmetry, and equivalence relations (see Table 1). All baseline relations were trained, before

Table 1. *Sequence of experimental phases presented to the MTO and OTM groups, and relations trained or tested per experimental phase, probability of consequences, minimum number of trials, and mastery criterion*

Phases	MTO Group	OTM Group	Probability of consequences (%)	Number of trials per block	Learning Criterion (%)
Training of Baseline					
1.	BA, CA, DA, EA, FA, GA	AB, AC, AD, AE, AF, AG	100	90	95
2.	BA, CA, DA, EA, FA, GA	AB, AC, AD, AE, AF, AG	50	90	95
3.	BA, CA, DA, EA, FA, GA	AB, AC, AD, AE, AF, AG	0	90	95
Test of Baseline and Emergent Relations					
4.	*(BSL)* BA, CA, DA, EA, FA, GA	*(BSL)* AB, AC, AD, AE, AF, AG	0	378	95
	(SYM) AB, AC, AD, AE, AF, AG	*(SYM)* BA, CA, DA, EA, FA, GA			
	(EQ) BC, BD, BE, BF, BG, CB, CD, CE, CF, CG, DB, DC, DE, DF, DG, EB, EC, ED, EF, EG, FB, FC, FD, FE, FG, GB, GC, GD, GE, GF	*(EQ)* BC, BD, BE, BF, BG, CB, CD, CE, CF, CG, DB, DC, DE, DF, DG, EB, EC, ED, EF, EG, FB, FC, FD, FE, FG, GB, GC, GD, GE, GF			

Note. The relations were presented concurrently (in random order) within every phase. BSL=baseline, SYM=symmetry, and EQ=Equivalence

all derived relations were tested, using a simultaneous protocol (see Iman, 2006). Baseline training for the OTM group involved AB, AC, AD, AE, AF, and AG relations; whereas baseline training for the MTO group involved the BA, CA, DA, EA, FA, and GA relations. The "A" stimulus thus functioned as node in both structures.

The present experiment termed the "A" stimulus as node in both training structures (as in R. R. Saunders et al., 1999; R. R. Saunders et al., 1988). Therefore, in the present experiment, the unique difference between groups was the training structures defining features: the nodal stimuli were presented as samples for the OTM and as comparisons for the MTO groups.

The whole procedure employed exclusively trials of simultaneous conditional discrimination with an observing response. That is, every trial started with a single stimulus presented in the center of the screen (sample). A mouse click on the sample (observing response) presented three other stimuli at the corners (comparisons). The comparisons were displayed along with the sample during the trial. The training and testing phases are detailed below.

Baseline training. In Phase 1, participants in both groups were trained on the baseline relations with 1.0 probability of programmed consequences for responses. The programmed consequence was a written word presented in the center of the screen for 500ms. The word "Excellent", "Great", "Very good", or "Right" was presented for correct response and "Wrong" was presented for incorrect responses. The consequences were followed by a 500ms blank screen (intertrial interval) and the beginning of a new trial.

Phase 1 employed blocks of 90 trials, with 18 trial types repeated 5 times in random order and varying the correct comparison position on the screen. In the OTM group, the trial types were: A1/B1B2B3, A2/B1B2B3, A3/B1B2B3 (AB); A1/C1C2C3, A2/C1C2C3, A3/C1C2C3 (AC); A1/D1D2D3, A2/D1D2D3, A3/D1D2D3 (AD), A1/E1E2E3, A2/E1E2E3, A3/E1E2E3 (AE); A1/F1F2F3, A2/F1F2F3, A3/F1F2F3 (AF), A1/G1G2G3, A2/G1G2G3, A3/G1G2G3 (AG). The underlined stimuli were the correct comparison. In the MTO group, the trial types were: B1/A1A2A3, B2/A1A2A3, B3/A1A2A3 (BA); CA (C1/A1A2A3, C2/A1A2A3, C3/A1A2A3 (CA); D1/A1A2A3, D2/A1A2A3, D3/A1A2A3 (DA); E1/A1A2A3, E2/A1A2A3, E3/A1A2A3 (EA);), F1/A1A2A3, F2/A1A2A3, F3/A1A2A3 (FA); G1/A1A2A3, G2/A1A2A3, G3/A1A2A3 (GA). The 90-trial block was repeated until the participant performed at least 95% correct responses in one block of trials (86/90, mastery criterion).

After mastering the baseline, the participant was exposed to two phases of gradual decreases in the probability of consequences: Phase 2 (with 50% probability of differential consequences) and Phase 3 (without any programmed consequence). These phases were conducted as preparation for the test, conducted under ex-

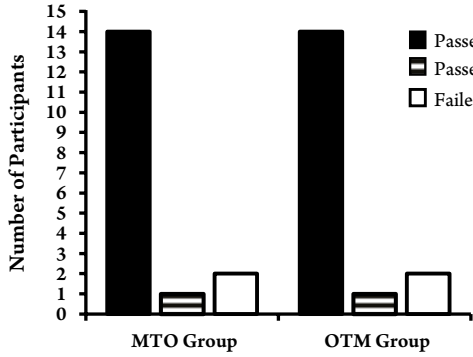

Figure 2. Number of participants who responded in accordance with stimulus equivalence (Passed) in Cycles 1 and 2 (i.e., immediate and delayed emergence, respectively) and who failed to respond in accordance with stimulus equivalence (Failed) after being exposed to the Many-to-One (MTO) and One-to-Many (OTM) training structures.

tinction. Phases 2 and 3 were identical to Phase 1, except for the probability of consequences.

Test of baseline and emergent relations. Once the participant mastered Phase 3, one block of 378 probe trials assessed the maintenance of baseline (54 trials) and the emergence of symmetry (54 trials) and equivalence (270 trials). For the OTM group, symmetric relations were BA, CA, DA, EA, FA, and GA; for the MTO group, symmetric relations were AB, AC, AD, AE, AF, and AG. For both groups, equivalence relations were BC, BD, BE, BF, BG, CB, CD, CE, CF, CG, DB, DC, DE, DF, DG, EB, EC, ED, EF, EG, FB, FC, FD, FE, FG, GB, GC, GD, GE, GF.

During the test, each trial type was presented 3 times, in random order, with no programmed consequences. If the participant did not reach the mastery criterion (at least 95% correct responses for baseline, 52/54; symmetry, 52/54; and equivalence, 256/270), s/he was exposed once more to all training and test phases (Cycle 2).

Results

The same number of participants: Fifteen out of seventeen participants, or 88% (see Figure 2) responded in accordance with equivalence in both groups. In the MTO and in the OTM group, 14 participants passed in the first test (Cycle 1), and one did so after being exposed to the second test (Cycle 2).

