EXPERIMENTAL ANALYSIS OF HUMAN BEHAVIOR BULLETIN

Volume 32 | Fall 2020 | Number 1

BRIEF REPORTS

RESEARCH IN PROGRESS

THE EXPERIMENTAL ANALYSIS OF HUMAN BEHAVIOR BULLETIN

About the Bulletin:

- The EAHB Bulletin is published by the Experimental Analysis of Human Behavior Special Interest Group (EAHB SIG), a group organized under the auspice of the Association for Behavior Analysis International (ABAI). Articles in the Bulletin represent the views of the authors. They are not intended to represent the approved policies of the SIG or ABAI, or the opinions of the membership of the SIG or ABAI. Publication costs are paid by dues of the SIG members. Donations my be made to the SIG by sending funds via PayPal to eahbsig@gmail.com.

Types of Manuscripts Accepted

- Brief Reports (≤ 2,000 words): This section is for manuscripts that make an original, valid, and significant contribution to the experimental analysis of human behavior. Examples include empirical reports, literature reviews, or theoretical treatises that address issues of interest the experimental analysis of human behavior. These submissions undergo a rigorous peer-review process and, once published, are considered final publications. As such, these reports have the status as work published in print journals.

- Research in Progress (≤ 1,000 words): This section is for manuscripts reporting work in development or progress. Examples include, but are not limited to, preliminary analyses of data and findings from pilot studies. These submissions are subject to editorial review only. These reports are not considered final publications. Authors are encouraged to acknowledge interim presentation of the data in the Bulletin in their final publication of the results.

- Laboratory Descriptions (≤ 2,000 words): This section is intended for brief descriptions of laboratory procedures. Examples include novel methods or preparations, innovative data analytic methods, and descriptions of research strategies. These submissions are subject to editorial review only. These reports are not considered final publications.

- Technical Information: This section is intended for brief descriptions of technical information unique to the study of the behavior of human participants. Examples include, but are not limited to, descriptions of particularly effective programming code, innovative methods for stimulus presentation, response detection, or contingency management with human participants.

2019 & 2020 Student Paper Award Competition

About the award:

- Graduate students who have completed empirical or theoretical papers in the experimental analysis of human behavior are invited to submit their papers to the EAHB-SIG student paper competition. This competition specifically seeks to reward EAHB scholarship on human operant research topics. All papers submitted are sent to reviewers with expertise in the topic area of the student paper (many of the reviewers serve on the editorial boards of top behavioral journals). All student authors receive their reviews (regardless of award status). As such, this is an excellent opportunity for talented students to receive early peer-review experience.

2019 SPC Award Winner and Abstract

- **Vanessa Ayres-Pereira, Oslo Metropolitan University,** "Different Response Patterns during Baseline Acquisition in Many-to-One and One-to-Many Training Structures and Performance in Tests for Equivalence Class Formation." - (Academic Sponsor: Dr. Erik Arntzen)

- The present experiment compared the outcomes of the One-to-Many (OTM) and Many-to-One (MTO) training structures on the emergence of three 9-member equivalence classes. Forty-two college students participated; half of them were exposed to the MTO and the other half to the OTM training structure. First, participants trained 24 baseline relations concurrently (the OTM group trained AB, AC, AD, AE, AF, AG, AH, and AI, and the MTO group trained BA, CA, DA, EA, FA, GA, HA, and IA). Next, they were exposed to one test block that assessed the emergence of 192 derived relations and the baseline maintenance in random order. After mastering the baseline, 86% of the participants in the MTO group passed in the test and 48% did so in the OTM group. In the MTO group, the baseline acquisition process was characterized by a uniform distribution of responses per comparison stimulus, and failure in the test was predicted by the requirement of greater amounts of trials to learn the baseline. In the OTM group, the baseline acquisition process was characterized by an uneven distribution of responses per comparison, and failure was predicted by non-programmed discrepancies in the number of reinforcements per nodal stimulus.

2020 SPC Award Winner and Abstract

- Student Paper Competition Award temporarily postponed due to the COVID-19 pandemic.

The EAHB SIG extends its warm thanks to all reviewers for the Student Paper Competition

2019 & 2020 EAHB Distinguished Contributions Award

About the award:

- Each year, SIG members vote on nominees for our Distinguished Contributions Award. This award is given to an individual who has made substantial, long-term contributions to the Experimental Analysis of Human Behavior and is presented each year at the ABAI annual convention.

- Past awardees include: Alan Baron, Murray Sidman, Joe Brady, Joe Spradlin, Charlie Catania, David Schmitt, Grayson Osborne, Travis Thompson, Jack Michael, Howie Rachlin, Nate Azrin, Harry Mackay, Deisy de Souza, Carol Pilgrim, Phil Hineline, and Mike Perone.

2019 DC Award Winner and Abstract

- Distinguished Contribution Award temporarily postponed due to the COVID-19 pandemic.

2020 DC Award Winner and Abstract

- Distinguished Contribution Award temporarily postponed due to the COVID-19 pandemic.

BRIEF REPORT

ON THE USE OF MORPHING TECHNIQUES IN CONDITIONAL-DISCRIMINATION PROCEDURES

Erik Arntzen, Richard K. Nartey, and Hanna Steinunn Steingrimsdottir
OSLO METROPOLITAN UNIVERSITY

A stimulus class is a set of stimuli that can be arrayed along some continuum, all of which occasion the same response after having been trained to occur in the presence of only some of the stimuli in the set (Fields & Reeve, 2001; Keller & Schoenfeld, 1950). Concept formation is inferred when the same response is occasioned by many stimuli in a set or by many of the relations among the stimuli in a set (Dinsmoor, 1995; Keller & Schoenfeld, 1950). These sets of stimuli which differ in terms of their physical attributes have been called perceptual classes (e.g., Fields et al., 1997; Lea & Ryan, 1984), relational classes (e.g., Lea & Ryan, 1984; Smoke, 1932), equivalence classes (e.g., Sidman, 1994; Sidman & Tailby, 1982), functional classes (e.g., Sidman et al., 1989), and semantic categories (e.g., Rosch & Mervis, 1975).

Stimulus equivalence is defined as novel responding after preliminary training of conditional discriminations that is in accord with the properties of reflexivity, symmetry, and transitivity (Sidman & Tailby, 1982). Previously unrelated stimulus relations will arise without direct training, and members of an equivalence class will become mutually substitutable (Green & Saunders, 1998).

Equivalence class formation has been demonstrated with both verbally competent people (Dugdale & Lowe, 2000), including adults and typically developing children (e.g., Arntzen & Vaidya, 2008; Pilgrim et al., 1995; Sidman & Tailby, 1982), and those with developmental disabilities or autism (e.g., Arntzen et al., 2010; LeBlanc et al., 2003). Responding according with stimulus equivalence has also been demonstrated after training with different stimulus modalities such as olfactory (Annett & Leslie, 1995), haptic (Belanich & Fields, 1999), tactile (O'Leary & Bush, 1996), and gustatory (Hayes et al., 1988). Stimulus equivalence has also been demonstrated with a variety of visual stimulus materials, such as different abstract stimuli (e.g., Sidman & Tailby, 1982), consonant-vowel-consonant syllables (e.g., Fields et al., 1997), three-dimensional objects (e.g., Devany et al., 1986), and meaningful pictures (e.g., Arntzen, 2004; Arntzen & Lian, 2010; Arntzen & Nikolaisen, 2011; Bentall et al., 1993; Holth & Arntzen, 1998; Smeets & Barnes-Holmes, 2005).

Research has also shown that responding in accordance with equivalence classes could be influenced by the type of stimuli used. For example, different types of pictorial stimuli (e.g., Bentall et al., 1993) or pronounceable stimuli (e.g., Mandell & Sheen, 1994) have enhanced the formation of equivalence classes. In a series of experiments adult participants trained to form larger equivalence classes with a linear series (LS) training structure have shown that only abstract shapes as stimuli produce little class enhancement, while the inclusion of at least one meaningful stimulus in a class of abstract shapes can influence the likelihood of forming an equivalence class among that set of stimuli (e.g., Arntzen & Mensah, 2020; Arntzen et al., 2015; Fields et al., 2012). This enhancing effect has also been shown in experiments with a many-to-one (MTO) training structure both with adults (Arntzen, 2004; Lyddy et al., 2000; Rustad Bevolden & Arntzen, 2018) and children (Arntzen & Lian, 2010; Smeets & Barnes-Holmes, 2005). In the present experiment, the MTO is used because it produces a high rate of yields (number of participants forming equivalence classes) (see Fields et al., 2020, for a discussion).

Corresponding author: Erik Arntzen, Oslo Metropolitan University, Department of Behavioral Science, P O Box 4 St. Olavs Plass, 0130 Oslo, Norway. E-mail: erik.arntzen@equivalence.net

Arntzen (2004) examined how responding in accordance with equivalence class formation changes as a function of the position of meaningful stimuli, pictures, and nonsense syllables among fifty college students using the MTO-training structure (AB, CB, DB, and EB). The main findings were that when A-stimuli were meaningful stimuli, 10 of 10 participants responded in accordance with equivalence, whereas five of 10 responded in accordance with equivalence when the meaningful stimuli were presented at the end of the training (as E-stimuli). In addition, when all stimuli were nonsense syllables, four of 10 participants responded in accordance with equivalence, and when all the stimuli were Greek/Arabic letters, only three of ten participants responded in accordance with equivalence. Furthermore, four of 10 participants responded in accordance with equivalence when the A-stimuli were meaningful stimuli, and participants were required to use keys on the keyboard. In another experiment, Arntzen and Lian (2010) examined the effect of meaningful stimulus as a node (C) in the formation of equivalence classes in children. The experiment was arranged as an MTO training structure training on six conditional discriminations (AC/BC) and testing for three 3-member equivalence classes. The main findings were that the children formed equivalence classes more likely when the nodal stimuli were pictures than when they were abstract. In sum, there seems to be a difference in outcome on equivalence tests between the reference groups (all abstract shapes and pictorial stimuli as a node) both for LS and MTO.

The present experiment was arranged to replicate and extend the outcome in Arntzen and Lian (2010) with larger classes by introducing sets of nodal stimuli degrading the effect of pictorial stimuli. Hence, we wanted to study discriminability (the degree to which a participant discriminates between stimuli (e.g., White et al., 1985)) when introducing mixed variants of abstract and pictorial stimuli. Since one way of degrading the effectiveness of the role of meaningful stimuli on stimulus equivalence class formation, we employed a morphing technique to change the stimuli used as nodes gradually. There are at least two ways to do morphing of stimuli. One way is to have two endpoints, and have the midpoint morphed to be 50% of each of the endpoints. Fields and colleagues (Fields, Matneja, et al., 2002; Fields, Reeve, et al., 2002) used this morphing technique to study linked perceptual classes and generalized categorization. Another morphing technique is when one endpoint is morphed into the other endpoint. Such a morphing technique was used in the present experiment where pictorial stimuli as one of the endpoints was morphed into abstract stimuli the other endpoint to study the effect of degrading of pictorial stimuli on the likelihood of forming stimulus equivalence classes.

We asked about the effects of degrading the pictorial stimuli from employing pictorial stimuli to abstract shapes as nodal stimuli in an MTO training structure (AE/BE/CE/DE) when children were trained conditional discriminations and testing for the emergence of three 5-member equivalence classes. Thus, a between-group design with five groups differing in the degree of morphing of the E-stimuli (nodal stimuli) in the stimulus set were used. Group 1 included all pictorial stimuli, and the pictures were gradually morphed into the abstract shapes through Groups 2, 3, 4, and finally, Group 5 with E-stimuli as abstract shapes.

METHOD

Participants

Fifty primary school pupils, 21 girls and 29 boys aged between 7–10 years (average 9.04 years) participated in the experiment. Their teachers and parents consented to participate after being briefed about what the experiment was about. The consent included an assurance that the children could withdraw from the experimental session at any time without any negative consequences. None of the children had previously taken part in any experiments. After finishing the experiment, the children were thanked and given small gifts, like pencils, markers, etc.

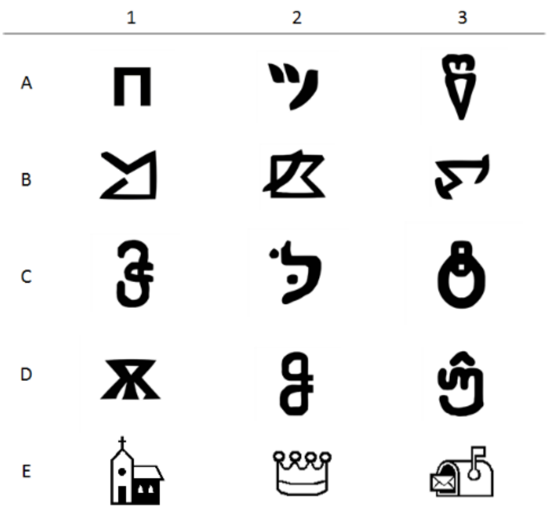
Setting and Apparatus

The experiment was conducted in a small room measuring approximately 5m2. There were two tables and two chairs on which children sat during the sessions. An HP Compaq nc6320 laptop computer running Windows 8 and a screen with a 16.8 in diagonal length and with a 16 X 9 horizontal-to-vertical ratio was used to conduct the experiment. The children used an external mouse to control the position of the cursor. A customized software developed in collaboration with the first author controlled the presentation of stimuli and the recording of all responses throughout the experiment.

Stimuli

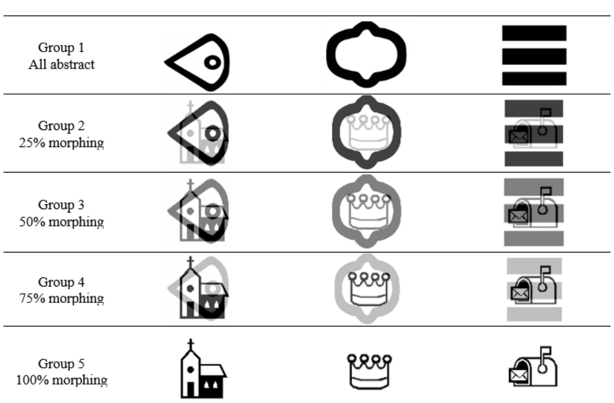
Abstract shapes and meaningful pictures were used as stimuli (see Figures 1 & 2).

Figure 1



Note. The Stimuli used in Experiment 1 for Group 5. The E-stimuli varied from one group to the other (shown in Figure 2).

Figure 2



Note. The variation of the E-stimuli used in the five conditions. This is the correct order.

Morphing

The morphing program used to create the stimuli was FantaMorph Version 5.4.1 (http://www.fantamorph.com/index.html). With the help of the program, one stimulus was "morphed" into another stimulus (see Figure 2).

Design

The experiment was arranged as a between-group design with children randomly assigned to five different groups, each with 10 participants. The difference between the stimuli used for each group was the degree of morphing of the E-stimuli in the stimulus set. For Group 1, the E-stimuli were pictorial stimuli. The meaningful stimulus was degraded across groups, from 25% into the abstract shapes for Group 2, then 50% morphing in Group 3, and 75% and 100% morphing in Groups 4 and 5 (ABS), respectively (see Figure 2). Each participant was trained with the baseline relations required to form three 5-member equivalence classes in an MTO training structure. All participants completed the experiment in a day.

Procedure

All baseline relations were trained in an MTO training structure and with a simultaneous protocol. Furthermore, the baseline relations were trained in a serialized basis, and programmed consequences were provided following the selection of comparisons for each trial (see Table 1). All conditional discriminations were established before introducing a test for all emergent relations, and the trials were randomly presented (see Table 2). Tables 1 and 2 show each of the trial representations; the first stimulus is the sample, and the other three are the comparison stimuli, whereas the underlined comparison is the correct comparison (e.g., A1/E1E2E3; see details in Table 1). The comparison stimuli were randomly presented in different positions in the four corners of the screen with one corner blank. Every correct response (experimenter-defined) to a sample-comparison pair, was followed by a textual stimulus on the screen as "awesome", "very good," etc. Every incorrect response (experimenter-defined) to a sample-comparison pair, was followed by the textual stimulus "incorrect" on the screen. The programmed consequences in a 36-trial block

Table 1

Overview of the Conditional Discrimination Training

Training Phase		Mastery Criterion	Likelihood of Programmed Consequences
Acquisition	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3	9/9	100
	B1/E1E2E3, B2/E1E2E3, B3/E1E2E3	9/9	100
	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3	17/18	100
	C1/E1E2E3, C2/E1E2E3, C3/E1E2E3	9/9	100
	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3, C1/E1E2E3, C2/E1E2E3, C3/E1E2E3	25/27	100
	D1/E1E2E3, D2/E1E2E3, D3/E1E2E3	9/9	100
	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3, C1/E1E2E3, C2/E1E2E3, C3/E1E2E3, D1/E1E2E3, D2/E1E2E3, D3/E1E2E3	33/36	100
Maintenance	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3, C1/E1E2E3, C2/E1E2E3, C3/E1E2E3, D1/E1E2E3, D2/E1E2E3, D3/E1E2E3,	33/36	75
	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3, C1/E1E2E3, C2/E1E2E3, C3/E1E2E3, D1/E1E2E3, D2/E1E2E3, D3/E1E2E3,	33/36	50
	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3, C1/E1E2E3, C2/E1E2E3, C3/E1E2E3, D1/E1E2E3, D2/E1E2E3, D3/E1E2E3,	33/36	25
	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3, C1/E1E2E3, C2/E1E2E3, C3/E1E2E3, D1/E1E2E3, D2/E1E2E3, D3/E1E2E3,	33/36	0

Table 2

Overview of Trial Types During Testing

Baseline	A1/E1E2E3, A2/E1E2E3, A3/E1E2E3, B1/E1E2E3, B2/E1E2E3, B3/E1E2E3, C1/E1E2E3, C2/E1E2E3, C3/E1E2E3, D1/E1E2E3, D2/E1E2E3, D3/E1E2E3,
Symmetry	E1/A1A2A3, E2/A1A2A3, E3/A1A2A3, E1/B1B2B3, E2/B1B2B3, E3/B1B2B3, E1/C1C2C3, E2/C1C2C3, E3/C1C2C3, E1/D1D2D3, E2/D1D2D3, E3/D1D2D3,
Equivalence	A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, A1/C1C2C3, A2/C1C2C3, A3/C1C2C3, A1/D1D2D3, A2/D1D2D3, A3/D1D2D3, B1/A1A2A3, B2/A1A2A3, B3/A1A2A3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3, B1/D1D2D3, B2/D1D2D3, B3/D1D2D3, C1/A1A2A3, C2/A1A2A3, C3/A1A2A3, C1/B1B2B3, C2/B1B2B3, C3/B1B2B3, C1/D1D2D3, C2/D1D2D3, C3/D1D2D3, D1/A1A2A3, D2/A1A2A3, D3/A1A2A3, D1/B1B2B3, D2/B1B2B3, D3/B1B2B3, D1/C1C2C3, D2/C1C2C3, D3/C1C2C3

Note. No programmed consequences were presented in the test.

consequence was displayed in the middle of the screen for 500 ms. Termination of the programmed consequence was followed with a 500 ms inter-trial interval. Between trials, the mouse cursor was returned to the center of the screen.

Instruction

Each session started with the participant seated facing the computer monitor and presented with the following instructions on the computer screen, which was read out loud by the experimenter:

"In a moment, a stimulus will appear in the middle of the screen. Click on this by using the computer mouse. Three stimuli will then appear in three corners of the screen. Choose one of them by clicking on it with the mouse. If you choose the stimulus we have defined as correct, words like "very good," "excellent," and so on will appear on the screen. If you press a wrong stimulus, the word "wrong" will appear on the screen. At the bottom of the screen, the number of correct responses you have made will be counted. During some stages of the experiment, the computer will not tell you if your choices are correct or wrong. However, based on what you have learned so far, you can get all the tasks correct. Please do your best to get everything right. Thank you and good luck!" No further instructions were given after this before and after the experiment had started.

Acquisition of Baseline Relations

AE relations were trained first in a block containing nine trials, three of each trial type (see Table 1). A total number of nine correct trials out of the nine trials in the training block was required to proceed to the training of the next relation. BE relations were then introduced with the same requirements as AE, followed by a mix of AE and BE relations with blocks of 18 trials with a mastery criterion of at least 17 correct. Then, participants were exposed to the CE relations with the same requirements as for AE and BE training, followed by a mix of AE, BE, and CE relations with blocks of 27 trials with a mastery criterion of at least 25 correct. The last relation trained was the DE relation with the same requirements as AE, BE, and CE training. The final acquisition of the baseline training involved a block of all the relations in a mixed training: AE, BE, CE, and DE relations. The block had three presentations of each 12 trial types adding up to 36 trials. Out of the 36 trials, a minimum of 33 correct trials was required to complete the acquisition of the baseline relations. If the mastery criterion was no met for any of the trained relations as described above, the participants repeated the block until they did so.

Maintenance of Baseline Relations

Programmed consequences followed the selection of any comparison in every trial in all the blocks during acquisition. In the maintenance training, the percentage of trials that produced programmed was reduced first to 75%, then 50%, 25%, and finally to 0% (see Table 1). If the mastery criterion was not reached in any of the blocks, the blocks were repeated until the criterion was reached. When the participants reached the mastery criterion on the last block with no programmed consequences, the test for emergent relations was introduced.

Test for Emergent Relations

A test block that contained 180 trials followed the last block with no programmed consequences (see Table 2). Of the 180 trials, there were 36 baseline trials, 36 symmetry trials, and 108 equivalence trials. All of the trials were randomly presented and without programmed consequences. The formation of equivalence classes was defined by the selection of at least 90% correct comparisons that were consistent with the experimenter-defined classes for each type of relation.

RESULTS

Acquisition of Baseline Conditional Discriminations

The mean of the number of trials required to establish baseline conditional discriminations was computed for each group (see Figure 3). A visual inspection of the graph shows an inverted u-function across different stimuli with the highest number of trials for Group 3 and the lowest number of trials less variation for Group 1. A Welch's ANOVA test showed an effect of stimulus material W(4, 45)=7.006, p=0.0011. T-tests with Welch correction showed a significant difference in the number of trials to mastery criterion when comparing Group 1 (meaningful stimuli as node group) and Group 5 (abstract stimuli as the node group) (p=.016). Also, statistical differences were shown between Group 1 and Groups 4 (p=.0023) and 3 (p=.016), but not for Group 2 (p=0.136). Thus, the degree of degrading the nodal stimulus did to some extent influence the speed of acquisition of the baseline relations.

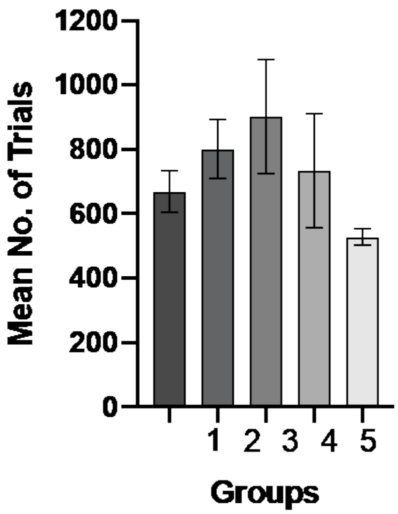
When the speed of acquisition (number of trials) was compared for all participants who went on to form equivalence classes and those who did not form classes, fewer trials were required to acquire the baseline relations for those who formed classes than those who did not (see Figure 4). However, a Welch's test showed that the difference was not statistically significant, t=1.779, p=.0855.

Equivalence Class Formation

As shown in Figure 5, there was a decrease in responding in accordance with stimulus equivalence as a function of morphing steps or degrading of the meaningful stimuli. In Group 1 (PIC as node), all children responded in accordance with stimulus equivalence, 70% of the children in Groups 2 and 3 formed the classes, and 50% of the children formed the experimenter defined classes in Group 4. When exposed to abstract shapes as the node, 10% of the children formed the experimenter-defined stimulus equivalence classes. A chi-square analysis showed the differences in yields to be statistically significant, X^2 (4) = 18.33, p =.001. Fisher Exact Tests indicated that equivalence yields of Group 1 participants differed significantly with Groups 5 (p=.000) and 4 (p=.033) and, but not Groups 2 and 3 (p=.211). The results suggest that the formation of equivalence was a function of degrading the pictorial stimulus as the nodal stimulus where greater yields were produced with meaningful stimuli and a decrease in the likelihood of equivalence class formation as the nodal stimulus becomes more and more abstract.

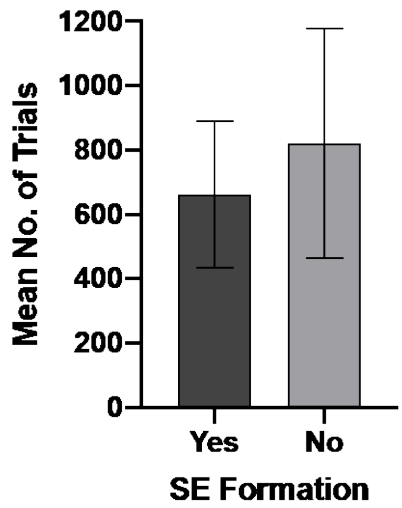
Figure 6 shows individual performance on separate relations in the test for emergent relations. In Group 1, all participants responded in accordance with experimenter-defined classes. In Group 2, only P4733 did not have the baseline relations intact, while two more participants (P4720 and P4719) did not respond in accordance with the experimenter-defined criterion for symmetry and equivalence trials. In Group 3, seven participants (P4705, P4713, P4716, P4736, P4706, P4722, and P4747) responded in accordance with the experimenter-defined criterion for all relations. In Group 4, five participants (P4741, P4737, P4727, P4703, and P4745) responded in accordance with the experimenter-defined criterion for all relations. For participants in Group 5, three participants (P4738, P4702, and P4721) had baseline relations intact during testing and P4702 had responded in accordance with the experimenter-defined criterion for symmetry, and P4721 also for equivalence.

Figure 3



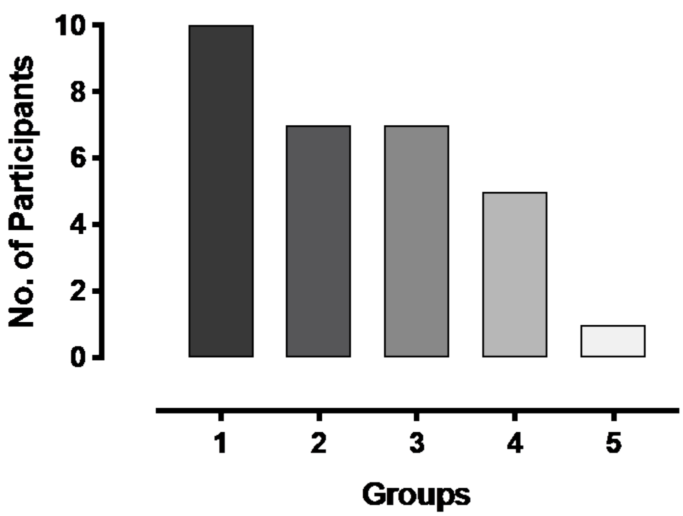
Note. Mean number of training trials across groups. Each line shows the standard error of the mean for the respective group

Figure 4



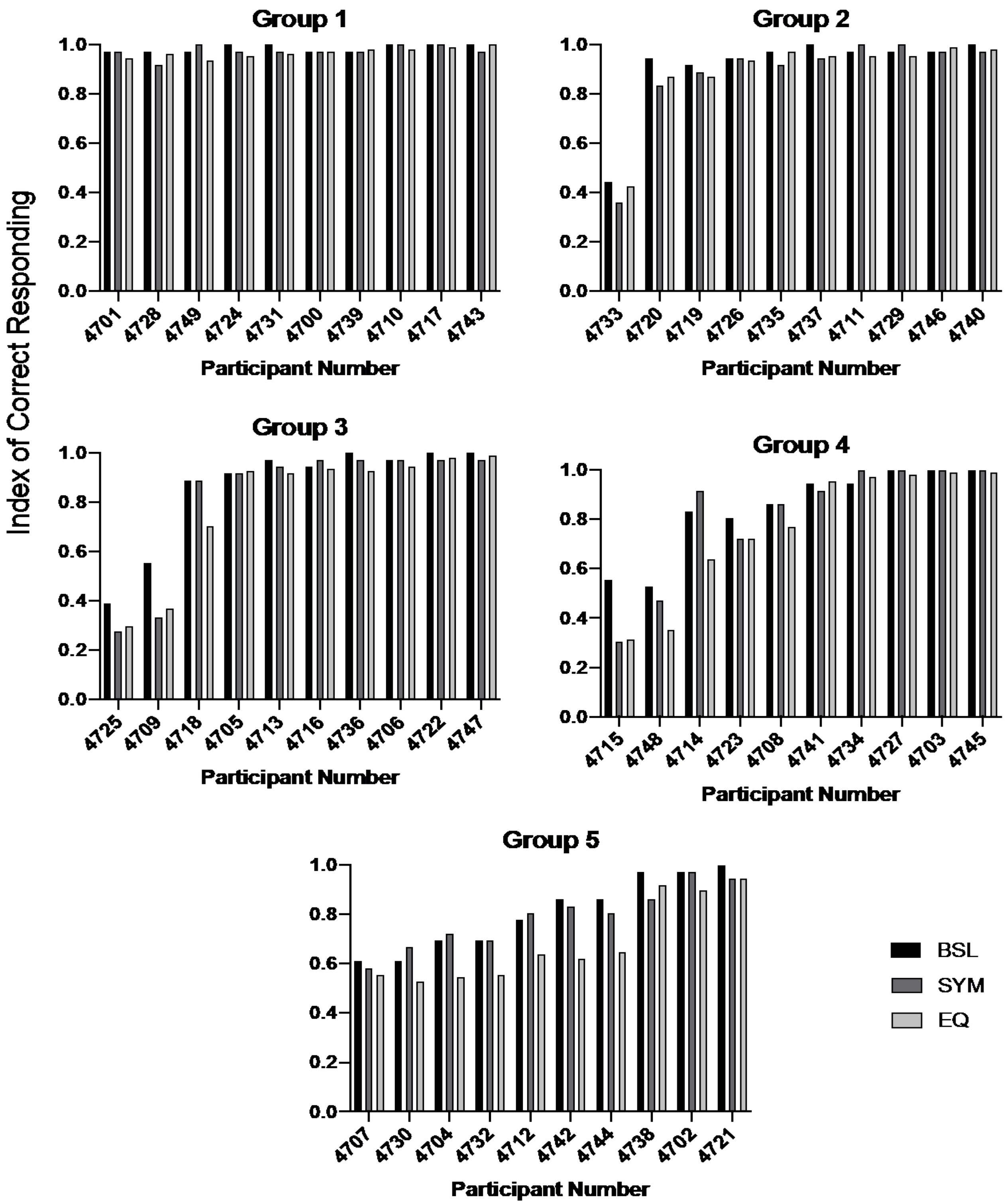
Note. Number of trials to criterion for participants who responded in accordance with equivalence (Yes) and participants who did not respond in accordance with equivalence (No). Each line shows the standard error of the mean

Figure 5



Note. Number of participants responding in accordance with stimulus equivalence.

Figure 6



Note. Index of correct responses across participants for baseline, symmetry, and equivalence trials in the test for emergent relations

DISCUSSION

The degrading of the pictorial stimuli which served as nodes were done by using a morphing procedure with gradually changing stimuli from meaningful or pictorial stimuli to abstract shapes. All children formed classes when the nodal stimuli were meaningful (the PIC group), while few children responded in accordance with equivalence when the nodal stimuli were abstract shapes (the ABS group). When the pictorial stimuli were degraded, there was a systematic reduction in yields across groups. Thus, the formation of equivalence classes was shown to be a function of the degrading of pictorial stimuli, which served as the nodal stimulus.

Number of Trials to Mastery Criterion

The results showed that the number of trials to meet the mastery criterion was lowest for the PIC group and highest for the ABS group. These findings are in accordance with Arntzen and Lian (2010). In the present experiment, there was no linear function decreasing number of training trials and the visibility of the pictorial stimuli as nodes. Thus, the speed (number of responses) at which participants acquired the baseline relations necessary for the formation of equivalence classes was only to some extent influenced by the degree of morphing of the nodal stimulus used. The variation of the number of trials to criterion is much smaller for the two reference groups compared to the three groups with degrading of the pictorial stimuli. Finally, it is also suggested that acquisition speed was not a determinant or predictor of subsequent equivalence class formation.

Effect of Degrading Meaningful Stimuli on Class Formation

A recent publication by Fields et al. (2020) discusses how yield has been used as a measurement in stimulus equivalence research, but also points out the necessity of analysis of individual data. The present experiment has included both group and individual data. Regarding the group data, the equivalence class formation for the reference groups (PIC and ABS groups) replicated the findings in Arntzen and Lian (2010) employing the same training structure. Also, these findings are consistent with results from the two reference groups in several experiments which have employed the linear series structure in testing for three 5-member equivalence classes in adult participants (e.g., Arntzen & Mensah, 2020; Arntzen & Nartey, 2018; Nartey et al., 2015).

The visual analysis of the individual data supports the notion about the analytic unit of analysis of equivalence relations (Sidman, 1994, 2000). One interpretation of Sidman's theory is that all relations should emerge if the necessary baseline conditional discriminations are established. Forty-seven of the 50 children showed a pattern of responding during testing in which all features of stimulus equivalence would prevail, if any. The three remaining children showed that the baseline performance was intact during testing.

Possible Mechanisms

There seem to be several possible mechanisms for the findings in the present experiment. Regarding the difference in class enhancement comparing the results from the PIC and the ABS groups, it has been argued that the enhancement could be related to that the pictorial stimuli have different behavioral properties. Hence, meaningful stimuli are stimuli that have at least two functions (e.g., Fields et al., 2012). The degree of meaningfulness may vary from one stimulus to the other, however, and also dependent on the participants learning history. Thus, the meaningfulness of a stimulus could be relative. Also, the enhancement effect of pictorial stimuli is because the meaningful stimuli are presumably members of already established categories prior to their use in the equivalence class formation research.

Another variable of importance in the comparison of the PIC and the ABS group could be the possibility of the naming of the meaningful stimuli presented as E-stimuli as for the participants in Group 1. This possibility of naming stimuli as a class enhancement mechanism is in accordance with other experiments (e.g., Ma et al., 2016). For the participants in Groups 2, 3, and 4 had a decreasing number of participants forming equivalence classes, the children could still be able to name stimuli even the degrading of pictorial stimuli.

Finally, the decreasing number of participants forming equivalence classes could be related to discriminability of stimuli. Other experiments have shown how the discriminability can influence the performance on training on conditional discriminations in non-humans (e.g., Jones & White, 1992; White et al., 1985) and humans (e.g., Doughty et al., 2014; Hayashi & Vaidya, 2008, 2012).

Design

In the present experiment, the arrangement of the experimental conditions was employed as a group design. We have presented individual data, however. Quite a high number of the earlier studies within stimulus equivalence research have been arranged as demonstrations without a strict experimental design, and lately, many experiments are arranged as group designs. Some of the reason for employing a group design in research within experiments on emergent relations is because of the possible effect of order and sequence when experiments are arranged as single-case research design, in particular the effect of order. We will emphasize the need for presenting individual data along with group data in experiments arranged as group design.

Generality of the Findings

Since the present experiment replicated the findings for different groups of participants, it will therefore be very important to successfully explore the effect of the inclusion of meaningful stimuli on the formation of equivalence classes as a single-case experimental design and with MTO. The suggestion of training structure is based on findings showing that MTO and OTM training structures produce higher yields than LS (see Arntzen, 2012, for an overview). An experiment arranged as a single-case experimental research design with different morphing steps will contribute greatly towards the generalization of the effect of meaningful stimuli on the formation of classes. In such an experiment, each participant will be exposed to a number of morphing steps, from 0% morphing to 100% morphing.

An LS training structure with more members could be useful in future experiments if the focus is on the effect of number of nodes. The LS will provide a sensitive measure for the experiment of the effect of the use of varying degrading the pictorial stimuli of the nodes on class formation within participants.

Also, experiments including eye-tracking technology, could give important information about the controlling variables for responding in conditions with morphed stimuli as in the three variants of degrading the pictorial stimuli as in the present experiment. Previous experiments have shown time spent observing positive stimuli than negative stimuli (e.g., Huziwara et al., 2016) and patterns as fixation time and transitions between stimuli (e.g., Sadeghi & Arntzen, 2018). Eye-movements analyses when presented the degrading pictorial stimuli could give essential information of what aspects of the stimuli participants are attending to.