Statistical Analyses

To compare the mean outcomes obtained over the first cycle of training and test between both groups, independent-samples *t*-test ($\alpha = .05$), confidence intervals,

Table 2. *Individual performances in phases of training and test for equivalence class formation of participants exposed to the MTO and OTM training structures*

Group	P	Cycle 1				Cycle 2			
		Training Trials	Correct Responses in Test Trials			Training Trials	Correct Responses in Test Trials		
			BSL	SYM	EQ		BSL	SYM	EQ
MTO	15264	450	54	54	270				
	15272	450	54	54	269				
	15291	630	54	54	268				
	15284	180	53	54	269				
	15252	450	54	54	267				
	15268	990	53	54	268				
	15254	450	54	53	267				
	15256	360	54	54	266				
	15274	360	52	54	267				
	15280	450	53	54	265				
	15258	540	52	54	265				
	15270	630	53	54	264				
	15260	1170	53	54	262				
	15288	450	53	54	262				
	15290	2070	53	52	253	90	51	50	248
	15276	1350	52	47	258	90	53	51	263
	15262	2970	51	53	251	90	54	54	261
OTM	15261	2430	54	54	270				
	15267	630	54	54	270				
	15273	450	54	54	270				
	15255	630	53	54	270				
	15271	450	54	54	269				
	15259	450	54	54	268				
	15275	270	54	54	268				
	15277	360	54	54	268				
	15281	270	54	54	268				
	15269	540	53	54	268				
	15279	1080	54	54	267				
	15285	450	52	54	267				
	15283	720	54	53	263				
	15263	450	54	53	262				
	15253	450	54	53	255	90	53	51	258
	15287	540	53	51	255	90	54	54	267
	15265	1890	54	50	232	90	52	51	243

Note. The table shows the number of correct responses in trials of test of Baseline (BSL), Symmetry (SYM), and Equivalence (EQ). Performances bellow the learning criterion (95% of correct responses) are written in bold. P = Participant.

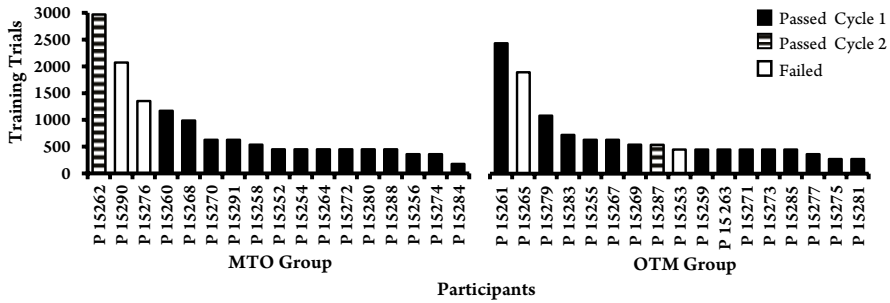

Figure 3. Number of training trials required up to mastering criterion in Cycle 1 by participants exposed to the Many-to-One (MTO) and One-to-Many (OTM) training structures. Participants are ordered according to the number of training trials.

and effect sizes (Hedge's *g* or Cohen's *d*) were calculated (cf. Lakens, 2013). See Table 2 for individual results.

Training trials. In Phase 1, there was no significant difference between groups in the mean number of trials required to learn the baseline ($MMTO$ = 820.6, $SDMTO$ = 725.5; $MOTM$ = 709.4, $SDOTM$ = 584.9, t(32) = 0.49, p = .63, 95% CI [349.22, 571.57], Hedges' gs = 0.16), nor in the number of correct ($MMTO$ = 586.2, $SDMTO$ = 510.4; $MOTM$ = 430.8, $SDOTM$ = 279.7, t(32) = 1.10, p = .28, 95% CI [132.10, 442.92], Hedges' gs = 0.37) or incorrect responses emitted over training ($MMTO$ = 234.4, $SDMTO$ = 219.2; $MOTM$ = 278.7, $SDOTM$ = 352.5, t(32) = -0.44, p = .66, 95% CI [160.82, 249.30], Hedges' gs = 0.15). Figure 3 presents the total number of training trials per participant and their results in the test (pass or fail). In the MTO group—participants whose responding failed to demonstrate stimulus equivalence in the first test were those who required more trials to meet the criterion.

Correct responses in the test of baseline and emergent relations. In the equivalence test, participants in the OTM group had significantly more correct responses in baseline probes ($MOTM$ = 53.7, $SDOTM$ = 0.6) than participants in the MTO group ($MMTO$ = 53.0, $SDMTO$ = 0.9), t(32) = 2.64, p = .01. The difference between the mean number of correct responses in baseline probes in the MTO and OTM group was 0.7 (and above the mastery criterion), but with a large effect size (i.e., 95% CI [0.16, 1.26], Hedges' gs = 0.91). The 16 incorrect responses in baseline probe trials in the MTO group were spread across 12 different relations (B2A1, C2A1, D3A1, F3A1, B1A2, C3A2, D3A2, E3A2, F1A2, B2A3, F1A3, and G2A3).

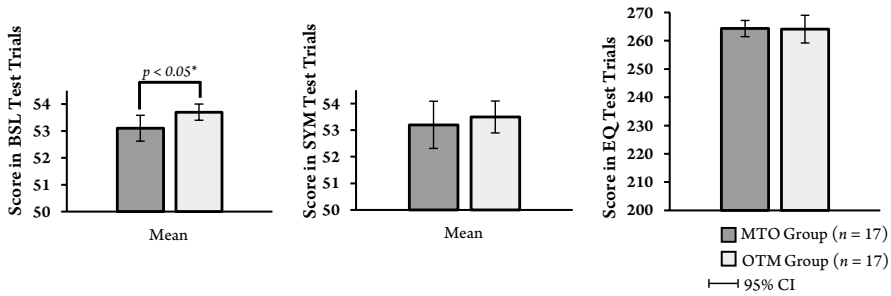

Figure 4. Mean number of correct responses (scores) in baseline (BSL), symmetry (SYM), and equivalence (EQ) test trials for the Many -to -One (MTO) and One -to -Many (OTM) groups in Cycle 1.* Independent-samples t-test (32 df, p < 0.05).

The five incorrect responses in the OTM group were spread across four different relations (A1D2, A1G2, A3C1, and A3G1).

There was no significant difference between groups or large effect sizes considering the number of correct responses in symmetry ($MMTO$ = 53.4, $SDMTO$ = 1.7; $MOTM$ = 53.4, $SDOTM$ = 1.2; t(32) = -0.11, p = .91, 95% CI [0.97, 1.09], Hedges' gs = 0.04) and equivalence probes ($MMTO$ = 264.2, $SDMTO$ = 5.5; $MOTM$ = 264.1, $SDOTM$ = 9.6; t(32) = 0.02, p = .98, 95% CI [5.38, 5.50], Hedges' gs = 0.01). Figure 4 summarizes the average performances of both groups in the test.

Response speed. Reaction time is the latency between presentation of the comparisons and a comparison selection. For statistical purposes (see Baron, 1985; Whelan, 2008), latency was transformed into response speed (1/latency).

The training structure had a large effect on baseline-trials response speed, which differed significantly between groups (see Figure 5). In Phase 1, the MTO group responded significantly faster than the OTM group and had a large effect size over the last five training trials with both correct ($MMTO$ = 0.78, $SDMTO$ = 0.15; $MOTM$ = 0.38, $SDOTM$ = 0.11; t(32) = 8.84, p < .01, 95% CI [0.31, 0.49], Hedges' gs = 2.97) and incorrect responses ($MMTO$ = 0.44, $SDMTO$ = 0.18; $MOTM$ = 0.25, $SDOTM$ = 0.10, t(32) = 3.61, p < .01, 95% CI [0.09, 0.29], Hedges' gs = 1.27). The difference of speed between groups remained significant and the effect remained large over the first five baseline probes ($MMTO$ = 0.57, $SDMTO$ = 0.15; $MOTM$ = 0.37, $SDOTM$ = 0.10; t(32) = 4.63, p < .01, 95% CI [0.11, 0.29], Hedges' gs = 1.53) and over the last five baseline probes with correct responses ($MMTO$ =

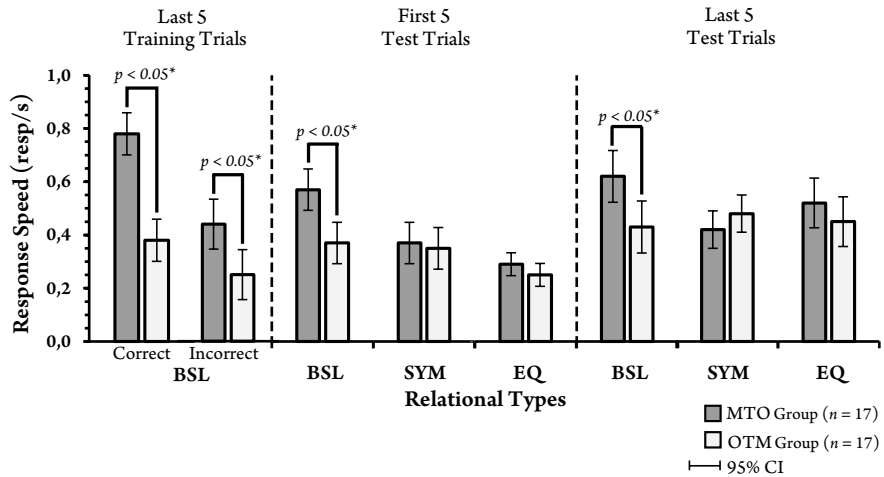

Figure 5. Comparison between the mean median response speed (resp/s)in the Many -to -One (MTO) and One -to -Many (OTM) groups for correct and incorrect responses over the last five baseline training trials; and for correct responses over the first five and last five baseline (BSL), symmetry (SYM), and equivalence (EQ) test trials in Test 1. *Independent-samples t-test (32 df, $p < 0.05$).

0.62, $SDMTO = 0.19$; $MOTM = 0.43$, $SDOTM = 0.14$; $t(32) = 3.31$, $p < .01$, 95% CI [0.07, 0.31], Hedges' $gs = 1.11$).

There was no significant difference between groups in response speed in symmetry trials, neither over the first five ($MMTO = 0.37$, $SDMTO = 0.10$; $MOTM = 0.35$, $SDOTM = 0.12$; $t(32) = 0.38$, $p = .71$, 95% CI [0.06, 0.10], Hedges' $gs = 0.18$), nor over the last five ($MMTO = 0.42$, $SDMTO = 0.14$; $MOTM = 0.48$, $SDOTM = 0.17$; $t(32) = 1.10$, $p = .28$, 95% CI [0.05, 0.17], Hedges' $gs = 0.37$). Likewise, there was no significant difference in equivalence probes, neither over the first five ($MMTO = 0.29$, $SDMTO = 0.08$; $MOTM = 0.25$, $SDOTM = 0.10$; $t(32) = 1.07$, $p = .29$, 95% CI [0.03, 0.09], Hedges' $gs = 0.43$), nor over the last five probes ($MMTO = 0.47$, $SDMTO = 0.18$; $MOTM = 0.43$, $SDOTM = 0.16$; $t(32) = 0.68$, $p = .28$, 95% CI [0.05, 0.17], Hedges' $gs = 0.37$).

Figure 6 shows trends to increase or maintain the mean response speed from the first to the last five probes of baseline, symmetry, and equivalence, within the MTO and OTM groups. Analysis of paired *t*-tests (α=.05) and effect size (Cohens' *d*av) were conducted. The exposure to the test did not produce a large effect on response

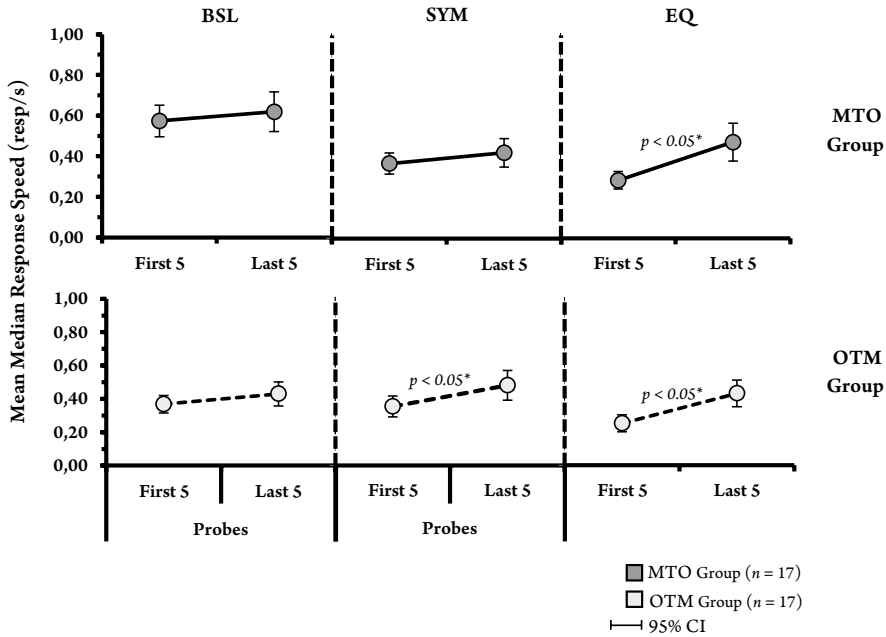

Figure 6. Comparison between the mean median response speed (resp/s) in the first five and last five probes for baseline (BSL), symmetry (SYM), and equivalence (EQ) in Test 1, for the Many -to -One (MTO) and One -to -Many (OTM) groups. *Independent-samples t-test (32 df, p < 0.05).