Summary

In the present experiment, we explored how pictorial stimuli compared to abstract shapes influenced the outcome on number of training trials to reach the mastery criterion of baseline conditional discrimination and tests for the emergence of equivalence classes. A morphing technique was used to present three variants of degrading of the pictorial stimuli, which serve as nodal stimuli. The main findings showed a decrease of in responding in accordance with stimulus equivalence across the variants of degrading of the pictorial stimuli.

REFERENCES

Annett, J. M., & Leslie, J. C. (1995). Stimulus classes involving olfactory stimuli. *The Psychological Record*, *45*, 439–450. https://doi.org/10.1007/BF03395153

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record*, *54*, 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*, 123–135. https://doi.org/10.1080/15021149.2012.11434412

Arntzen, E., Halstadtro, L. B., Bjerke, E., & Halstadtro, M. (2010). Training and testing theoretical music skills in a boy with autism using a matching-to-sample format. *Behavioral Interventions*, *25*(2), 129–143. https://doi.org/10.1002/bin.301

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures as nodes. *The Psychological*

Record, *60*, 659–677.
https://doi.org/10.1007/BF03395738
Arntzen, E., & Mensah, J. (2020). On the effectiveness of including meaningful pictures in the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior*, *113*(2), 305–321.
https://doi.org/10.1002/jeab.579
Arntzen, E., & Nartey, R. K. (2018, Sep). Equivalence class formation as a function of preliminary training with pictorial stimuli. *Journal of the Experimental Analysis of Behavior*, *110*(2), 275–291.
https://doi.org/10.1002/jeab.466
Arntzen, E., Nartey, R. K., & Fields, L. (2015). Enhanced equivalence class formation by the delay and relational functions of meaningful stimuli. Journal of the *Experimental Analysis of Behavior*, *103*, 524–541.
https://doi.org/10.1002/jeab.152
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.
https://doi.org/10.1080/15021149.2011.11434358
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.
http://www.eahb.org/NewSitePages/Bulletin-Homepage.htm
Belanich, J., & Fields, L. (1999). Tactual equivalence class formation and tactual-to-visual cross-modal transfer. *The Psychological Record*, *49*, 75–91.
https://doi.org/10.1007/BF03395308
Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *46B*, 187–214.
https://doi.org/10.1080/14640749308401085
Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior*, *46*, 243–257.
https://doi.org/10.1901/jeab.1986.46-243
Dinsmoor, J. A. (1995). Stimulus control: Part I. The Behavior Analyst, *18*(1), 51–68.
http://www.abainternational.org/TBA.asp
Doughty, A. H., Brierley, K. P., Eways, K. R., & Kastner, R. M. (2014). Effects of Stimulus Discriminability on Discrimination Acquisition and Stimulus-Equivalence Formation: Assessing the Utility of a Multiple Schedule. *The Psychological Record*, *64*(2), 287–300.
https://doi.org/10.1007/s40732-014-0001-7
Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior*, *73*, 5–22.
https://doi.org/10.1901/jeab.2000.73-5
Fields, L., Arntzen, E., & Doran, E. (2020). Yield as an essential measure of equivalence class formation, other measures, and new determinants. *The Psychological Record*, *70*, 175–186.
https://doi.org/10.1007/s40732-020-00377-3
Fields, L., Arntzen, E., Nartey, R. K., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior*, *97*, 163–181.
https://doi.org/10.1901/jeab.2012.97-163
Fields, L., Matneja, P., Fitzer, A., & Shamoun, K. (2002). The formation of linked perceptual classes. *Journal of the Experimental Analysis of Behavior*, *78*, 271–290.
https://doi.org/10.1901/jeab.2002.78-271
Fields, L., & Reeve, K. F. (2001). A methodological integration of generalized equivalence classes, natural categories, and cross-modal perception. *The Psychological Record*, *51*, 67–87.
https://doi.org/10.1007/BF03395387
Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior*, *68*, 67–92.
https://doi.org/10.1901/jeab.1997.68-67
Fields, L., Reeve, K. F., Matneja, P., Varelas, A., Belanich, J., Fitzer, A., & Shamoun, K. (2002). The formation of a generalized categorization repertoire: effect of training with multiple domains, samples, and comparisons. *Journal of the Experimental Analysis of Behavior*, *78*, 291–313.
https://doi.org/10.1901/jeab.2002.78-291
Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), Handbook of research methods in human operant behavior (pp. 229–262). Plenum Press.
Hayashi, Y., & Vaidya, M. (2008, 2008/12/01). Effects of Stimulus Discriminability on the Acquisition of Conditional Discriminations in Adult Humans. *European Journal of Behavior Analysis*, *9*(2), 173-183.
https://doi.org/10.1080/15021149.2008.11434303
Hayashi, Y., & Vaidya, M. (2012). A comparison of prompting versus feedback in the development of conditional discriminations in adult humans. *European Journal of Behavior Analysis*, *13*(1), 25–37.
https://doi.org/10.1080/15021149.2012.11434402
Hayes, L. J., Tilley, K. L., & Hayes, S. C. (1988). Extending equivalence and membership to gustatory stimuli. *The Psychological Record*, *38*, 473–482.
https://doi.org/10.1007/BF03395040
Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus

equivalence or consistent nonequivalence. *The Psychological Record*, *48*, 81–110. https://doi.org/10.1007/BF03395260

Huziwara, E. M., de Souza, D. d. G., & Tomanari, G. Y. (2016). Patterns of eye movement in matching-to-sample tasks. *Psicologia: Reflexao e Critica*, *29*, 2–10. https://doi.org/10.1186/s41155-016-0010-3

Jones, B. M., & White, K. G. (1992, 07). Sample-Stimulus Discriminability and Sensitivity to Reinforcement in Delayed Matching to Sample. *Journal of the Experimental Analysis of Behavior*, *58*(1), 159–172. https://doi.org/10.1901/jeab.1992.58-159

Keller, F. S., & Schoenfeld, W. N. (1950). Principles of Psychology. Copley Publishing Group.

Lea, S. E. G., & Ryan, C. M. E. (1984). Feature analysis of pigeons' acquisition of concept discrimination. In T. G. Bever & H. S. Terrace (Eds.), *Animal Cognition* (pp. 233–261). Erlbaum.

LeBlanc, L. A., Miguel, C. F., Cummings, A. R., Goldsmith, T. R., & Carr, J. E. (2003, 11). The effects of three stimulus-equivalence testing conditions on emergent US geography relations of children diagnosed with autism. *Behavioral Interventions*, *18*(4), 279–289. https://doi.org/10.1002/bin.144

Lyddy, F., Barnes-Holmes, D., & Hampson, P. J. (2000). The effect of stimulus meaningfulness on the formation of equivalence classes. *European Journal of Behavior Analysis*, *1*, 71–87. https://doi.org/10.1080/15021149.2000.11434157

Ma, M. L., Miguel, C. F., & Jennings, A. M. (2016, May). 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

Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Processes*, *32*, 29–46. https://doi.org/10.1016/0376-6357(94)90025-6

Nartey, R. K., Arntzen, E., & Fields, L. (2015, Mar). Enhancement of equivalence class formation by pretraining discriminative functions. *Learning & Behavior*, *43*(1), 20–31. https://doi.org/10.3758/s13420-014-0158-6

O'Leary, C. A., & Bush, K. M. (1996). Stimulus equivalence in the tactile modality. *The Psychological Record*, *46*, 509–517. https://doi.org/10.1007/BF03395180

Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior*, *63*, 225–238. https://doi.org/10.1901/jeab.1995.63-239

Rosch, E., & Mervis, C. B. (1975). Families resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rustad Bevolden, K., & Arntzen, E. (2018). How robust is the correspondence between responding on matching-to-sample tests and sorting tests? *Norwegian Journal of Behavior Analysis*, *45*, 125–142. http://www.nta.atferd.no/journalissue.aspx?IdDocument=729

Sadeghi, P., & Arntzen, E. (2018). Eye-movements, training structures, and equivalence class formation. *The Psychological Record*, *68*, 461–476. https://doi.org/10.1007/s40732-018-0290-3

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

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*, 5–22. https://doi.org/10.1901/jeab.1982.37-5

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior*, *52*, 261–274. https://doi.org/10.1901/jeab.1989.52-261

Smeets, P. M., & Barnes-Holmes, D. (2005, 05/20/). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes*, *69*, 281–293. https://doi.org/10.1016/j.beproc.2004.12.009

Smoke, K. L. (1932). An objective study of concept formation. *Psychological Monographs*, *42*, 157. https://doi.org/10.1037/h0093297

White, K. G., Pipe, M.-E., & McLean, A. P. (1985). A note on the measurement of stimulus discriminability in conditional discriminations. *Bulletin of the Psychonomic Society*, *23*, 153–155. https://doi.org/10.3758/BF03329810

BRIEF REPORT

A PRELIMINARY TWO-PHASE TEST OF HOW INEQUITY AVERSION IS MODULATED BY PREVIOUS DYADIC INTERACTIONS

Marcelo Benvenuti[a,b], José de Oliveira Siqueira[a], Carla Jordão Suarez[a,b], Cesar Augusto Villela Silva do Nascimento[a,b], Karen M. Lionello-DeNolf[c], and Josele Abreu-Rodrigues[d]

[a]DEPARTMENT OF EXPERIMENTAL PSYCHOLOGY, UNIVERSITY OF SÃO PAULO, SÃO PAULO, BRAZIL
[b]NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY ABOUT COGNITION, BEHAVIOR AND TEACHING, SÃO CARLOS (HEAD OFFICE), BRAZIL
[c]ASSUMPTION UNIVERSITY, MASSACHUSETTS, UNITED STATES
[d]DEPARTMENT OF BASIC PSYCHOLOGICAL PROCESSES, UNIVERSITY OF BRASÍLIA, BRASÍLIA, BRAZIL

Inequity aversion can be defined as the refusal of gains or strong, negative emotional behavior when there is an unfair distribution of outcomes (Brosnan & de Waal, 2014; Fehr & Schmidt, 1999). Aversion to inequity may be investigated by the use of an inequity game, an experimental procedure derived from the economic games literature (see McAuliffe, Blake, Steinbeis, & Warneken, 2017). In this game, an allocation is distributed between two players (Player One and Player Two) by an experimenter. If Player One accepts the allocation, both players receive their allocated payoffs. If Player One rejects the allocation, neither player receives a payoff. Rejection of an unequal distribution may be used as a measure of inequity aversion. Inequity aversion may occur in situations of disadvantageous inequity (DI), in which one rejects an outcome that is less than that of a partner. Inequity aversion may also occur in situations of advantageous inequity (AI), in which one rejects an outcome that is more than that of a partner.

With respect to the ontogenesis of humans' "sense of fairness", there seem to be important differences between aversion to DI and AI (Blake et al., 2015; McAuliffe, Blake, Kim, Wrangham, & Warneken, 2013; Corbit, McAuliffe, Callaghan, Blake, & Warneken, 2017). Children across diverse societies show aversion to DI as young as 4 years old (Blake & McAuliffe, 2011; Blake et al., 2015; McAuliffe et al., 2013; Shaw & Olson, 2012). In contrast, emergence of aversion to AI is more variable. Blake et al. (2015), for example, found evidence of AI in older children (8 years old) in some countries (the US, Canada, and Uganda), but not in other countries (India, Mexico, Peru, and Senegal). Aversion to AI may be more related to social cues and cultural context than aversion to DI, in that it is strongly observed and varies less between people from different cultures and individuals from different species, such as humans and monkeys (e.g., Blake et al., 2015; Brosnan & de Waal, 2014). In this paper, we ask if flexibility of aversion to DI may be best investigated with experimental procedures with more

Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant # 465686/2014-1), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Grant # 2014/50909-8), and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Grant # 88887.136407/2017-00). Carla Suarez was supported by a Master's fellowship from CNPq (grant # 134205/2012) and Cesar Augusto was supported by Master's fellowship from CAPES (grant # 1579255).
We are thankful to Dr. Debora Hollanda Souza for comments on the first version of this manuscript and to Dr. Paulo Sergio Silveira for his assistance with charts and data analysis.
Correspondence should be sent to Marcelo Benvenuti, marcelobenva@gmail.com.

long-lasting opportunities to interact with a partner.

There is interesting empirical evidence showing that more long-lasting opportunities to interact with a partner may have a strong influence on social behavior in test conditions (Abreu-Rodrigues, Natalino, & Aló, 2002; Avalos, Ribes-Iñesta, Ortiz, & Serna, 2015; Marwell & Schmitt, 1975; Ribes-Iñesta, Rangel, Pulido, Valdez, Ramírez, Jiménez, & Hernández, 2010; Schmitt, 1998; Silverstein, Cross, Brown, & Rachlin, 1998). Typically, there is no programmed cost for social interactions, but outcomes can vary due to some aspects of a partner's behavior and the experimental condition. In this kind of experimental strategy, there are usually at least two experimental conditions, and participants may experience repeated interactions with a given partner under stable conditions before there is a change in experimental conditions. For experimental purposes, dyadic interactions may be controlled when one member of the dyad is a confederate.

An example of a procedure that involves a long-lasting opportunity to interact with a partner is a puzzle task that can be shared with another person (Avalos, Ribes-Iñesta, Ortiz, & Serna, 2015; Ribes-Iñesta, Rangel, Pulido, Valdez, Ramírez, Jiménez, 2010). The participant and a partner (who is a research confederate) need to solve puzzles, presented on individual computer screens that show both the participant's puzzle and the confederate's puzzle. The participant and confederate can put pieces on their own puzzle as well as on the other puzzle. If the participant or the confederate places a piece on their own puzzle, he/she receives 10 points. In addition, if the participant or confederate places a piece on the other person's puzzle, they may both receive 10 points (points delivery varied between studies). Participants (college students) rarely put pieces on the confederate's puzzle when confederates put pieces only on their own puzzles in baseline sessions. Across experimental conditions, the percentage of pieces that the confederate placed on the participant's puzzle varied from 0 to 25, 50, 75, and 100%, in ascending or descending order. Results showed that participants placed pieces on the confederate's puzzle in the same proportion as the confederate placed pieces on the participants' puzzle. The flexibility of cooperative strategies has also been investigated by Silverstein, Cross, Brown, and Rachlin (1998). The study used a two-phase procedure with an iterated prisoner's dilemma game. Participants were initially assigned to one of four experimental conditions in which they played with a confederate, and the confederate's strategy varied: (1) tit-for-tat, (2) play randomly, (3) always cooperate, or (4) always defect. In a second phase, participants played the prisoner's dilemma game with each other (instead of with the confederate). During this latter condition, cooperation was the predominant strategy mainly for those participants previously exposed to the tit-for-tat condition.

Collectively, experimental results indicate that different cooperative behaviors may be flexible in social situations, which requires a special analysis regarding the learning mechanisms involved when one person's decisions may be affected by the other person's behavior. The main aim of the present study was to devise a two-phase experiment to investigate flexibility of aversion to DI in young adults as a result of dyadic interactions with AI. We investigated whether aversion to DI could be modulated by a previous experimental history in which a confederate acted in a "friendly" manner that produced AI. We compared this situation with two control situations: one in which participants interacted with an "unhelpful" partner who did not permit AI and one in which participants did not have previous experience with a partner (they were exposed directly to the DI test).

METHOD

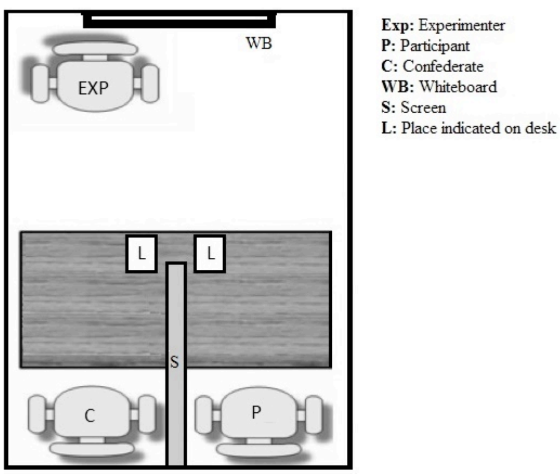
Participants

Fifty-nine college students, ranging in age from 18 to 27 years, were recruited from a university campus: 33 were female and 26 were male. All participants signed informed consent forms that had been approved by a Brazilian research ethics committee (process CAAE: 19646713.4.0000.5561).

Experimental Environment and Materials

The experiment was conducted in a research room at the University of São Paulo. In the room, there were two tables and two chairs, a folding screen, and a whiteboard (Figure 1). The experimenter was stationed next to the whiteboard and had visual access to the participant and confederate. The participant and the confederate sat at separate tables, and visual contact between them was limited by a folding screen positioned between them. They could only see each other's hands, which card the other person played on each trial, and the outcome presented by the experimenter on the whiteboard. The participant and confederate each had a blue and a green card, and there was a space marked on each table indicating where they needed to place the chosen cards on each trial. They also had access to a pencil and notepad, on which they could write whatever they wanted. In a pre-experimental phase with four trials, all participants learned general rules about choosing cards, combinations of cards and outcomes, and the payoff matrix for different combinations of cards.

Figure 1



Note. An overhead illustration of the experimental setting.

Procedure

The participants were randomly assigned to one of three experimental groups: friendly confederate (FRICON), unhelpful confederate (UNHCON), or control/no previous history (NOHIST). Fifteen participants in the FRICON group and all participants in the UNHCON group completed two experimental phases: In the first phase (labeled Phase AI), a history with a friendly or unhelpful confederate was manipulated. In the second phase (labeled Phase DI), the production of DI was tested.

There were 29 participants in the FRICON group, 15 participants in the UNHCON group, and 15 participants in the NOHIST group. In the FRICON group, 14 participants were excused from the experiment before Phase DI due to failure to meet the Phase AI criterion (see below). Data from these participants were not included in the overall data analysis. Participants in the NOHIST group only completed the test for DI (Phase DI).

Experimental Task and Payoff Matrix

On each trial, the outcomes for the participant and confederate were determined by the combined choices of blue and/or green cards. When the participant and confederate both chose the blue card, there was an inequitable outcome. When one or both choose a green card, there was an equitable outcome. In Phase AI, inequity was advantageous to the participant; in Phase DI, inequity was disadvantageous to the participant (Table 1).