speed in baseline probes, which did not differ significantly between the first and the last five probes in neither the MTO group ($MFIRST_5$ = 0.57, $SDFIRST_5$ = 0.15; $MLAST_5$ = 0.62, $SDLAST_5$ = 0.19; t(16) = .93, p = .37, 95% CI [-0.05, 0.15], Cohens' $d$av = 0.26) nor the OTM group ($MFIRST_5$ = 0.37, $SDFIRST_5$ = 0.10; $MLAST_5$ = 0.43, $SDLAST_5$ = 0.14; t(16) = 1.87, p = .08, 95% CI [-0.01, 0.13], Cohens' $d$av = 0.50). The exposure to the test produced a large effect on response speed during equivalence probes, which increased significantly from the first to the last five trials in both groups. In the MTO group, the mean speed in equivalence trials increased from 0.29 ($SDFIRST_5$ = 0.08) to 0.47 ($SDLAST_5$ = 0.18), t(16) = -4.14, p < .01, 95% CI [0.09, 0.28], Cohens' $d$av = 1.33). In the OTM group, the speed increased from 0.25 ($SDOTM$ = 0.10) to 0.43 ($SDOTM$ = 0.16), t(16) = -4.20, p < .01, 95% CI [0.09, 0.27], Cohens' $d$av = 1.39.

Only in the OTM group did exposure to the test have a large effect on response speed in symmetry probes, which increased significantly ($MFIRST_5$ = 0.35, $SDFIRST_5$ = 0.12; $MLAST_5$ = 0.48, $SDLAST_5$ = 0.17; t(16) = -2.29, p = .04, 95% CI

[0.01, 0.24], Cohens' *d*av = 0.85). In the MTO group, exposure to the test did not produce significant differences or large effects on response speed in symmetry trials ($MFIRST_5$ = 0.37, $SDFIRST_5$ = 0.10; $MLAST_5$ = 0.42, $SDLAST_5$ = 0.14; t(16) = -1.30, *p* = .21, 95% CI [-0.03, 0.14], Cohens' *d*av = 0.44).

Individual Analysis of Participants who Failed in the Equivalence Test

Trial-by-trial analyses of the baseline training of the six participants who did not respond in accordance with equivalence in the first test were conducted. The analyses attempted to unveil spurious controlling variables associated to the baseline acquisition process under both structures. Figures 7, 8, and 9 show the performance of three of the six participants who failed in the test—P15262 (MTO), P15265 and P15287 (OTM), respectively. Their results illustrate patterns of acquisition that were also presented by other participants (as will be detailed below). The Figures consist of cumulative records of responses to comparisons over the 18 baseline trial types.

MTO group (P15262, P15290, and P15276). The three participants who failed in the MTO group had similar baseline acquisition processes; Figure 7 illustrates the performance of one of them (P15262). P15262[1] required 2970 training trials. He rapidly discriminated the comparisons A1, A2, and A3 under 13 different sample stimuli, but had persistent errors before other five samples (B1, B3, D1, D2, and F2; see graphs with gray background in Figure 7). Three of these stimuli (B1, D2, and F2) were later involved in trials with errors over testing, but accounted for only 12 of the 23 errors (52%) in the test (see Table 3 for all errors). P15290 required 2070 training trials, and had persistent errors before four samples (B2, E3, F2, and G3). All these stimuli were later involved in trials with errors in the test but accounted for only 12 out of 20 (60%) errors. P15276 required 1350 training trials and had persistent errors before six samples (B1, E1, E2, F2, F3, and G2). Four of

1 P15262 should have been exposed to every trial type 165 times, but the number of presentations for some trial types was unbalanced in one of the blocks due to an experimental error (D1/A1A2A3 was present 161 times; E1/A1A2A3, F1/A1A2A3, F2/A1A2A3, and G1/A1A2A3 were presented 164 times each; B1/A1A2A3, C2/A1A2A3, and G3/A1A2A3, 166 times each; and C3/A1A2A3 and E2/A1A2A3, 167 times each).

2 P15290 should have been exposed to every trial type 115 times, but the number of presentations for some trial types was unbalanced in one of the blocks due to an experimental error (B1/A1A2A3, B2/A1A2A3, C1/A1A2A3, C3/A1A2A3, E2/A1A2A3. F1/A1A2A3, and F2/A1A2A3 were presented 114 times each; D1/A1A2A3, D2/A1A2A3, D3/A1A2A3, and E3/A1A2A3, 116 times each; G1/A1A2A3 and G2/A1A2A3, 117 times each).