When the participant entered the experimental room, the confederate was already sitting in one of the chairs, behind the screen. Written instructions were given to the participant and confederate simultaneously. The experimenter asked that instructions be read in silence. The instructions were:

> *This study is not about intelligence, and it is not about assessing your intellectual abilities. When you're done, you'll get more explanations. You will be working with a partner, and both of you will have an identical task to perform. You and your partner will receive two cards (one blue and one green). When the experimenter says the word "Attention," you must make a choice: place your hand on the blue card or place your hand on the green card. After your choice, the experimenter will say the word, "Now!" At this point, put the chosen card in the place indicated on your desk so that you and your partner can see each other's choices. On each trial, you will receive a certain number of points. The number of points you will receive depends on your choice and the choice of your partner. The experimenter will notify you when the session is finished. Please remain seated and do not talk to your partner or the experimenter*

Table 1.
Payoff Matrix for Participants and Confederates in Phases AI and DI

	Card combinations	Points P	Points C	Confederate's Choice	Trials
PHASE AI — Advantageous Inequity to the Participant					
FRICON	Blue-Blue*	5	2	Blue	15
UNHCON	Blue-Blue*	5	2	Green	15
NOHIST	(not exposed to this experimental phase)				
PHASE DI — Disadvantageous Inequity to the Participant					
ALL GROUPS	Blue-Blue*	2	5	Blue	12

*Any other combination produced equal outcomes: 2 points to both players.
"P" refers to participant and "C" refers to confederate.

> *during the session. All instructions are contained on this sheet. If you have questions, reread the instructions (do not ask the experimenter any questions). When you're ready to begin, raise your right hand.*

After returning the paper with the general instructions, the participants received the following specific instructions, also printed on paper:

> *If you choose the blue card and the participant beside you also chooses the blue card (combination: blue–blue), you will earn five points and the participant next to you will earn two points. If you or your partner choose the green card (combinations: blue–green, green–blue, or green–green), you both will earn two points.*

Experimental Design

After the participant read the general and specific instructions, there was a pre-experimental phase that consisted of four trials. The confederate alternately chose the green and blue cards on these trials (i.e., green-blue-green-blue). The outcome on these trials was the same as that in the next phase: inequity favorable to the participant in cases in which both players chose the blue card, and equity with any other card combination. The objective of this pre-experimental phase was to permit the participants to test the instructions about card choices and points distributions. After the four trials, the experimental phase was initiated without any notification.

Phase AI: advantageous inequity to the participant. In this phase, the confederate's behavior varied depending on the participants' experimental group. The confederate's behavior was pre-determined in order to permit or prevent inequity favorable to the participant. In the FRICON group, the confederate used the blue card and allowed the participant to earn five points while the confederate earned two points on every trial. In the UNHCON group, the confederate used the green card and did not allow the participant to earn five points (i.e., both the participant and confederate earned two points) on every trial. There were 15 trials in this phase. For participants in the FRICON group, there was a criterion to finish the phase: Participants were only exposed to the next phase if they played the blue card on at least 10 trials, and the blue card was played on the last three trials.

Phase DI: disadvantageous inequity to the participant. At the beginning of this phase, the experimenter provided additional written instructions to the participant and confederate on how to earn points. These instructions indicated that the payoff matrix was reversed: now blue-card choices resulted in the confederate earning five points and the participant earning two. On every trial for all groups, the confederate always chose the blue card. There were 15 trials in Phase DI. This phase included the NOHIST group that received only the preliminary, general instructions about gains; also, this group was only exposed to Phase DI and thus did not have an experimental history with a friendly or unhelpful partner.

Data Analysis

The data file was organized in long format. Each data file record (each choice for each participant) contained the following variables: participant identification (ID), participant's choice (green, blue) (PC), phase (1, 2) (Phase), group (FRICON, UNHCON, NOHIST) (Group), and block (1, 2, 3) (Block). The total number of records was 1125.

The dependent (outcome) variable was PC, and category Green was the reference category. A full factorial generalized linear mixed model with binomial distribution and logit link function (repeated measure logistic regression) was used to test the main variables. Fixed effects for factors Phase and Group was controlled by random effect of ID.

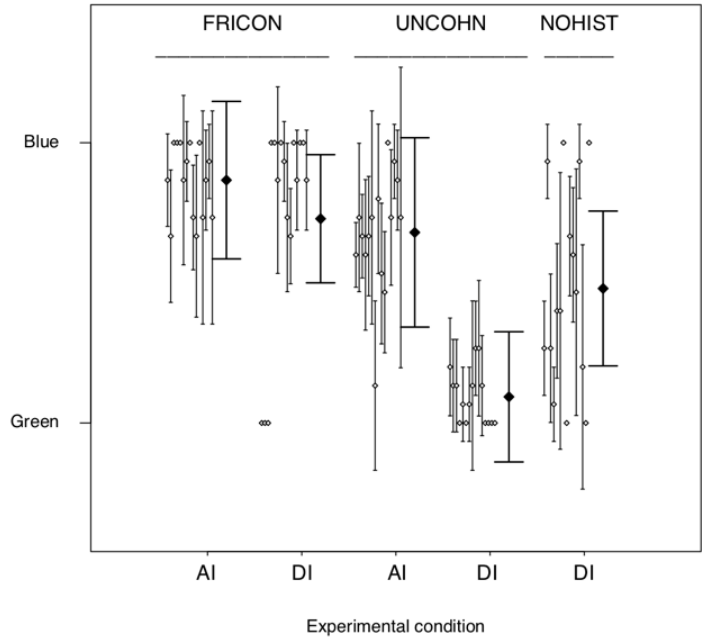
RESULTS

Our main purpose with this experiment was to see if dyadic interactions with a confederate constrain aversion to DI for FRICON participants (compared to participants in UNCOHN and NOHIST groups). Our first analysis compared participants in the three experimental groups in two consecutive phases (AI and DI).

Figure 2 depicts the results from the two phases for participants in FRICON and UNCOHN groups and from Phase DI for participants in NOHIST group. Closed markers show estimated marginal means and open markers show individual participant data in each phase. Considering participants in the FRICON group, there was a small decrease in blue choices in Phase DI. For these participants, blue choices were still more frequent than green choices in the second phase. For Participants in the UNCOHN group, in contrast, the percentage of trials on which participants chose the blue card dropped to close to 40% in Phase DI, indicating an unwillingness to produce DI. The NOHIST group showed strong variability in choices: Some participants choose the blue card on most of the trials, but other participants choose the green card on most of the trials. This result clearly illustrates the importance of previous experience with AI regarding the more consistent data in Phase DI for participants in the FRICON and UNCOHN groups.

For statistical comparisons, we adopted a significance level of 0.05. The interaction effect was significant, $F(1, 1120) = 25.354$, $p < 0.001$. There was a simple main effect of Group for both Phases DI and AI, $F(1, 1120) = 6.691$, $p = 0.01$ and $F(2, 1120) = 46.787$, $p < 0.001$, respectively. Using pairwise contrasts in Phase DI, the differences among the levels for groups FRICON, UNHCON, and NOHIST were significant (FRICON - UNHCON: $t(1120) = 9.373$, Sidak adjusted $p < 0.001$, FRICON - NOHIST: $t(1120) = 2.535$, Sidak adjusted $p = 0.011$, and UNHCON - NOHIST: $t(1120) = 3.639$, Sidak adjusted $p = 0.001$). The estimated marginal means (adjusted proportion of blue card choices) were 0.928 and 0.689 in Phase AI for groups FRICON and UNHCON, respectively, and were 0.799, 0.083, and 0.482 in Phase DI for groups FRICON, UNHCON, and NOHIST, respectively.

Figure 2



Note. Blue and green choices in Phases AI and DI for participants in the FRICON and UNCOHN groups and in Phase DI for participants in the NOHIST. Closed markers show estimated marginal means and open markers show individual participant data in each phase. Error bars indicate confidence intervals of 95%.

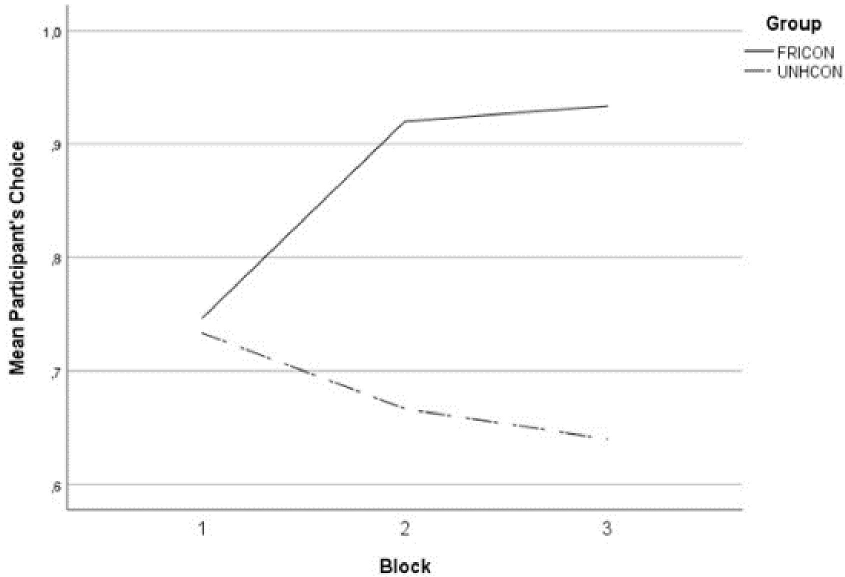
Figure 3 depicts the difference between groups in Phase AI in three blocks of five trials. Examining the data in smaller trial blocks shows whether differences between the FRICON and UNHCON groups in Phase AI occurred at the beginning of the experiment or were established during that phase. At the beginning of the experiment, participants from both groups chose the blue card on a similar number of trials initially, but they differed in the second and third block. Specifically, participants in the FRICON group began to make more blue-card choices, whereas those in the UNHCON group tended to make slightly fewer blue-card choices.

DISCUSSION

Our results clearly show that previous personal history (Phase AI) affected decisions in a situation with DI. We clearly constrained aversion to DI for FRICON participants. We were able to see this by comparing FRICON participants with UNHCON participants and also by comparing these groups with the NOHIST group (participants who were not exposed to Phase AI). These results are very strong and consistent, even with a relatively small number of participants in each experimental group. Another way to conceptualize these results is that participants who experienced a friendly confederate in Phase AI (i.e., a partner who permitted AI to the participant) produced DI to themselves in Phase DI.

Our experimental strategy was successful in demonstrating how flexibility in inequity aversion may be produced in a two-phase experiment. The primary contribution of this strategy is that a majority of previous reports in the experimental literature test inequity aversion using just a few trials presented in a single condition that are part of between-group strategies. There are, however, some limitations in our analysis that may be improved in future investigations. The main limitation is related to the criteria for advancing participants in the FRICON group to Phase DI: this criterion was used only for participants in that group, which may have biased the comparisons between groups in the Phase AI. The use of the same criteria for all participants in

Figure 3



Note. Difference between groups in Phase AI in three blocks of five trials

the initial phase may produce more comparable results in different conditions.

The results of Phase DI resemble a "tit-for-tat" situation that is common in behavioral games like the iterated prisoner's dilemma (Axelrod, 1984). Research on reciprocity has created an interesting discussion related to the evolution of cooperation and has contributed to quantitative models of social behavior and cultural evolution (Axelrod & Dion, 1988; Axelrod & Hamilton, 1981; Boyd & Richerson, 1985). Developmental and cultural mechanisms may explain changes in inequity aversion at different age stages or group levels, respectively, but they are less predictive in dealing with the fact that inequity aversion may be established or constrained within the repertoire of a given individual. For this reason, cross-cultural variation about fairness is sometimes hard to interpret and open to discussion about which psychological or cultural mechanisms are at work (Delton, Krasnow, Cosmides, & Tooby, 2010).

Our results may help integrate contributions from learning principles (usually described as content-independent processes) with evolutionary mechanisms that promote sociality (usually described as content-dependent processes; Tooby & Cosmides, 1992). Questions about learning are usually best investigated by using procedures that permit long-lasting interactions between participants in a cooperative task before a test. The literature on associative learning phenomena, for example, has repeatedly illustrated that learning rarely occurs in just one or a few trials and often requires long-lasting interactions (Rescorla & Wagner, 1972). Reciprocity and inequity aversion may be strongly explained by evolved mechanisms. However, at the same time, the results from Phases AI and DI suggest a cumulative effect of learning during the dyadic interactions. This effect may be also partially explained by the principles of stimulus control (Sidman, 2000; Urcuioli, 2013) because in Phase AI the confederate's behavior (blue or green choices) is a condition associated with different rates of point's delivery. The effects of arousal (Killeen, Hanson, & Osborne, 1978) may also aid in understanding differences between participants in FRICON and UNHCON groups: Arousal refers to the cumulative activation of behavior by the presentation of outcomes (e.g., points gained on each trial of a game) that can only be fully observed once participants have had multiple exposures to the same type of trial (Killeen & Sitomer, 2003).

As in other fields of psychology, the question is not about "innate" versus "acquired," but, rather, is a matter of identifying mechanisms and how those different mechanisms work and are integrated (Tooby, Cosmides, & Barrett, 2005). This is an exciting avenue to explore because it permits reconciliation of a genetic disposition to behave in a cooperative manner with the role of personal experience.

REFERENCES

Abreu-Rodrigues, J., Natalino, P. C., & Aló, R. M.(2002). Instruções e iniquidade de reforços: efeitos sobre o comportamento competitivo [Instructions and inequity of reinforcement: Effects over competitive behavior]. *Psicologia: Teoria e Pesquisa*, *18*, 83–94. http://dx.doi.org/10.1590/S0102-37722002000100010

Avalos, L. P., Ribes-Iñesta, E., Ortiz, I. L., & Serna, B. L. (2015). Interacciones altruístas totales como función de la inducción de reciprocidad. *Revista Mexicana de Análisis de la Conducta*, *41*, 32–52. http://dx.doi.org/10.5514/rmac.v41.i1.63688

Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science*, *242*, 1385-90. Retrieved from https://www.jstor.org/stable/1702320.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390-96. Retrieved from http://www.jstor.org/stable/1685895.

Axelrod, R. (1984). The evolution of cooperation. New York, NY: Basic Books.

Blake, P. R., & McAuliffe, K. (2011). "I had so much it didn't seem fair": Eight-year-olds reject two forms of inequity. *Cognition*, *120*, 215-24. http://dx.doi.org/10.1016/j.cognition.2011.04.006

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., Kleutsch, L., Kramer, K. L., Ross, E., Vongsachang, R., Wrangham, R., & Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*, 258-62. https://doi.org/10.1038/nature15703

Boyd, R., & Richerson, P. J. (1985). Culture and the evolutionary process. Chicago, IL: University of Chicago Press.

Brosnan, S. F., & de Waal, F. B. M. (2014). Evolution of responses to (un)fairness. *Science*, *346*(6207), 1-19. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4451566/pdf/nihms683162.pdf.

Corbit, J., McAuliffe, K., Callaghan, T., Blake, P. R., & Warneken, F. (2017). Children's collaboration induces fairness rather than generosity. *Cognition*, *168*, 344-56. https://doi.org/10.1016/j.cognition.2017.07.006.

Delton, A. W. (2010). A psychological calculus for welfare tradeoffs. (Doctoral dissertation). University of Santa Barbara, California.

Delton, A. W., Krasnow, M., Cosmides, L., & Tooby, J. (2010). Evolution of fairness: Rereading the data. *Science*, *329*(5990), 389. https://doi.org/10.1126/science.329.5990.389-a

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817-868. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.101.1.411.

Killeen, P. R., Hanson, S. J., & Osborne, S. R. (1978). Arousal: Its genesis and manifestation as response rate. *Psychological Review*, *85*(6), 571–581. https://doi.org/10.1037/0033-295X.85.6.571

Killeen, P. R., & Sitomer, M. T. (2003). MPR. *Behavioural Processes*, *62*, 49–64. https://doi.org/10.1016/S0376-6357(03)00017-2

Marwell, G., & Schmitt, D. R. (1975). Cooperation: An experimental analysis. New York, NY: Academic Press.

McAuliffe, K., Blake, P. R., Kim, G., Wrangham, R. W., & Warneken, F. (2013). Social influences on inequity aversion in children. *PLoS ONE*, *8*(12), 1-11. https://doi.org/10.1371/journal.pone.0080966

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, 1, 1-9. http://doi.org/10.1038/s41562-016-0042

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*, 1560–1563. https://doi.org/10.1126/science.1133755

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*, 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Rescorla, R. A., & Wagner, A. R. (1972). Theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II. New York, NY: Appleton-Century-Crofts.

Ribes-Iñesta, E., Rangel, N., Pulido L., Valdez U., Ramírez, E., Jiménez, C., & Hernández, M. (2010). Reciprocity of responding as a determinant of a partial-altruistic behavior in humans. *European Journal of Behavior Analysis*, *11*, 105-114. https://doi.org/10.1080/15021149.2010.11434337

Schmitt, D. R. (1998). Social behavior. In: K. Lattal & M. Perone (Eds.), Handbook of research method in human operant behavior. New York, NY: Plenum Press.

Selten, R., & Stoecker, R. (1986). End behaviour in sequences of finite prisoner's dilemma supergames: A learning theory approach. *Journal of Economic Behavior and Organization*, *7*, 47-70. https://doi.org/10.1016/0167-2681(86)90021-1

Shaw, A., Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: Genera*l, *141*, 382-395. Retrieved from https://europepmc.org/abstract/med/23882227

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

Silverstein, A., Cross, D., Brown, J., & Rachlin, H. (1998). Prior experience and patterning in a prisoner's dilemma game. *Journal of Behavioral Decision Making*, *11*, 123–138. https://doi.org/10.1002/(SICI)1099-0771(199806)11:2<123::AID-BDM283>3.0.CO;2-5

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In: J. H. Barkow, L. Cosmides L, & J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture. New York, NY: Oxford University Press.

Tooby, J., Cosmides, L., & Barrett, H. C. (2005). Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In: P. Carruthers, S. Laurence, & SD. Stich (Eds.), The innate mind: Structure and content. New York: Oxford University Press

Urcuioli, P. J. (2013). Stimulus control and stimulus class formation. In G. J. Madden (Ed.), APA handbook of behavior analysis, *Volume 1*. Washington, DC: APA.