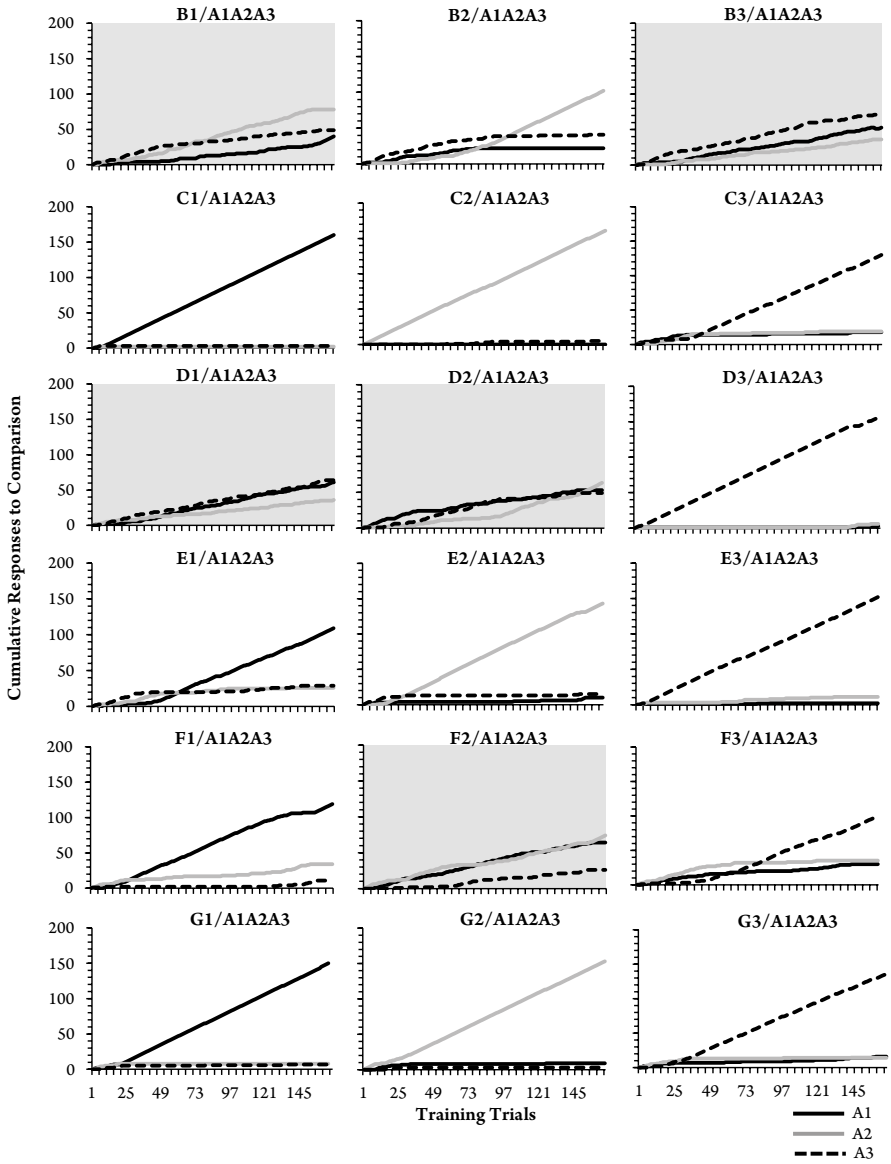

Figure 7. Cumulative responses to comparison stimuli in each trial type over the training of baseline relations, by P15262 (exposed to the Many-to-One training structure). The gray backgrounds indicate trials with persistent incorrect responses.

Table 3. *Trial types with persistent errors over baseline training, and incorrect matchings in Test 1 for participants who failed to respond in accordance with stimulus equivalence. In the "Incorrect Matchings" column, the first alphanumerical term represents the sample stimulus, the second term indicates the comparison selected, and the third term (in parentheses) the comparison defined as correct*

Group	P	Training Trials with Persistent Errors	Test Incorrect Matchings
MTO	15262	B1/A1A2A3, B3/A1A2A3, D1/A1A2A3, D2/A1A2A3, F2/A1A2A3	Baseline: B1A2(A1), F1A2(A1), G2A3(A2) Symmetry: A1D2(D1) Equivalence: C2B3(B2), C2D3(D2), C2D3(D2), C2D3(D2), E3D2(D3), F2D3(D2), G2D3(D2), G3D2(D3), G3D2(D3), C2F3(F2), E2F1(F2), C2E3(E2), D3C2(C3), D3C2(C3), F1E2(E1), F1G2(G1), E3F1(F3), E1G2(G1), F3D1(D3)
	15290	B2/A1A2A3, E3/A1A2A3, F2/A1A2A3, G3/A1A2A3	Baseline: C2A1(A2) Symmetry: A2B3(B2), A2D3(D2) Equivalence: B2C3(C2), B3G2(G3), C2B1(B2), C2B3(B2), C2B3(B2), D2B3(B2), D2B3(B2), E3G2(G3), F2C3(C2), F2D3(D2), G1E3(E1), B3C2(C3), B3C2(C3), B3D2(D3), B3D2(D3),G1E2(E1), C3D1(D3)
	15276	B1/A1A2A3, E1/A1A2A3, E2/A1A2A3, F2/A1A2A3, F3/A1A2A3, G2/A1A2A3	Baseline: D3A1, F1A3 Symmetry: A1E3(E2), A3F1(F3), A3F1(F3), A2D3(D2), A3C1(C3), A3G1(G3), A3G1(G3), Equivalence: F2B1(B2), F2B1(B2), F3G1(G3), G1B3(B1), G1E3(E1), G1E3(E1), F1D3(D1), F1D3(D1), F1D3(D1), G1C3(C1), G1D3(D1), G1D3(D1)
OTM	15265	A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, A1/D1D2D3, A2/D1D2D3, A3/D1D2D3, A2/G1G2G3	Baseline: D2A3(A2), G3A2(A3), C3A2(A3), C3A2(A3) Equivalence: B2C3(C2), B3C1(C3), B3F1(F3), B2G3(G2), C2D3(D2), C2G3(G2), C3B2(B3), C3D2(D3), D2E3(E2), D2F3(F2), D2F3(F2), D2G3(G2), D2G3(G2), D3B1(B3), E1B3(B1), E2G3(G2), E3D2(D3), F2G3(G2), F2G3(G2), F3B1(B3), F3G2(G3), F3G2(G3), G2B1(B2), G2D1(D2), G3B2(B3), G3B2(B3), G3B2(B3), G3D2(D3), G3C2(C3), G3E2(E3), G3F2(F3), G3F2(F3), G3F2(F3), C2F1(F2), C3E1(E3), C3F2(F3), C3F2(F3), F3C2(C3)
	152871	A2/B1B2B3, A3/B1B2B3, A1/D1D2D3, A3/ G1G2G3	Baseline: A1G2(G1) Symmetry: B3A2(A3), B3A2(A3), D2A3(A2) Equivalence: B3C2(C3), B3E2(E3), B3E2(E3), B3G2(G3), D2G3(G2), D2G3(G2), E1B3(B1), C3D2(D3), C3D2(D3), D2C3(C2), D2E3(E2), F1D2(D1) C3E2(E3), F1C2(C1), F1C2(C1).
	15253	None	*Symmetry*: D3A1 Equivalence: B2E3, B1F2, B1F2, B2F3, C2B1, C3E1, C2F1, D1B3, D2C3, D3E1, D1G3, E2D3, F1E2, F2D1, F1E2

Note. Stimulus involved in trials with persistent errors over the baseline training were underlined and written in bold. P = Participant.

[1] P15287 presented low frequency of selections (≤ 25%) of B2, D1, and G3 over the initial training trials.

these stimuli (B1, E1, F2, and F3) were involved in nine out of 21 (43%) errors in the test. In summary, the three participants who failed in the MTO group required an increased number of training trials due to persistent errors before approximately a quarter of the samples. Stimuli involved in trials with increased errors over training were not necessarily involved in errors over testing and could account for only, approximately, half of the later errors.

OTM Group (P15265, P15287, and P15253). The three participants who failed in the OTM group presented more varied baseline acquisition patterns. P15253 required 450 training trials and did not have particularly increased number of errors within any trial type. P15265 required 1890 training trials and had persistent errors in trial types involving the same subsets of comparisons (B and D; see Figure 8), suggesting difficulties to discriminate within the two subsets. He also had persistent errors in the trial type A2/G1G2G3, when G3 was selected repeatedly. Altogether, the B, D, and G3 stimuli could account for 35 out of his 42 (83%) errors in the test.

P15287 required 540 training trials and had persistent errors in four trial types (A2/B1B2B3, A3/B1B2B3, A1/D1D2D3, and A3/G1G2G3). A closer inspection indicated that she rarely selected B2, D1, and G3 among the comparisons (≤ 25% of all choices; see arrows in Figure 9). Although differential response frequencies were not observed in the test, 14 out of her 19 (74%) errors in the test involved B1, B3, D2, D3, or G3.

P15261 passed the test, but her results were analyzed in detail due to the increased number of training trials (2430). In this case, fast responses (ranging from 1076ms to 1941ms per block) were distributed by chance from the 3rd to the 19th block. Only over the last four blocks, correct responses increased systematically, and the average reaction time was greater than 3792ms.

Discussion

The present results did not confirm the main prediction of the DiAn that, the MTO is more effective than the OTM on producing stimulus equivalence, particularly, with increased class sizes (R. R. Saunders & Green, 1999). In the present experiment, 15 out of 17 participants exposed to both structures passed in the test and responded with more than 95% accuracy for more than one hundred emergent relations, tested 3 times each. According to the DiAn, consistent results such as these could not be reached following the OTM arrangement, if the simple dis-

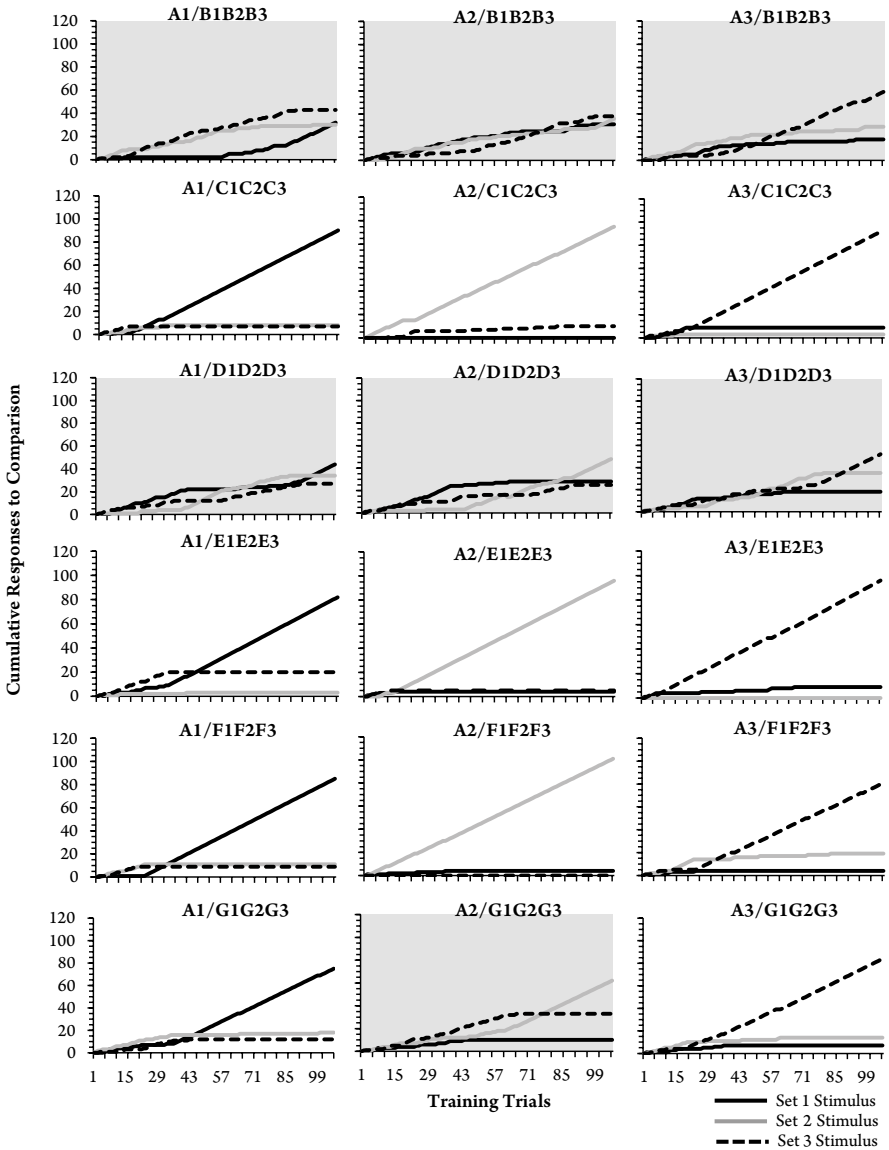

Figure 8. Cumulative responses to comparison stimuli in each trial type over the training of baseline relations by P 15265 (exposed to the One-to-Many training structure). The gray backgrounds indicate trials with persistent incorrect responses.