BRIEF REPORT

AN ANALYSIS OF INDIVIDUAL ELEMENTS OF COMPOUND STIMULI

Live F. Braaten and Erik Arntzen

OSLO METROPOLITAN UNIVERSITY, NORWAY

Compound stimuli consist of two or more elements that can be separated and potentially control behavior individually (Stromer, McIlvane, & Serna, 1993). Even though multiple elements of stimuli are present in the environment when the three- or four-term contingencies are learned, not all aspects of the compound stimuli are necessarily a controlling part of the contingency. Lack of stimulus control by the different components of compound stimuli has been referred to as selective attention (e.g., Ploog, 2011; Ray, 1969), overselectivity (e.g., Dube & McIlvane, 1999; Schneider & Salzberg, 1982) or restricted stimulus control (e.g., Dube & McIlvane, 1997; Ribeiro et al., 2015; Stromer, McIlvane, & Serna, 1993). Experiments have shown that stimulus control often restricted in non-human animals (Born & Peterson, 1969; Reynolds, 1961) and humans diagnosed with developmental disabilities and autism (e.g., Dickson et al., 2006; Lovaas et al., 1971; Stromer, McIlvane, Dube, et al., 1993) when established with simple discrimination training. Lovaas et al. (1971) found that in children without developmental disabilities and autism, all aspects of the compound stimuli controlled responding when tested separately in a simple successive discrimination training procedure. Perez et al. (2015) found similar results in a simple simultaneous discrimination procedure with college students. Whereas, restricted stimulus control has been shown after conditional stimulus control has been established in matching-to-sample (MTS) training procedure in the same population (Braaten & Arntzen, 2019; Stromer & Stromer, 1990a).

Corresponding author: Live F. Braaten, Oslo Metropolitan University, Department of Behavioral Science, P O Box 4 St. Olavs Plass, 0130 Oslo, Norway. E-mail: livefb@oslomet.no

MTS training is an efficient procedure to establish conditional discriminations among stimuli. In this procedure, participants match one of several comparison stimuli to a sample stimulus. In MTS, the initial relation between the sample stimulus and the comparison stimuli can either be identical or arbitrary. In identity MTS, participants match stimuli that are identical or have a physical resemblance to each other. Whereas, in arbitrary MTS, the stimuli are different and do not have physical similarities. Based on the programmed consequences given in the arbitrary MTS procedure, participants learn specific four-term contingencies between stimuli, defined by the experimenter.

Braaten and Arntzen (2019) tested the preference for the individual elements of four different compound stimuli in adult participants after an identity MTS procedure. The compound stimuli in these experiments were made up of simple shapes superimposed on a colored background. In both experiments, many participants repeatedly responded, in a forced-choice test, to only one aspect of the compound stimuli. The element from the compound stimuli that controlled responding (color or shape) varied across participants. The uniform responding to one element might reflect that the chosen stimulus had acquired stimulus control and not both elements of the compound stimulus. A limitation in this study was the forced-choice set-up where participants had to choose between the two correct comparison stimuli. Hence, the authors suggested as a future experiment to establish conditional discriminations with abstract and compound stimuli in an arbitrary MTS format and to test each element of the compound stimuli individually. Such an experiment would test if adults show restricted stimulus control in MTS or if the result of the Braaten and Arntzen study was an artifact of the procedure (forced-choice).

In an MTS procedure, once participants have learned several arbitrarily related stimulus-stimulus relations, one can test for the formation of stimulus equivalence classes. Stimulus equivalence is verified by the emergence of novel relations between the sample and comparison stimuli. Sidman and Tailby (1982) described that when participants relationally respond to three untrained properties: reflexivity, symmetry, and transitivity, stimulus equivalence classes have emerged. Some experiments have used compound stimuli in arbitrary MTS procedures. These experiments have tested for emergent relations by separating and rearranging the compound stimuli in different ways (e.g., Stromer & Stromer, 1990a, 1990b). Stromer and Stromer (1990a) trained 18 college students in an MTS procedure using compound stimuli and tested for equivalence relations. The compound stimuli used in the experiment were tone and color stimuli presented at the same time. Stromer and Stromer reported that 14 out of 18 participants responded consistently with the two 5-member equivalence classes and four participants that did not. These results suggest that both elements of the compound stimuli do not control responding in arbitrary MTS for some participants. Stromer and Stromer (1990b) extended their procedure by training each component of the compound stimuli (tone and color) to abstract stimuli first, before training compound stimuli to a new abstract stimulus. The results showed that stimulus control by all elements of the compound stimuli was established for a higher number of participants (13 out of 14). Hence, these results might indicate that different manipulations of the MTS arrangement can affect stimulus control to compound stimuli.

The studies mentioned have investigated compound stimuli in arbitrary MTS in humans with compound sample stimuli and simple (one element) comparison stimuli. Hayashi and Vaidya (2008) investigated the effect of establishing control with compound stimuli as a sample stimuli, comparison stimuli, or both in an MTS procedure. In their conclusion, they suggested that conditional discriminations might be easier to establish if the sample stimulus is a simple stimulus and comparison stimuli are compound stimuli than the other way around. One way to investigate the compound stimuli's function as sample or comparison stimuli and test for emergent responding can be done by comparing One-to-Many (OTM) and Many-to-One (MTO) training structures. For example, in a three-member stimulus class where the compound stimulus is the nodal stimulus, the compound stimulus would be the sample stimulus in an OTM training structure. In the MTO training structure, the compound stimulus would be the comparison stimuli.

Investigating restricted stimulus control to elements of compound stimuli in arbitrary MTS and testing for emergent responding with an additional manipulation of the training structure in adult participants has, as far as the authors know, not yet been studied. Hayashi and Vaidya's (2008) results may predict that participants trained with an MTO training structure will use fewer training trials to learn conditional discriminations. However, they did not test for emergent relations, so we do not know how such manipulation would affect emergent responding. Saunders and Green (1999) argue that the number of simple discriminations learned in conditional discrimination training predicts the outcome of a test for stimulus equivalence relations. In an MTS procedure, participants are exposed to simple discrimination when learning conditional discriminations. Based on their analysis, the MTO training structure presents all simple discriminations between all stimuli in training, whereas the OTM structure does not. This discrepancy between the number of simple discriminations presented in training might result in higher yields in stimulus equivalence tests following the MTO training structure than the OTM training structure. Saunders and Green's assumption is based on single element stimuli and not compound stimuli.

The present experiment's primary purpose is to investigate restricted stimulus control with compound stimuli in adult participants in arbitrary MTS and emergent responding. To do so, participants will learn conditional discriminations with some compound stimuli in an arbitrary MTS procedure with a 0-s delay, followed by a test for equivalence class formation with elements of the compound stimuli tested individually. A secondary purpose is to investigate if the function of the compound stimuli either as a sample stimulus or comparison stimulus in the MTS procedure would affect potential restricted stimulus control. For this purpose, two groups, one trained with OTM training structures and one with MTO, will be compared.

METHOD

Participants

Twenty-six female and four male (age 19–53) university students participated in the experiment. Participants signed a consent form where they were informed in general terms about the experimental situation, their rights, and the experiment's approximated duration (one hour). All participants were shown their data and thoroughly debriefed about the research.

Setting

Two rooms were used to conduct the experiment. One room was $13m^2$, had two windows covered with blinds, and furnished with chairs and tables. The second room was without windows and organized with dividing walls creating two small cubicles. Each cubical was 2.7 m^2 and equipped with one table and one chair

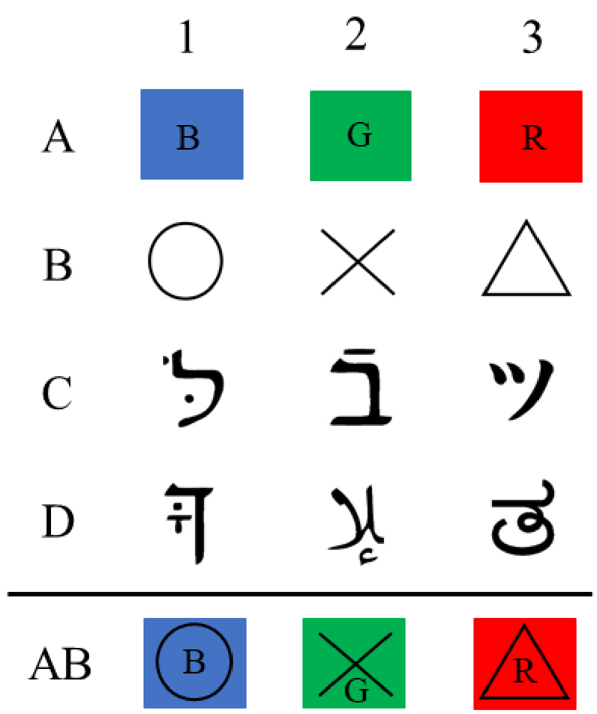
Apparatus and Stimuli

Participants were trained and tested on laptop computers with Windows 8 operating system and connected with an external mouse. One computer had a 15.6-inch screen, and the other one had a 17-inch screen. Custom-made matching-to-sample software ran the training and testing, controlled stimuli presentation, and registered participants' responses.

The same set of stimuli were used in both groups (see Figure 1) and consisted of 12 stimuli, potentially partitioned into three classes. The stimuli-set contained three color stimuli (A), three shape stimuli (B), and six abstract stimuli (C and D). In training, one compound stimulus was trained to two abstract stimuli. The compound stimuli consisted of shape stimuli superimposed on color stimuli AB (see Figure 1, the letters indicate which background color each stimulus had, and were not on the stimuli). The three compound stimuli were a circle on a blue (B) background, a cross on a green (G) background, and a triangle on a red (R) background. In the test, elements of each compound stimuli were separated and presented individually (stimuli A and B in Figure 1). All stimuli were approximately 5 cm x 5 cm on the screen.

The participants were asked to sort laminated printouts of the stimuli before training to ensure that they were not familiar with the experimenter-defined stimulus classes.

Figure 1
Experimental Stimuli in Both Conditions



Note. The stimuli used in training and testing. The letters on the left side denote class members, and the number on the top denotes the class. AB stimuli on the bottom are the compound stimuli made up of A and B stimuli merged on top of each other. The letters on the A and AB stimuli represents the color of the stimuli, and were not on the stimuli used in the experiment. B=blue, G=green, and R=red.

Design

Participants were randomly assigned into two groups. In one group, participants were trained with an OTM training structure, hereafter called OTM-group. Here, the compound stimuli were the nodal stimuli, always presented as sample stimuli. Participants in the other group were trained with an MTO training structure, hereafter called MTO-group. Again, the compound stimuli were nodal stimuli, but the compound stimuli served as comparison stimuli in this condition.

Procedure

Instruction

Participants were seated in front of a computer and presented with the following instruction on the screen (translated from Norwegian):

"A stimulus will appear on the screen. You must click on this with the mouse. Three stimuli will appear. Select one of these by clicking the mouse. If you choose the one we have defined as correct, words like "good," "super," etc., will appear on the screen. If you press incorrectly, "wrong" will display on the screen. During the experiment, the computer will not provide feedback on whether your choices are correct or incorrect, but based on what you've learned, you can get all the tasks right. Do your best to get everything correct. Good luck!"

To advance to training, participants had to press the "START" button below the instruction.

Training

The purpose of the training phases was to establish six conditional discriminations. Participants in the OTM-group were taught A1B1-C1, A2B2-C2, A3B3-C3, A1B1-D1, A2B2-D2, A3B3-D3 relations, whereas participants in the MTO-group learned C1-A1B1, C2-A2B2, C3-A3B3, D1-A1B1, D2-A2B2, D3-A3B3 relations. Each trial began with a sample stimulus in the middle of the screen, and after a mouse click on the sample stimulus, it disappeared, and three comparison stimuli appeared on the screen with a 0-s delay. Here, participants had to choose one of three comparison stimuli. If they chose the experimenter-defined correct comparison, words like "correct," "good," etc., appeared in the middle of the screen. If they chose the experimenter-defined wrong comparison, the word "wrong" appeared on the screen. The programmed consequences were on the screen for 1 s, followed by a 0.5 s intertrial interval before the next trial. For each trial, the three comparison stimuli appeared randomly in the four corners of the screen. The baseline relations were established concurrently in blocks of 30 trials. Each relation was presented five times in a block and in random order. After one block with 90% correct or more, the probability of programmed consequences was reduced to 75%, 25%, and then 0% in the consecutive blocks. If the mastery criterion of 90% correct was not reached, the last block was repeated.

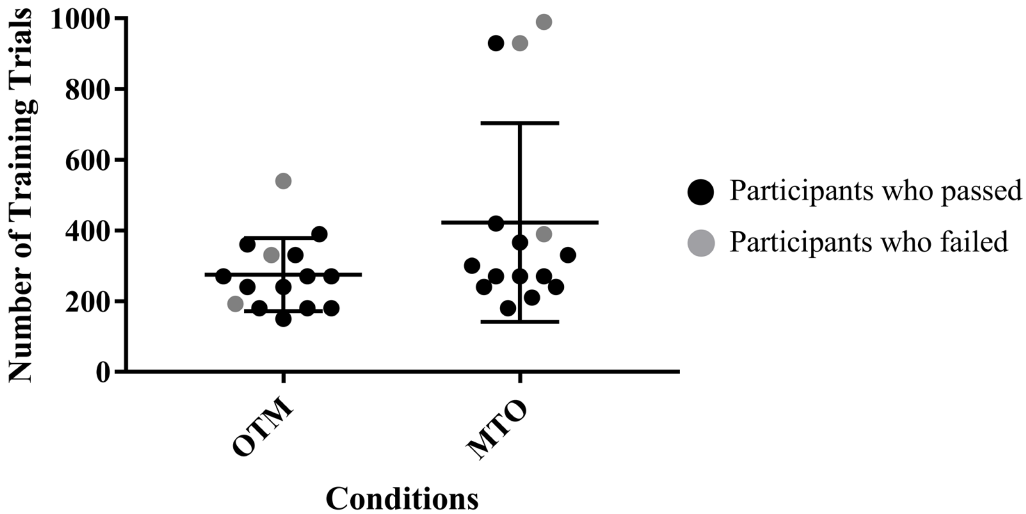
Testing

Testing Phase 1. The first testing phase's purpose was to test for baseline (BSL) and symmetry (SYM) relations with the compound stimuli. In this test, all compound stimuli were as in training, made up of shape and color. The test trials had the same set-up as in training. The test consisted of 60 trials: 30 BSL trials same as in training, and 30 SYM trials: C1-A1B1, C2-A2B2, C2-A3B3, D1-A1B2, D2-A2B2, D3-A3B3 for the OTM-group, and A1B1-C1, A2B2-C2, A3B3-C3, A1B1-D1, A2B2-D2, A3B3-D3 for the MTO-group. All trials were presented in random order.

Testing Phase 2. The purpose of the second testing phase was to test BSL, SYM, and equivalence (EQ) relations presenting each element of the compound stimuli at the time to evaluate if one part of the compound stimuli exerted more control of responding than the other. The second test phase started immediately after the first test, independently of the performance in Test 1. In Test 2, compound stimuli were separated, and each element was presented individually on the screen in each trial, shown as stimuli A and B in Figure 1. All participants, regardless of training structure, were tested for 30 trials similar to BSL relations with only color stimuli (BSL-C), 30 trials similar to BSL relations with only shape stimuli (BSL-S), 30 trials similar to SYM with only color stimuli (SYM-C), 30 trials similar to SYM with only shape stimuli (SYM-S), and 30 EQ trials: C1D1, C2D2, C3D3, D1C1, D2C2, D3C3 for both groups.

Figure 2

Number of Training Trials in Each Condition



Note. The black dots represent participants who passed Test 1, and the grey dots represent participants who failed Test 1 in the OTM and MTO condition. OTM=One-to-many, MTO=one-to-many.

In total, there were 150 test trials. The set-up was the same as in training and Test 1.

Mastery Criterion in Test 1 and Test 2. The mastery criterion for both tests was 95% correct for each relation. Participants with mastery below 95 percent in either baseline or symmetry relations on Test 1 were excluded from the experiment's last part. The criterion was set to exclude participants that did not establish the stimulus classes with the compound stimuli in Test 2. This way, to a higher degree of certainty, one could conclude that any incorrect responding in Test 2 was due to the separation of the compound stimuli.

RESULTS

Fifteen participants in each group finished the experiment with an average of 275 training trials in the OTM condition (range=150–540) and an average of 422 training trials in the MTO condition (range=180–990). Figure 2 displays the number of training trials for each participant. The black and grey circles denote those participants who passed and failed Test 1, respectively. An independent-samples t-test was conducted to compare the number of training trials in the OTM and MTO conditions. The test showed no significant difference in the number of training trials for the OTM (M=274.8, SD=103.1) and the MTO (M=422.4, SD=280.9) conditions; p=0.0664.

Results from Test 1 and 2 for participants in both groups are shown in Table 1. Here, performance above the criterion (95% correct) is written in bold. When presented with BSL and SYM test trials with the compound stimuli, three participants in each group did not reach criterion in one or both relations. These participants were excluded from further analysis. Eight participants in the OTM-group passed Test 2, and seven participants passed in the MTO-group. Fisher's Exact Test indicate a non-significant difference in test outcome on Test 2 between the groups (p=1). The four participants who did not reach the criterion in the OTM-group responded incorrectly on trials testing the color or the shape aspect of the compound stimuli. P17185 had incorrect responses when presented with the color stimuli in BSL trials, whereas P17171 responded incorrectly when the shape stimuli were presented in BSL trials. P17178 had a total of 42

Table 1
Overview of Results

OTM

Participant #	Test 1 - Compound BSL	Test 1 - Compound SYM	Test 2 - Compund Separated BSL Color	BSL Shape	SYM color	SYM shape	EQ
17151	**30**	**30**	**30**	**29**	**29**	**30**	**30**
17159	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17160	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17162	**30**	**30**	**30**	**29**	**30**	**30**	**30**
17167	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17180	**29**	**30**	**30**	**30**	**30**	**30**	**30**
17181	**30**	**30**	**29**	**30**	**30**	**30**	**29**
17170	**30**	**30**	**30**	**29**	**30**	**30**	**29**
17158	**30**	**29**	28	**30**	**30**	**30**	**29**
17171	**29**	**30**	**30**	28	**30**	**30**	**30**
17178	**29**	**30**	8	**30**	10	**29**	**30**
17172	**29**	**30**	27	28	28	28	28
17154	**30**	27					
17175	**30**	27					
17164	28	28					

MTO

Participant #	Test 1 - Compound BSL	Test 1 - Compound SYM	Test 2 - Compund Separated BSL Color	BSL Shape	SYM color	SYM shape	EQ
17152	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17163	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17166	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17177	**30**	**30**	**30**	**30**	**30**	**29**	**29**
17168	**30**	**30**	**30**	**30**	**29**	**30**	**30**
17173	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17179	**30**	**30**	**30**	**30**	**30**	**30**	**30**
17161	**30**	**29**	**30**	**30**	**30**	27	**30**
17169	**30**	**30**	**30**	27	**30**	28	**29**
17156	**29**	**30**	**29**	26	**29**	28	28
17165	**30**	**30**	**30**	10	**30**	10	**30**
17176	**29**	**29**	**30**	28	**30**	28	**30**
17157	28	**29**					
17153	26	25					
17155	24	19					

Note. The table shows the number of correct responses made in Test 1 and Test 2 for each relation. Performance above the mastery criterion (95%) is in bold. The three last participants in each group did not meet the criterion in Test 1 and did not advance to Test 2. BSL=baseline, SYM=symmetry, EQ=equivalence.

incorrect trials in Test 2. All incorrect trials were made regarding color stimuli, both in trials testing BSL and SYM relations. P17172 responded below the criterion in all relations tested, including EQ relations.