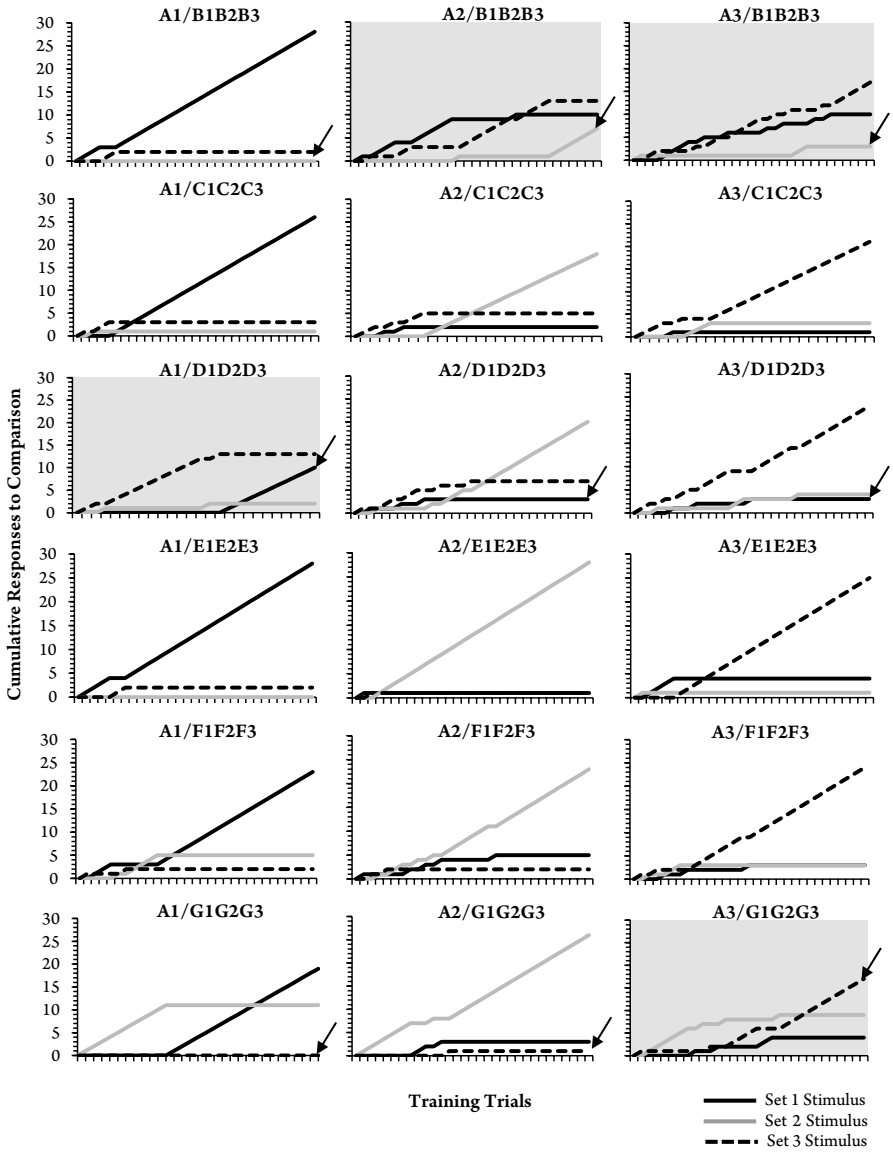

Figure 9. Cumulative responses to comparison stimuli in each trial type over the training of baseline relations by P15287 (exposed to the One-to-Many training structure). The gray backgrounds indicate trials with persistent incorrect responses. The arrows indicate the comparison stimuli selected in less than 25% of the occasions for response.

criminations have not been established. These findings add to the growing body of research indicating similar yields of class formation following both arrangements on the emergence of three equivalence classes with four, five, and six members (Arntzen et al., 2010; Arntzen & Hansen, 2011; Arntzen & Vaidya, 2008; Sadeghi & Arntzen, 2018). Although substantial evidence supports the DiAn prediction that the LS is the least effective training (e.g., Arntzen & Holth, 1997), the DiAn has not been sufficient to account for the variability in the effectiveness of the MTO and the OTM with normal adults.

If considered in isolation, the present results concerning the yields of class formation raise two possibilities about the DiAn: either the training of simple discriminations is not actually critical for the emergence of new relations in general or the critical discriminations were here established in the OTM group by other means. Some authors have speculated that the latter might be the case for participants with extensive pre-experimental learning histories related to simple discriminations (as suggested by Arntzen & Vaidya, 2008) or differential verbal responses (as suggested by R. R. Saunders & Green, 1999). In these cases, sophisticated repertoires could foster the simple discriminations canceling out training structures differential effects.

In fact, most of the experiments supporting the MTO superiority were conducted with children and populations in atypical development (e.g. R. R. Saunders et al., 1999; R. R. Saunders et al., 1988; Spradlin & Saunders, 1986). Alongside, experiments with adults tended to find smaller differences between the MTO and OTM, although some of them were significant (e.g., Arntzen & Hansen, 2011; Fields et al., 1999; Fiorentini et al., 2013). There are also exceptional results, however, indicating great superiority of the MTO even with adults (Hove, 2003) and of the OTM with children and adults (Arntzen & Holth, 1997; Smeets & Barnes-Holmes, 2005), suggesting other relevant interacting variables with the training structures than age. Therefore, it is still relevant to demonstrate if age or an extensive verbal repertoire is really a determinant of differences in the degrees of class formation between the MTO and OTM and, if demonstrated, to explain why they are not a factor for equivalence when combined to the LS or in the above-mentioned exceptional results.

Our results are not only in contrast with predictions of the DiAn but also with a set of experiments on the emergence of two equivalence classes with larger sizes, in which the MTO was superior to the OTM (e.g., Arntzen & Nikolaisen, 2011; Fields et al., 1999; R. R. Saunders et al., 1999; R. R. Saunders et al., 1988; Smeets & Barnes-Holmes, 2005; Spradlin & Saunders, 1986). The results of one of these

experiments (Fields et al., 1999) is particularly noteworthy in relation to the present study, because similar populations (normal adults), class size (7-member), and procedures were employed in both investigations. The present experiment produced higher yields of class formation in general, as expected due to the use of three- rather than two-choice procedures (Carrigan & Sidman, 1992). Our results, however, did not replicate the MTO superiority. The variability between Fields et al.'s (1999) and our results cannot be accounted by the DiAn or by arguments related to population differences.

An explanation for differences between the efficacy of the MTO and OTM when using two-choice procedures was originally posed by Sidman (1994, pp. 527–528), who suggested that the establishment of contextual control by negative stimuli could be more likely in the OTM rather than the MTO structure. Further research is necessary to demonstrate the effect that, if confirmed, could weaken part of the empirical support in favor of the DiAn as an explanation for differences between the training structures. It could suggest, for example, that both structures actually train the overall simple discriminations and the inferiority of the OTM observed in several experiments would not be due to the nontraining of simple discriminations, but to the establishment of spurious sources of control.

The major contribution of the present research was brought by analyses of the baseline training trials compared against test performances. In the present experiment, the greater the number of training trials in the MTO group, the greater were the chances of failure in the equivalence test. It could be argued that the relation between increased number of MTO training trials and failure in the test in two out of three participants was the result of an experimental error that generated a slightly unbalanced number of training trials across trial types. However, the same relation, although unnoticed up to date, has occurred in other experiments (see Experiments 1 and 3 of Arntzen & Holth, 2000).

The trial-by-trial analysis indicated that the increased number of training trials in the MTO group resulted from persistent errors before approximately a quarter of the samples, which supports the interpretation that increased errors over the MTO baseline training is indicative of confusion regarding successive discriminations between some of the samples (K. J. Saunders et al., 1993), as anticipated by the DiAn. However, the relation between persistent errors and failure in the test expands our knowledge by revealing that errors due to the difficulty of successive discriminations between a few sample stimuli might establish irrelevant stimulus control topographies that lead to failure in the equivalence test (cf. Dube & McIlvane, 1996; McIlvane & Dube, 2003), a downside of an increased amount of successive discrim-

inations that was not anticipated by the DiAn. The precise mechanism of the alternative topography of control remains to be elucidated. Here, no particular response pattern was identified in the test, and the sample stimuli involved with persistent errors over training could account for approximately half of the errors in probes. In sum, the results support the notion that the training of simple discriminations is embedded in the training of conditional discriminations (as suggested by the DiAn), but the results suggest that the training of simple discriminations can influence the emergent responding in previously unexpected ways. The generalizability of the relation between training trials and test results, however, is constrained especially due to the limited number of participants who failed in the test. It is fundamental to verify the replicability of the results.