All five participants who did not reach the mastery criterion in the MTO group responded below the criterion in trials testing shape stimuli. P 17161 responded only incorrectly in SYM-S trials, where all the others did so in both BSL-S and SYM-S trials. All of those had two to four incorrect trials, whereas P17165 had 40 incorrect trials in both relations testing with shape stimuli. P17156 was the only participant that did not reach the criterion on EQ trials.

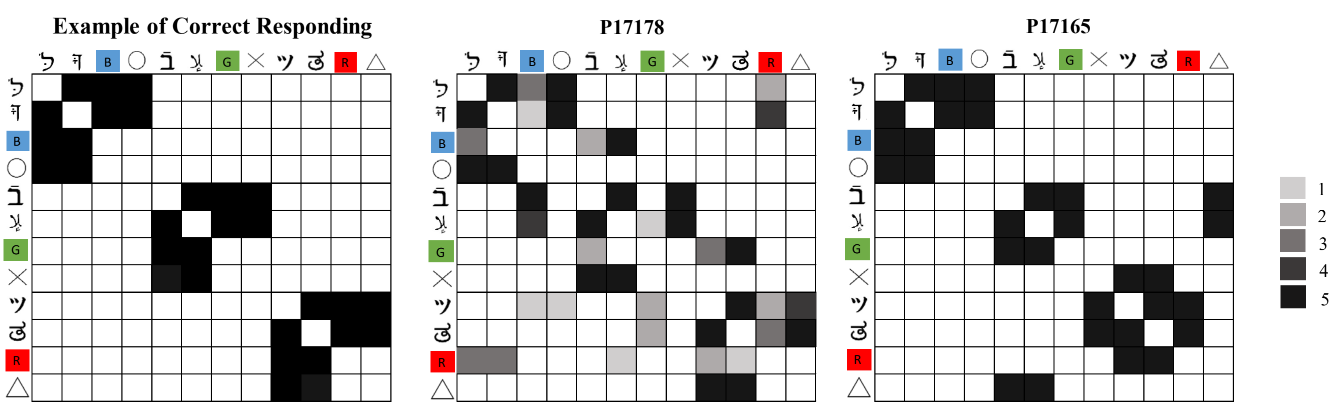
The test performance of the two participants with the most incorrect trials in Test 2 is displayed as a response matrix in Figure 3. Here, the vertical stimuli on the left side of the matrix represent sample stimuli, and the horizontal stimuli on the top of the matrix represent chosen compound stimuli. The different colors, from light grey to black, represent the number of responses illustrated on the right side of the figure. The left matrix exemplifies how the matrix would look if a participant responded correctly to all trials. The responses of P17178 in the OTM group are shown in the middle matrix. This participant had 22 incorrect test trials in BSL-C and 20 incorrect trials for SYM-C relations, and the incorrect responses to color stimuli were random without any pattern of responding. This response patten show a lack of stimulus control to all color stimuli. The right matrix shows responses of P17165 from the MTO group. This participant had 20 incorrect trials in both BSL-S and SYM-S relations, and systematically responded to the cross when the triangle was correct and vice versa. This response pattern shows participant-defined stimulus control.

DISCUSSION

In both groups, 12 out of 15 participants responded within the criterion on baseline and symmetry relations in Test 1. Thought, when exposed to Test 2, where the compound stimuli were separated, 33.3% of participants in the OTM-group and 42% in the MTO-group did not reach the test's mastery criterion (95%). For these participants, both elements of the compound stimuli did not control responding.

Figure 3

Response Matrix for Two Participants



Note. The response matrix display responses made in Test 2. Stimuli on the left side of each matrix represent sample stimuli, and the stimuli on the top represent the comparison stimuli chosen by the participants. The matrix on the left is an example of 100 % correct responding. The different colors illustrate the number of responses made for each relation, labeled on the right side.

In this experiment, one might have expected a higher number of participants reaching the criterion in Test 2 for several reasons. Firstly, OTM and MTO training structures often result in high yields in stimulus equivalence relations (Arntzen, 2012). Also, variables such as small and few classes (Arntzen & Holth, 2000), meaningful stimuli (Fields & Arntzen, 2018), and training and testing with a 0-s delay (Arntzen, 2006; Bortoloti & de Rose, 2009) generally increase the probability of responding in accordance with stimulus equivalence. Lastly, stimuli used in the three compound stimuli are very familiar to the participants. Therefore, it is surprising and interesting that more than one-third of the total number of participants responded incorrectly when the compound was separated, and the elements were tested individually. The present experiment results differ from Lovaas et al. (1971) and Perez et al. (2015) that did not show restricted stimulus control in healthy children or adults, respectively. On the other side, the results support Braaten and Arntzen (2019) and Stromer and Stromer (1990a) that some adult participants show restricted stimulus control.

The present results show no statistical difference between the MTO and the OTM groups regarding equivalence class formation in Test 2. These results oppose Saunders and Green's simple discrimination analysis (1999), which predicts higher yields in equivalence class formation following training with an MTO training structure than an OTM training structure. Saunders and Green's discrimination analysis are based on simultaneous matching between sample and comparison stimuli. In the present experiment, a 0-s delay between the offset of the sample stimulus and the onset of the comparison stimuli was used. A delayed MTS procedure creates successive discriminations instead of simultaneous discrimination between the sample stimulus and the comparison stimuli. Saunders and Green do not discuss this variance of the procedure or how this would affect the outcome. Though, they do argue that simple successive discriminations are more difficult than simple simultaneous discrimination. Saunders and Green emphasize that the discrepancy between the number of simple discriminations embedded in the training structures increases when the class size and number of classes increase, leading to a greater difference in outcome between the two training structures. They also write that when training with few and small classes, differences between outcome might not be as evident (p.129), which might be the case for the present experiment.

There are differences between the two groups in the present experiment regarding participants' responding to the compound stimuli' elements for those who failed Test 2. Participants in the MTO-group only responded incorrectly to shape stimuli, not color. Whereas in the OTM-group, participants responded incorrectly to both color and shape stimuli. The main difference between the two conditions is that the compound stimuli serve as sample stimuli in the OTM training structure and as comparison stimuli in the MTO training structure. Thus, the compound sample stimuli are successively discriminated from each other in OTM, and the compound comparison stimuli are simultaneously discriminated from each other in MTO. Arguably, presenting all the compound stimuli on the screen together, the color stimuli are the most immediate visually discriminable feature of the compound stimuli. All the shape stimuli are black and, though familiar, maybe not visually impactful. The present results indicate that simultaneous discrimination of the compound stimuli might have resulted in an increased probability of stimulus control to the more outstanding or salient part of the compound stimuli, the color. Contrary, when compound stimuli are successively discriminated as sample stimuli, the compound stimuli are not pitted against one another. Hence, whether participants responding are under control of the shape or color might be the result of participants' preference (learning history) for the color or shape of that particular compound stimuli and not because of a comparison of the compound stimuli as sample stimuli, resulting in more variation in what aspect of the compound stimuli controlled behavior. Future research could vary the compound stimuli functions and the stimuli that compose the compound stimuli to investigate this potential effect of simultaneous and successive discrimination on restricted stimulus control. Also, one could include more complex or unfamiliar stimuli as the compound stimuli.

The BSL and SYM relations in Test 2 are of most interest in terms of evaluating restricted stimulus control. However, the equivalence trials are interesting to assess whether participants fully formed equivalence classes. Only P17172 in OTM-group and P17156 in MTO-group did not reach the criterion in EQ trials. They had incorrect trials in other relations tested, see Table 1. Both responded correctly in Test 1, which shows that the separation of the compound stimuli disrupted class formation. All the other participants establish 3 three- or four-member equivalence classes with one or both aspects of the compound stimuli, respectively, as a part of the class.

Two participants stand out due to a high number of incorrect trials in Test 2, illustrated in Figure 3. P17165, in the MTO group, had 20 incorrect trials. This participant systematically responded to the cross when the triangle was correct and the triangle when the cross was correct. Such a pattern of responding is an example of participant-defined classes as opposed to experimenter-defined classes. The participant responded correctly to the circle stimuli, as defined by the experimenter. In the OTM-group, P17178 had 23 incorrect trials, mostly to color stimuli. This participant responded incorrectly to all colors indicating a general lack of stimulus control and not participant-defined class formation. Individual differences as to what aspect of the compound controls responding have been shown in pigeons (Reynolds, 1961) and humans (Braaten & Arntzen, 2019).

The results from this experiment show that participants used, on average, approximately 50% more training trials to learn the six conditional discriminations with an MTO training structure compared to the OTM training structure. This result was not significant, though it might indicate that learning conditional discriminations with an MTO training structure with compound stimuli as comparison stimuli were more challenging. Hayashi and Vaidya (2008) argued that discriminability is a more critical feature then complexity. "…(T)he stimuli that are more readily discriminated should be positioned as the sample and those less readily discriminated as the comparison stimuli" (p.182). In the present experiment, the simple stimuli were abstract shapes unfamiliar to the participants (see Figure 1), and the compound stimuli were well-known shapes and colors. In terms of discriminability, it is difficult to conclude that the compound stimuli in the present experiment are more difficult to discriminate than the abstract, unfamiliar shape; actually, it might be the opposite. Participants have a long history with squares, triangles, and circles and the colors; blue, green, and red, making those stimuli potentially easier to discriminate than the abstract stimuli. Therefore, it is challenging to draw any conclusions on whether current results support or oppose Hayashi and Vaidya.

Finally, the present results have valuable contributions by elucidating that restricted stimulus control occurs under specific conditions in conditional discrimination procedures in adult humans without a diagnosis, which has practical implications that should be considered when establishing stimulus control to complex stimuli. Simultaneously, this experiment shows that small manipulations of the MTS procedure and a fine-grained analysis can increase knowledge regarding the stimulus function in a four-term contingency and the role of simple simultaneous and successive discriminations in conditioned discriminations and stimulus equivalence class formation.

REFERENCES

Arntzen, E. (2006). Delayed matching to sample: Probability of responding in accord with equivalence as a function of different delays. The Psychological Record, *56*(1), 135–167. https://doi.org/10.1007/BF03395541

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., & Holth, P. (2000). Probability of stimulus equivalence as a function of class size vs. number of classes. *The Psychological Record*, *50*(1), 79–104. https://doi.org/10.1007/BF03395344

Born, D. G., & Peterson, J. L. (1969). Stimulus control aquired by components of two color-form compound stimuli. *Journal of the Experimental Analysis of Behavior*, *12*(3), 437–442. https://doi.org/10.1901/jeab.1969.12-437

Bortoloti, R., & de Rose, J. C. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential. *The Psychological Record*, *59*(4), 563–590. https://doi.org/10.1007/BF03395682

Braaten, L. F., & Arntzen, E. (2019). On the role of compound stimuli in a matching-to-sample procedure. *Behavior Analysis: Research and Practice*, *20*(1), 58–70. https://doi.org/10.1037/bar0000167

Dickson, C. A., Wang, S. S., Lombard, K. M., & Dube, W. V. (2006). Overselective stimulus control in residential school students with intellectual disabilities. *Research in Developmental Disabilities*, *27*(6), 618–631. https://doi.org/10.1016/j.ridd.2005.07.004

Dube, W. V., & McIlvane, W. J. (1997). Reinforcer frequency and restricted stimulus control. *Journal of the Experimental Analysis of Behavior*, *68*(3), 303–316. https://doi.org/10.1901/jeab.1997.68-303

Dube, W. V., & McIlvane, W. J. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis*, *32*(1), 25–33. https://doi.org/10.1901/jaba.1999.32-25

Fields, L., & Arntzen, E. (2018). Meaningful stimuli and the enhancement of equivalence class formation. *Perspectives on Behavioral Science*, *41*(1), 69–93. https://doi.org/10.1007/s40614-017-0134-5

Hayashi, Y., & Vaidya, M. (2008). Effects of stimulus discriminability on the acquisition of conditional discriminations in adult humans. *European Journal of Behavior Analysis*, *9*(2), 173–183. https://doi.org/10.1080/15021149.2008.11434303

Lovaas, O. I., Schreibman, L., Koegel, R., & Rehm, R. (1971). Selective responding by autistic children to multiple sensory input. *Journal of Abnormal Psychology*, *77*(3), 211–222. https://doi.org/10.1037/h0031015

Perez, W. F., Endemann, P., Pessôa, C. V. B. B., & Tomanari, G. Y. (2015). Assessing stimulus control in a discrimination task with compound stimuli: Evaluating testing procedures and tracking eye fixations. *The Psychological Record*, *65*(1), 83–88. https://doi.org/10.1007/s40732-014-0092-1

Ploog, B. O. (2011). Selective attention to visual compound stimuli in squirrel monkeys (Saimiri sciureus). *Behavioural Processes*, *87*(1), 115–124. https://doi.org/10.1016/j.beproc.2010.12.015

Ray, B. A. (1969). Selective attention: The effects of combining stimuli which control incompatible behavior. *Journal of the Experimental Analysis of Behavior*, *12*(4), 539–550. https://doi.org/10.1901/jeab.1969.12-539

Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior*, *4*(3), 203–208. https://doi.org/10.1901/jeab.1961.4-203

Ribeiro, D. M., Miguel, C. F., & Goyos, C. (2015). The effects of listener training on discriminative control by elements of compound stimuli in children with disabilities. *Journal of the Experimental Analysis of Behavior*, *104*(1), 48–62. https://doi.org/10.1002/jeab.161

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior*, *72*(1), 117–137. https://doi.org/10.1901/jeab.1999.72-117

Schneider, H. C., & Salzberg, C. S. (1982). Stimulus overselectivity in matching-to-sample paradigm by severely retarded youth. *Analysis and Intervention in Developmental Disabilities*, *2*(4), 273–304. https://doi.org/10.1016/0270-4684(82)90025-8

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

Stromer, R., McIlvane, W. J., Dube, W. V., & Mackay, H. A. (1993). Assessing control by elements of complex stimuli in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, *59*(1), 83–102. https://doi.org/10.1901/jeab.1993.59-83

Stromer, R., McIlvane, W. J., & Serna, R. W. (1993). Complex stimulus control and equivalence. *The Psychological Record*, *43*(4), 585–598. https://doi.org/10.1007/BF03395901

Stromer, R., & Stromer, J. B. (1990a). The formation of arbitrary stimulus classes in matching to complex samples. *The Psychological Record*, *40*(1), 51–66. https://doi.org/10.1007/BF03399571

Stromer, R., & Stromer, J. B. (1990b). Matching to complex samples: Further study of arbitrary stimulus classes. *The Psychological Record*, *40*(4), 505–516. https://doi.org/10.1007/BF03399537

BRIEF REPORT

AN EXPLORATORY ASSESSMENT OF HUMAN TOKEN ACCUMULATION

Sean Regnier, Nicholas Van Zandt, and Anthony DeFulio

WESTERN MICHIGAN UNIVERSITY

The accumulation of reinforcers is prevalent in humans and non-human animals. For example, some animals store food, and many more engage in central place foraging, described in models of foraging such as the marginal value theorem (Charnov, 1976), which models the relationship between travel cost and time spent in patches. Humans also accumulate reinforcers, including conditioned reinforcers, such as money, but also others like beanie-babies and toilet paper.

Another example of conditioned reinforcers that are commonly accumulated are tokens. The tokens provided in token economies can serve as a bridge between a response and a reinforcer (Kazdin & Bootzin, 1972), making behavior less sensitive to delays to terminal reinforcement. Token accumulation research has often focused on the manipulation of the (a) token production schedule and (b) the exchange production schedule. This typically involves altering the relevant response requirement. Specifically, the token production schedule is the response requirement for producing tokens, and the exchange production schedule is the response requirement for producing the opportunity to exchange tokens. In a study by Yankelevitz, Bullock, and Hackenberg (2008), pigeons' key pecks were reinforced by delivery of tokens exchangeable for food. The token production schedule varied between fixed-ratio (FR) 1 and FR-10. Pecks to a separate key initiated an exchange period. The exchange production schedule ranged from FR-1 to FR-250. The authors found that accumulation was a combined function of the token production and the exchange production schedules. Token accumulation was positively correlated with increases to the exchange production schedule, and negatively correlated to the token production schedule.

The exchange production schedule is often described as a type of travel or procurement cost (Charnov, 1976; Hackenberg, 2018) and its effect on accumulation is well-documented. For example, Killeen (1974) measured the effects of travel distance between the lever and food dispenser on the accumulation of pellets before consuming them. Rats' lever presses were reinforced with food on an FR-1 schedule. Distance between the food dispenser and the response lever was increased across conditions. Killeen found that as the distance increased, the number of times the rats pressed the lever before traveling to the pellets increased. McFarland & Lattal (2001) conducted a similar study in which they manipulated the FR food schedule and the distance between the earn and collect lever. Overall, accumulation was highest when the earn and collect levers were furthest apart, and the FR food schedule was at its lowest value. This result is consistent with the effects of distance on accumulation (Killeen, 1974), and the effects of the exchange and token production schedules on accumulation (Yankelevitz et al, 2008.)

Over the past ten years, accumulation research has been extended to applied behavior analysis and behavior therapy. This line of research has focused on the preference for accumulated, delayed terminal reinforcers compared to immediate, distributed reinforcers. DeLeon et al. (2014) measured the efficacy of distributed and accumulated backup reinforcers on task completion by increasing the exchange production schedule in a token economy. Response rates were highest when participants were given access to accumulated reinforcers (i.e. several minutes of access to a video game) contingent on multiple response requirements rather than shorter access (i.e. thirty seconds) contingent upon a single response requirement. In addition to increased efficacy, participants also preferred the former over the latter when the reinforcer was an activity (4/4 participants) and an edible (3/4 participants). Participants typically engage in fewer problematic behaviors during accumulated reinforcer conditions as well (Fulton et al., 2020; Robinson & Peter, 2019) and are often successful in skill acquisition programs (Frank-Crawford et al., 2019). Another study assessed the extent to which preferences for larger exchange production schedules was moderated by the token production schedule (Falligant et al., 2020). Preference for larger exchange production schedules was higher during dense token production schedules, but reversed as the token production schedule was increased. These findings are consistent with studies of pigeons in which increases to token production and exchange production schedules had opposing effects on accumulation (Yankelevitz et al., 2008).

Another variable that affects how participants behave in a token economy is token generalizability. Generalizability can be manipulated by varying the number and type of back-up reinforcers available for each token during the exchange period. According to Skinner (1953) behavior maintained by a generalized reinforcer is likely to be under the control of multiple states of deprivation. For example, if a student were to earn a token that can be exchanged exclusively for potato chips, then a motivational operation that relates specifically to potato chips is required for the tokens to serve as effective reinforcers. However, if the student can exchange tokens for a large menu of items, then a much broader set of motivational operations will support the efficacy of tokens as reinforcers.

In one of the first demonstrations of generalized token efficacy, DeFulio et al. (2014) assessed the reinforcing value of three different types of tokens under conditions of water deprivation: food tokens (exchangeable for only food), water tokens (exchangeable for only water), and generalized tokens (exchangeable for food or water) with pigeons. Subjects produced more generalized than specific tokens across several increasing token production schedules, demonstrating a higher reinforcing efficacy for generalized tokens. A similar study extended the procedure of DeFulio et al. by measuring effects of increased token production requirements on the production of generalized and specific reinforcers. As the price of tokens exchangeable for specific reinforcers (food and water) increased, the production of generalized tokens (exchangeable for food or water) increased, demonstrating that generalized token reinforcers are substitutes for specific token reinforcers (Andrade & Hackenberg, 2017). Tan & Hackenberg (2015) used a similar arrangement to assess the efficacy of generalized token reinforcers with progressive ratio schedules, and preference procedures, and by manipulating response requirements to generate a demand function. This study similarly illustrated the substitutability of generalized tokens with specific food and water tokens.

A key gap in human reinforcer accumulation research is that the effects of generalizability on accumulation are not well understood. At most, previous research on humans has shown how preferences for accumulated reinforcers are affected by one of the token component schedules. Therefore, the present experiment was designed to investigate determinants of human token accumulation. Most importantly, token generalizability was manipulated across conditions by increasing the variety of goods that could be purchased with tokens. In addition, the exchange production schedule was manipulated by increasing the distance between the computer that delivered the token production task and the store where tokens were exchanged for other items. The token production schedule was also manipulated.

METHOD

Subjects

Five undergraduate students at a large midwestern university were recruited to participate in this study. Subjects were eligible to join this study if they were (1) at least 18 years old; (2) able to use a computer and be able to perform simple math; and (3) were not colorblind.

Subjects were excluded if they are (1) are suspected of being under the influence of recreational drugs or alcohol before, during, or immediately after any session; (2) had known allergies to any items included in the token exchange center.

Apparatus

A scale was used to weight all food items to the gram. Hot coffee and chocolate were measured using a programmable single cup coffee maker set at 4 ounces. All food items were distributed using paper plates, and sports drinks were poured into a plastic cup.

Procedure

Design

The study featured a single subject, repeated measures design. Each participant received all conditions. Participants completed the study over a minimum of seven and maximum of nine sessions, with no more than one session conducted per day.

Overview of Procedure

Subjects earned tokens on a computerized paint by number task. Completing a full screen of mathematical problems produced one token. Token production requirements were manipulated by altering the number of problems on each screen. All mathematical problems were simple addition, adding two numbers between 0 and 9 (excluding 0 + 0). Each problem was contained in a box on the screen. To fill each box, participants selected a color from an array on the left side of the screen which they dragged into the box using the computer mouse. There were six colors in the array, thus each color represented a three-value range of answers to the problem (e.g., the values 1 to 3 were represented by the color teal, while 4 to 6 was represented by green). In the present experiment, for each paint by number task, participants were asked to complete 100, 200, or 300 paint by number mathematical problems per screen, depending on the condition. The number of tokens participants accumulated were displayed as a running tally in top left corner of the paint by number screen. Participants were informed token production requirements, distance to the store, and menu items available for purchase at the beginning of each session. The menu was placed next to the participant throughout each session. While working on the paint by number task participants could pause at any time to exchange their tokens. All sessions lasted one hour, not including exchange periods. When a participant decided to pause their session to make an exchange, the researcher paused the one-hour timer. Timer pausing was designed to prevent travel time from affecting session duration. There was no time limit or requirement for the token exchange. Each subject could consume back-up reinforcers at any time during the experimental session, including while working on the task. The session timer was not paused for consumption. The token production schedule, exchange production schedule, and generalizability were manipulated as described below.

Phase 1: Token Production Schedule Manipulation

The FR token production schedule began at FR-100 and was increased by 100 responses across sessions, to a maximum of FR-300. Exchange production schedules and generalizability were held constant at their lowest values across these sessions. All subjects started on an FR-100 which increased each subsequent session. Any participant who failed to accumulate tokens at FR-100 or FR-200 was moved immediately to the next experimental phase instead of experiencing higher token production schedules.

Phase 2: Exchange Production Schedule Manipulation

The exchange production schedule was manipulated by increasing the walking distance required to exchange tokens. One distance was used per session. Token exchange centers were at the following locations: (a) next to the participant's work space, in the same room as the paint-by-number game (labeled the "No Walk" condition); (b) in the opposite corner of the experimental room (participants will have to stand up and move approximately 3 m to exchange their tokens, labeled the "Short Walk" condition); (c) across the hall, approximately 10 m to another room, labeled the "Long Walk" condition. The token production schedule and token generalizability were held constant during all exchange production schedule manipulations.

Phase 3: Token Generalizability Manipulation

For this phase token production and exchange production values were set based on the results of the prior phases. Specifically, values were selected such that the minimum amount of accumulation would be expected. These values (FR-200 token production and "No walk") were held constant across the three session of this phase. For the first session tokens could be exchanged for any of eight different snacks. Seven different kinds of salty chips (e.g., potato chips, Doritos®) were included on the menu, along with Welch's® fruit snacks. The second session included 15 items. Eight of these were identical to the eight offered in the previous session. The additional items included chocolate, breakfast cookies, juice, a sports drink, fruit-flavored candy, popcorn, and cheese-flavored crackers. The final condition included 19 items on the menu. Trail mix, beef jerky, coffee, hot chocolate were added to the 15 items included in the prior session.

Data Analysis

The primary outcomes in this study were the number of tokens spent and number of tokens available at each exchange period. The number of tokens available was measured to account for participants that do not spend all available tokens during a given exchange. Mean tokens spent and % multiple exchanges were calculated and averaged across all participants. Mean tokens spent was calculated by averaging the number of tokens spent for each condition, for each participant. A multiple exchange was any instance in which a participant spent more than one token during an exchange period.

RESULTS

Figure 1 contains the mean number of tokens available, tokens spent, and the percentage of exchanges during each condition where the participant spent more than one token. Excluding TA06, participants made multi-token exchanges at least 50% of the time across all conditions. For example, 100% of TA02's exchanges were of at least two tokens across all experimental conditions. However, multiple exchanges were more sensitive to manipulations for participants who spent their tokens more frequently (i.e. TA05 & TA06).

In general, participants' mean tokens available was more sensitive to all three manipulations than mean tokens spent. Three of five participants' tokens available decreased as the token production schedule increased. However, one participant showed a reverse trend. Excluding TA08, whose accumulation was insensitive to all manipulations, mean tokens available increased as the exchange production schedule increased. Tokens spent increased for three participants as exchange production schedule increased. During exchange production manipulations accumulation was highest in the Long Walk condition.

Mean tokens spent and available increased as generalizability increased for two of five participants (TA06, TA07). Three of five participants increased the number of multi-token exchanges as generalizability increased. The remaining two participants made multi-token exchanges across all generalizability manipulations. Overall, accumulation was highest, and multi-token exchanges were most prevalent, in the 19-item menu condition.

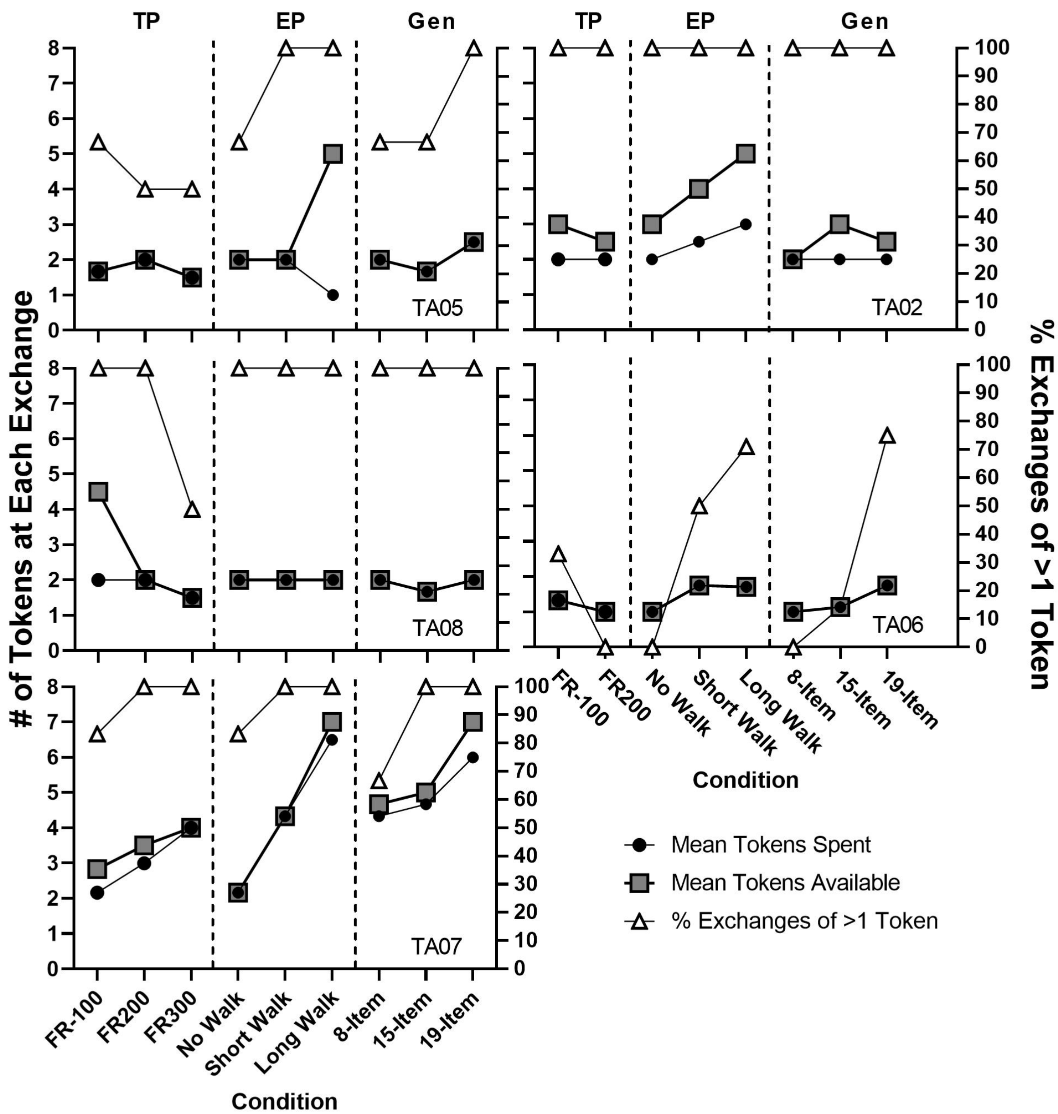
DISCUSSION

In this study, token accumulation was primarily a product of the exchange production schedule. Three of five participants had an immediate increase in accumulation when moving from the Short Walk to the Long Walk condition, while four of five participants increased their accumulation by the Long Walk condition. The effects of generalizability on accumulation were not substantial. Two participants increased their accumulation as generalizability was increased, with the largest increase in the token 19-item condition. Overall, participants' accumulation was unaffected by the token production schedule. One participant (TA08) was insensitive to all experimental contingencies.

The positive relationship between exchange production schedule and accumulation observed in this study is consistent with previous research on reinforcer accumulation (Yankelevitz et al., 2008; Killeen, 1974). The method used to manipulate the exchange production response requirement in the present study was similar to the method used by Killeen, though Killen's procedure did not incorporate tokens. In contrast, the present study and the Yankelevitz study both featured token economies, but differ-

Figure 1

Token accumulation across all conditions.



Note. Primary outcome variables for all study participants across all conditions. The left y-axis corresponds to the data paths with circle and square symbols, which represent mean tokens spent and mean tokens available for each participant, respectively. The right y-axis corresponds to the data path with triangle symbols and shows the percent of all exchanges in which more than one token was exchanged

ed in the method used for manipulating the exchange production response requirement.

The lack of relationship between token production schedule and accumulation is not consistent with prior research. Yankelevitz et al. (2008) found an orderly decrease in accumulation as the token production schedule increased from an FR-1 to an FR-10. In the present study, there were no observed accumulation trends across participants. TA06 was the only participant to have an immediate decrease in accumulation from the FR-100 (1.33/exchange) to FR-200 (1.0/exchange). In opposition to the predicted effect, a linear increase in accumulation was observed with TA07 as the production schedule increased.

A procedural difference between Yankelevitz et al. and the present study may explain the lack of token production schedule effect. Two participants did not run an FR-300 token production schedule because their accumulation had already been eliminated at the FR200 level (e.g., TA06). As a matter of efficiency, the FR300 condition was not conducted since that condition tends to reduce accumulation. Since they did not run an FR300 schedule, their low levels of tokens spent at each exchange did not contribute to the overall participant mean for the FR300. Not including these sessions likely increased the mean tokens spent and % multiple exchanges on the FR-300 schedule, which would have then been like the results on the FR-200 schedule.

Another important difference between this study and Yankelevitz et al. is that the magnitude of FR schedules used in this study were much higher. The FR-100 schedule produced low baseline accumulation, which led to a floor effect. It is possible that the parameter space within which humans would accumulate tokens on this task lies below the FR 100 response requirement. Perhaps more importantly, all token production manipulations were done with the smallest exchange production response requirement. Accumulation may be much more sensitive to token production schedule changes in the context of a larger exchange production schedule.

A possible limitation of the present study was that container size appeared to be a determinant of participants' behavior. For example, participant TA08 spent two tokens at a time during most sessions. Informal conversation with this participant indicated that the reason for this was that two tokens roughly equaled an individual-sized bag of chips. Thus, the spending of tokens may have been under the partial antecedent control of the commercial packaging of the food items on the menu. Chips were exchanged at the rate of one token for 14 grams, but this was roughly equivalent to one-half of a bag, which may have led to self-generated rules such as, "two tokens equals a bag." Further, participants observed the weighing of the food products on the scale. Such issues could easily be avoided in future studies by, for example, using larger bags of snacks rather than single serving bags. Consideration of stimulus control related to commercial packaging, and procedures designed to eliminate it should be a design consideration in future human operant studies of token systems.

A second potential limitation of this study was that the order of conditions was identical for all participants. All participants underwent each condition in the same order. This leaves open the possibility of an undetected sequence effect. For example, the No Walk condition may have served as an "anchor" for the short walk and long walk conditions (Tversky & Kahneman, 1974). A participant may have made accumulation decisions based on the magnitude of the first schedule. Counterbalancing could be used in an attempt to wash out sequence effects, but such a technique would also obscure any sequence effect rather than revealing it. The effects of condition sequencing on accumulation could be addressed in future studies by directly comparing a limited number of alternative sequences.

Research on reinforcer accumulation has significant applied value and may improve the quality of token economies as interventions. For example, in one study that used token systems to promote appropriate behavior, participants who save their tokens show performance decline over time (Winkler, 1973). In Subramaniam et al. (2017), however, the authors found that participants who held a higher balance during a therapeutic workplace intervention for adherence to naltrexone also tended to have higher rates of heroin and cocaine abstinence. Although the conclusions of these studies indicate opposite effects, they both indicate that there are conditions under which accumulation can mediate the effects of token interventions. It is thus possible that interventions that target accumulation specifically could improve overall outcomes in clinical applications of token economies. Given the robust effects of token component schedules on accumulation, these variables would be strong candidates for inclusion in future studies designed to investigate accumulation as a mediating variable in token economy interventions.

Generalized reinforcers are part of everyday human life and come in many forms ranging from verbal behavior to money (Skinner, 1953). We join Tan & Hackenberg (2015) in the view that despite the obvious translational value of generalized reinforcement studies, the literature remains limited. Generalizability of tokens did not produce a robust effect in the present study. Nevertheless, the possibility remains that generalizability may moderate the relationship between performance and accumulation. For example, the incentives used in Subramaniam et al. were paychecks, which are highly generalized reinforcers. Participants were able to save large sums of money during that study to pay for high cost bills such as rent. However, in Winkler et al. (1973) the tokens participants were working towards were restricted to privileges, meals, and beverages in an inpatient ward in which basic needs were met independent of the participants' performance. It is possible that participants were saving their tokens in that study partially because the relevant motivational operations fairly weak. This could serve to enhance the effects of the component schedules on accumulation and is also consistent with the reduction in earning responses observed in the study. Thus, a parametric analysis of the effects of token production schedule, exchange production schedule; and token generalizability on accumulation is warranted. This would constitute a systematic replication of Yankelevitz et al. (2008), with human participants and the addition of a token generalizability manipulation.