Our results support that successive discriminations might lead to increased errors across training, but they do not demonstrate that the MTO necessarily produces more errors than the OTM, as suggested by the DiAn and as also previously reported (e.g. Hove, 2003; R. R. Saunders et al., 2005; Smeets & Barnes-Holmes, 2005). A note of caution is due here since five participants exposed to the MTO training structure (whose data were excluded from the analysis) did not learn the baseline within a four-hour session, could not attend another meeting, and were dismissed from the experiment.

Failure following the OTM training structure differed from the MTO in a number of ways. First, there was no relation between the amount of training trials and results in the equivalence test in the OTM group. Second, there were no similarities in the baseline acquisition processes, therefore, no obvious baseline pattern predictive of negative test results. It is possible that in two cases (P15265 and P15287) idiosyncrasies related to previous learning histories could have determined the failure. P15265 had one persistent incorrect matching and P15287 emitted few responses towards some of the comparisons over the first trials, suggesting the generalization of pre-experimentally defined stimuli functions. Besides, P15265 had difficulty performing simultaneous discriminations within two subsets of comparisons, an unexpected result in normal adults. Only additional manipulations involving corrective procedures (e.g., substitution of the experimental stimuli), however, could have supported the conclusion about the role of idiosyncrasies on failure. In contrast, another participant who failed in the OTM group provided an interesting datum's baseline learning process was nearly perfect, leaving no indication of other causes for failure than the structure itself.

The MTO and OTM training structures produced not only different baseline acquisition processes in participants who failed in the test, but substantial differ-

ences in response speed over baseline training and test trials, replicating previous reports (e.g., Arntzen & Hansen, 2011). The MTO produced faster responses over baseline (Arntzen & Hansen, 2011; Hove, 2003; R. R. Saunders et al., 1988), but only in the OTM group the response speed in symmetry trials increased significantly from the first to the last five test trials, indicating an increase in stimulus control on emergent responding over testing, what was not observed in the MTO. This result is coherent with R. R. Saunders and Green's (1999) speculation that the increase in response speed could reflect the acquisition of further simple discriminations over testing. However, it is not consistent with another result, namely, that the response speed on equivalence trials of participants in general increased (Arntzen & Hansen, 2011; Hove, 2003; R. R. Saunders et al., 1988), even those in the MTO group (who would have already learned all the potential simple discriminations).

In sum, the present results do not support the main prediction of the DiAn, that the MTO training structure is more effective than the OTM in producing stimulus equivalence with larger classes. Although this study did not fully confirm the DiAn prediction, the results suggest that the training of simple discriminations is embedded in the training of conditional discriminations, but such training can influence the emergent responding in previously unexpected ways. The most significant finding of this study was the indication of a potential negative impact of the number of successive discriminations, typically observed in the MTO training structure, on equivalence class formation.

References

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record,* 60, 437–462. doi:10.1007/BF03395720

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

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*, 309–320. doi:10.1007/BF03395227

Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record, 50*, 603–628. doi:10.1007/BF03395374

Arntzen, E., & Nikolaisen, S. L. (2011). Establishing equivalence classes in children using familiar and abstract stimuli and many-to-one and one-to-many training structures. *European Journal of Behavior Analysis*, 12, 105–120. doi:10.1080/15021149.2011.11434358

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

Arntzen, E., & Vaidya, M. (2008). The effect of baseline training structure on equivalence class formation in children. *Experimental Analysis of Human Behavior Bulletin, 29*, 1–8. Retrieved from https://pdfs.semanticscholar.org/cdb6/1f6d0c4631e9560033490413a49f613541ea.pdf?_ga=2.201926673.1649493850.1525959049-1371614197.1525959049

Baron, A. (1985). Measurement scales and the age-complexity hypothesis. *Experimental Aging Research, 11*, 193–199. doi:10.1080/03610738508259187

Brady, N., & Saunders, K. J. (1991). Considerations in the effective teaching of object-to-symbol matching. *Augmentative and Alternative Communication, 7*, 112–116. doi:10.1080/07434619112331275773

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: a theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 183–204. doi:10.1901/jeab.1992.58-183

Dube, W. V., & McIlvane, W. J. (1996). Some implications of a stimulus control topography analysis for emergent behavior and stimulus classes. *Advances in Psychology, 117*, 197–218. doi:10.1016/S0166-4115(06)80110-X

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. doi:10.1007/BF03395336

Fields, L., & Moss, P. (2007). Stimulus relatedness in equivalence classes: interaction of nodality and contingency. *European Journal of Behavior Analysis, 8*, 141–159. doi:10.1080/15021149.2007.11434279

Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: a methodological analysis. *Journal of the Experimental Analysis of Behavior, 42*, 143–157. doi:10.1901/jeab.1984.42-143

Fiorentini, L., Vernis, S., Arismendi, M., Primero, G., Argibay, J. C., Sánchez, F., Tabullo, A., Segura, E., & Yorio, A. A. (2013). Relaciones de equivalencia de estímulos y relaciones de equivalencia-equivalencia: efectos de la estructura de entrenamiento. *International Journal of Psychology and Psychological Therapy, 13*, 233–242. Retrieved from ri.conicet.gov

Hove, O. (2003). Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record, 53*, 617–634. doi:10.1007/BF03395456

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

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1–12. Retrieved from doi: 10.3389/fpsyg.2013.00863

McIlvane, W. J., & Dube, W. V. (2003). Stimulus control topography coherence theory: foundations and extensions. *The Behavior Analyst, 26*, 195–213. doi:10.1007/BF03392076

Plazas, E. A., & Villamil, C. (2016). Effects of between-classes negative relations training on equivalence class formation across training structures. *The Psychological Record, 66*, 489–501. doi:10.1007/s40732-016-0189-9

Sadeghi, P., & Arntzen, E. (2018). Eye-movements, training structures, and stimulus equivalence class formation. *The Psychological Record*, 1–16 doi:10.1007/s40732-018-0290-3

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: extending the analysis of equivalence. *The Psychological Record, 43*, 725–744. doi:10.1007/BF03395909

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. doi:10.1007/BF03395526

Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior, 71*, 195–214. doi:10.1901/jeab.1999.71-195

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcome. *Journal of the Experimental Analysis of Behavior, 72*, 117–137. doi:10.1901/jeab.1999.72-117

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 49*, 95–115. doi:10.1901/jeab.1988.49-95

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

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

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

Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities, 6*, 41–58. doi:10.1016/0270-4684(86)90005-4

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*, 475–482. doi:10.1007/BF03395630

Zentall, T. R., Wasserman, E. A., & Urcuioli, P. J. (2014). Associative concept learning in animals. *Journal of the Experimental Analysis of Behavior, 101*, 130–151. doi:10.1002/jeab.55

Recibido Julio 1, 2018 /
Received July 1, 2018
Aceptado Marzo 12, 2019 /
Accepted March 12, 2018