The relationship between the token exchange schedule and token accumulation also warrants further study. Yankelevitz et al. (2008) held the token exchange schedule constant at FR-1 when manipulating token production and exchange production schedules. In applied settings, the token exchange schedule is typically the number of tokens required to purchase a backup reinforcer. Increasing the token exchange schedule may promote accumulation. However, the price of the items was not experimentally manipulated in the current study. Increasing the token exchange schedule by simply increasing the cost of all backups would inflate accumulation by requiring participants to save more tokens to spend them. However, possessing a number of tokens that is greater than one but less than the minimum necessary to purchase the least costly backup item should not be conceptualized as accumulation. Although price is a common independent variable in the field of economics, the effects of the token exchange schedule on accumulation has yet to be explored in an operant framework.

The present study was the first investigation of token generalizability and accumulation with human subjects and one of the first to indicate that the relationship between generalizability and accumulation merits further inquiry. Additionally, this study provides additional support for the exchange production schedule findings from Yankelevitz et al. (2008) This finding could have translational value, especially if saving undermines the effectiveness of token-based interventions.

REFERENCES

Andrade, L. F., & Hackenberg, T. D. (2017). Substitution effects in a generalized token economy with pigeons. *Journal of the Experimental Analysis of Behavior*, *107(1)*, 123–135. http://dx.doi.org.libproxy.library.wmich.edu/10.1002/jeab.231

Ayllon, T., & Azrin, N. H. (1965). The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior*, *8(6)*, 357–383. http://dx.doi.org.libproxy.library.wmich.edu/10.1901/jeab.1965.8-357

Charnov, E. (1976). Optimal foraging: The marginal value theorem. *Theoretical Population Biology*, *9(2)*, 129–136. https://doi.org/10.1016/0040-5809(76)90040-X

Cole, M. R. (1990). Operant hoarding: A new paradigm for the study of self-control. *Journal of the Experimental Analysis of Behavior*, *53*(2), 247-261. https://doi.org/10.1901/jeab.1990.53-247

DeLeon, I. G., Chase, J. A., Frank-Crawford, M. A., Carreau-Webster, A. B., Triggs, M. M., Bullock, C. E., & Jennett, H. K. (2014). Distributed and accumulated reinforcement arrangements: Evaluations of efficacy and preference. *Journal of Applied Behavior Analysis*, *47*(2), 293-313. https://doi.org/10.1002/jaba.116

Deslauriers, B. C., & Everett, P. B. (1977). Effects of Intermittent and Continuous Token Reinforcement on Bus Ridership. *Journal of Applied Psychology*, *62*(4), 369–375. https://doi.org/10.1037/0021-9010.62.4.369

Falligant, J. M., Pence, S. T., & Bedell, S. B. (2020). Preferences for token exchange-production schedules: Effects of task difficulty and token-production schedules. *Behavioral Interventions*, *35*(2), 234-248. https://doi.org/10.1002/bin.1706

Frank-Crawford, M., Borrero, J. C., Newcomb, E. T., Chen, T., & Schmidt, J. D. (2019). Preference for and efficacy of accumulated and distributed response–reinforcer arrangements during skill acquisition. *Journal of Behavioral Education*, *28*(2), 227-257. https://doi.org/10.1007/s10864-018-09312-7

Fulton, C. J., Tiger, J. H., Meitzen, H. M., & Effertz, H. M. (2020). A comparison of accumulated and distributed reinforcement periods with children exhibiting escape-maintained problem behavior. *Journal of Applied Behavior Analysis*, *53*(2), 782-795. https://doi.org/10.1002/jaba.622

Hackenberg, T. D. (2018). Token reinforcement: Translational research and application. *Journal of Applied Behavior Analysis*, *51*(2), 393–435. https://doi.org/10.1002/jaba.439

Kazdin, A. E. (1982). The token economy: A decade later. *Journal of Applied Behavior Analysis*, *15*(3), 431–445. https://doi.org/10.1901/jaba.1982.15-431

Kazdin, A. E., & Bootzin, R. R. (1972). The token economy: An evaluative review. *Journal of Applied Behavior Analysis*, *5*, 343–372. https://doi.org/10.1901/jaba.1972.5-343

Killeen, P.R. (1974). Psychophysical distance functions for hooded rats. *The Psychological Record*, *24*, 229–235. https://doi.org/10.1007/BF03394238

Lovaas, O. I. (1968). Some studies on the treatment of childhood schizophrenia. In J. M. Shlien (Ed.), Research in psychotherapy (pp. 103-121, Chapter vii, 615 Pages) American Psychological Association, Washington, DC. https://doi.org/10.1037/10546-005

McFarland, J. M., & Lattal, K. A. (2001). Determinants of reinforcer accumulation during an operant task. *Journal of the Experimental Analysis of Behavior*, *76*(3), 321-338. https://doi.org/10.1901/jeab.2001.76-321

Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *The Analysis of Verbal Behavior*, *27*, 3–22. https://doi.org/10.1007/BF03393089

Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1971). Achievement Place: modification of the behaviors of pre-delinquent boys within a token economy. *Journal of Applied Behavior Analysis*, *4*, 45–59. https://doi.org/10.1901/jaba.1971.4-45

Robinson, N., & St Peter, C.,C. (2019). Accumulated reinforcers increase academic responding and suppress problem behavior for students with Attention-Deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, *52*(4), 1076-1088. https://doi.org/10.1002/jaba.570

Skinner, B.F. (1953). Science and human behavior, New York: Macmillan. https://doi.org/10.1002/sce.37303805120

Skinner, B.F. (1957). Verbal Behavior, New York: Macmillan. https://doi.org/10.1037/11256-000

Sousa, C., & Matsuzawa, T. (2001). The use of tokens as rewards and tools by chimpanzees (Pan troglodytes). *Animal Cognition*, *4*(3-4), 213-221. https://doi.org/10.1007/s100710100104

Winkler, R. C. (1973). An experimental analysis of economic balance, savings and wages in a token economy. *Behavior Therapy*, *4*(1), 22-40. https://doi.org/10.1016/S0005-7894(73)80072-3

Yankelevitz, R. L., Bullock, C. E., & Hackenberg, T. D. (2008). Reinforcer accumulation in a token-reinforcement context with pigeons. *Journal of the Experimental Analysis of Behavior*, *90*(3), 283-99. https://doi.org/10.1901/jeab.2008.90-283

BRIEF REPORT

DIFFERENTIATED REINFORCEMENT ALTERS CHOICE PREFERENCES OF HUMANS

Craig A. Marrer and Benjamin N. Witts

DEPARTMENT OF COMMUNITY PSYCHOLOGY, COUNSELLING, AND FAMILY THERAPY, ST. CLOUD STATE UNIVERSITY

Understanding preference for choice is valuable in accounting for how different organisms respond to different environmental conditions. Part of this understanding comes from studying preference (or non-preference) for choice under differing and changing conditions, and in-the-moment flexibility to chain arrangements and reinforcer delivery will aid in this pursuit. When an arrangement contains two or more concurrently available alternatives, each of which functions as a discriminative stimulus (SD), the arrangement is called free choice whereas, when only one SD is available, it is considered a restricted choice (Martin, Yu, Martin, & Fazzio, 2006). Generally, both human and non-human organisms tend to prefer stimulus arrangements containing choice compared to arrangements with no choice (e.g., Catania, 1975, 1980; Catania & Sagvolden, 1980; Fisher, Thompson, Piazza, Crosland, & Gotjen, 1997; Sellers et al., 2013; Skowronski & Carlston, 1982; Tiger, Hanley, & Hernandez, 2006.)

Concurrent-chains schedules of reinforcement arrange two or more affixed simple schedules and are often used to study preference between free and restricted choice arrangements (see Fisher & Mazur, 1997). In a concurrent-chains design, the first simple schedule or initial link is signaled by the presence of two stimuli; one paired to the free choice arrangement and one to the restricted choice arrangement. When the schedule requirement under either of the initial links is met, the corresponding terminal link schedule is presented, and satisfying the terminal link requirement produces the consequence: a putative reinforcer. When terminal link work requirements and outcomes are equal, and only reinforcer presentation differs (free versus restricted choice), initial link responding can be used as a measure of preference. For example, Schmidt et al. (2009) found that when eight typically developing children were provided the opportunity to choose between 5 identical preferred items (free choice) or receive the same but therapist-selected item (restricted choice), responding was generally allocated towards the free choice initial link. As preference was not distributed randomly between options, these results suggest that the opportunity to choose is reinforcing beyond the terminal reinforcement available.

However, when it becomes advantageous to change preference due to some variation between the free and restricted terminal link outcomes, organisms tend to allocate their responding to whichever schedule terminates in quantitatively or qualitatively more valuable reinforcement (Hayes, Kapust, Leonard, & Rosenfarb, 1981; Karsina et al., 2011). Fisher, Thompson, Piazza, Crosland, and Gotjen (1997) found that while three children in an inpatient program initially preferred free choice arrangements when receiving contingent access to either high preference or low preference items, their preferences changed to the restricted choice arrangement when low preference items were delivered contingent on selecting the free choice initial link and high preference items contingent on selecting the restricted choice initial link. Thus, while free choice arrangements can be preferable, likely due to the opportunity

Address correspondence to: Craig Marrer (camarrer@gmail.com) or Benjamin Witts (bnwitts@stcloudstate.edu), Department of Community Psychology, Counseling, and Family Therapy, St. Cloud State University, 720 4th Ave S, St Cloud, MN 56301

to choose, the putative reinforcer can develop selective control over subsequent response allocation.

The exact mechanisms responsible for the general finding that organisms prefer free choice arrangements are not clear. Additionally, there have been studies in which participants have shown a restricted choice preference. Karsina, Thompson, and Rodriguez (2011) demonstrated that for seven college-aged students who preferred restricted choice to free choice (4 participants) or did not show a preference (3 participants), choice preference was amendable to a differential reinforcement procedure. Much like research that has shown free choice preference to be amenable to change when terminal link variables are altered, these researchers demonstrated the same for participants that showed a restricted choice preference. Additionally, Karsina et al. found that the conditioned free choice preferences for 5 of the 7 participants persisted during a withdrawal to baseline conditions where reinforcement for both the free and restricted choice contingences was equal.

The results of the above-mentioned studies support that both free and restricted choice preferences are amendable to prevailing contingencies when adequate selection pressure is placed on responding (e.g., via differential reinforcement). However, few studies have presented equivalent procedures to participants preferring free or restricted choice arrangements within the same experimental study. The current study presents greater flexibility in researching choice through the use of a computer-based game with built-in algorithms that allow for intervention condition assignments to be carried out for each participant based on baseline choice arrangement preferences.

A parametric differentiated reinforcement procedure (see van Haaren, 2017, for a review of differentiated reinforcement) was used in which points for the non-preferred choice arrangement progressively increased while diverging from the points available for the preferred choice arrangement. Through the use of an algorithm and an automated software function, an equivalent intervention procedure was presented across both groups of participants. It was anticipated that regardless of baseline preference, each participant's responding would come under the control of the prevailing contingencies, as has been demonstrated previously within the research. Additionally, in keeping with previous research findings, it was anticipated that most participants would show a preference for the free choice arrangement during the baseline condition. Lastly, it was anticipated that there would be no clear differentiation between the response patterns of those with a free choice preference and of those with a restricted choice preference, showing that regardless of choice arrangement preference in a given context, contingent differentiated reinforcement would come to control response allocation.

METHOD

Participants and Setting

Twelve undergraduate students (8 female, 4 male; M age = 21.90; range = 19-27) enrolled at a mid-sized Midwest university participated. All sessions were conducted in an approximately 6.5 m by 2.6 m research room. The participant space consisted of two long tables (1.21 m and 1.05 m), each with a computer monitor and a chair. Participants completed the study one at a time. All participants were compensated $12.00 for completion of the study. IRB approval was obtained, and informed consent procedures were followed for each participant. Due to the use of deception, each participant was given a debriefing statement at the end of the study, which explained how and why deception was used.

Materials

A computer program, built using Java programming language, was used to present the informed consent, demographic questionnaire, participant training, training quiz, and the experimental procedure. Participants used a standard computer mouse to input information into the program. The computer program was hosted on a private server and was accessible via a web address. The software recorded all mouse clicks related to the demographic questionnaire, training quiz, and responses towards initial and terminal links for each trial into a comma-separated value format stored on the program server. Participants were assigned a username and a password. A research assistant was present to log each participant into the program.

Design and Program Accuracy

An A-B-A or withdrawal design was used. Thirty-one baseline trials were presented, followed by 30 intervention trials and 30 withdrawal trials, resulting in a total of 91 trials. Due to the automatization of the data collection procedure, program accuracy was assessed by taking response data from screen-recorded pilot studies and comparing these data to the automatically compiled data from the software. Data collection accuracy for the software was found to be 100% across three consecutive pilot study sessions.

Procedure

A concurrent-chains arrangement was used to measure participant preference between free and restricted choice arrangements. Trials were presented via a computer-based game in which three different colored squares - blue, red, and green - each measuring 5.08 by 5.08 cm or 600 by 600 pixels, were displayed in a quasi-randomized order from left to right and served as the free choice, restricted choice, and control arrangements, respectively, for every participant. All initial link stimuli were represented as a single-celled square and were activated when the participant clicked the mouse cursor one time anywhere within the cell area (see Figure 1).

Once an initial link response was recorded, the selected array moved to the center of the screen and displayed as a 100-celled array for free and restricted choice terminal link arrangements or a single-celled array for the control arrangement. The free choice terminal link required the participant to select 3 of the 100 available cells of his or her choosing with the mouse cursor. Selections were indicated by a darkening of the selected cell. Similarly, when the restricted choice array was selected, the participant was required to click the mouse cursor three times. However, when the participant clicked the mouse, a single random cell was darkened, indicating it was activated and that the mouse's cursor position at the moment of a

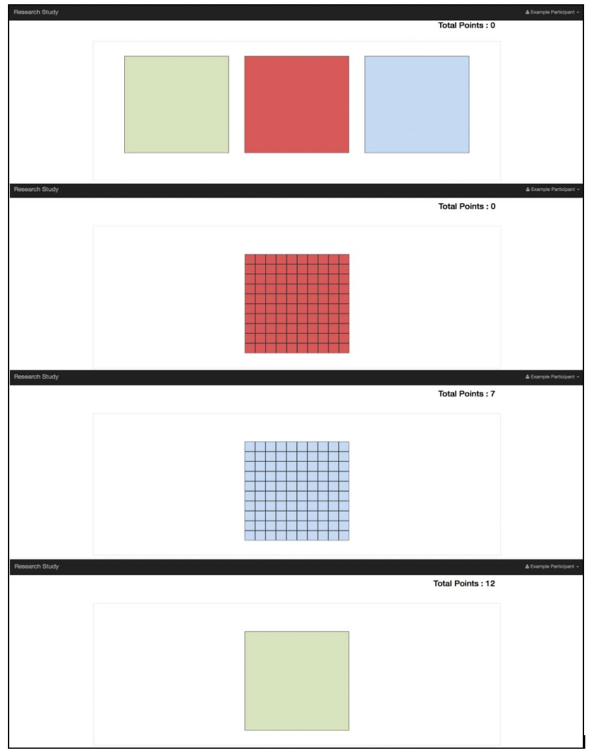


Figure 1. Screen shot of an initial link choice situation (top) with corresponding choice type and display color shown within figures. Screen shot of terminal link choice situation for the forced choice (top), free choice (middle) and control (bottom) array types.

click had no influence on what cell would be activated. Lastly, when the control array was selected, the participant was required to click one additional time anywhere within the array to complete the terminal link requirement. Following the next mouse click, the entire array darkened to indicate that the response requirement had been met. A control array was included to account for non-discriminated scrolling responses (i.e., selecting an array without first visually attending to the stimulus arrangement).

Participants were awarded points when meeting the terminal link requirement of the free or restricted choice arrays. Selection of the control condition always resulted in no points being awarded. Total trial point values were shown to the participant at the end of each terminal link trial, and the sum of points earned across trials was shown in a score box in the upper right-hand corner of the computer screen. Each response during the terminal link of the choice sequence was assigned an individual cell point value, which ranged from 0 to 4 points. The total point value per trial for the free and restricted choice arrays was always the sum of the three individual cell values activated and ranged from 0 to 12 points. Selection of the control array always resulted in 0 points.

During the baseline phase, total trial points were awarded at an equal probability for both choice arrangements based upon predetermined ratio assignments (see below). After meeting the initial link schedule requirement for the free, restricted, or control arrangements (FR-1) the participant was presented with the terminal link component. The terminal link schedule requirement for the free and restricted choice arrangements was FR-3, while the control array terminated following a single additional response (FR-1).

Baseline terminated following 31 trials, at which time the computer software analyzed the proportion of responding to each choice arrangement and assigned a participant to one of two groups—the differentiated reinforcement of free choice (DRFC) or the differentiated reinforcement of restricted choice (DRRC). Participants that allocated an equal number of responses to each choice arrangement (due to the selection of the control array) were automatically assigned to the DRRC group.

The differentiated reinforcement procedure was a parametric procedure in which total trial point values for the preferred and non-preferred choice arrangements progressively diverged over the course of the experimental condition. For example, if a participant showed a preference for free choice during baseline, restricted choice terminal links terminated with increasingly higher point totals over the course of the intervention condition. The inverse was true for participants that showed a preference for restricted choice during baseline. Regardless of baseline preference, all participants received the same intervention condition in relation to points available for selecting between their preferred and non-preferred choice arrangements. A withdrawal to baseline conditions was conducted for 30 trials to assess the maintenance of the differentiated reinforcement procedure.

Points

To determine individual cell point values, a probability of occurrences out of 10 was set for each possible point value with actual occurrence generated via a randomization formula. During baseline and withdrawal, these probabilities were set to occur at 10% occurrence for 0 and 4 points, 20% for 1 and 3 points, and 40% for 2 points. By arranging probabilities in this manner, total trial point values clustered around the median of 6 points. This minimized the likelihood of any participant receiving relatively high or low points during any trials during the baseline and withdrawal conditions, therefore, minimizing the relative reinforcer value of either choice arrangement over the other.

During the intervention condition, seven ratio modifications occurred over 30 trials (see Table 1). To determine individual cell point values for the intervention condition, 90 numbers (0-4) were generated using the randomization formula (30 trials of 3 numbers each) and assigned to the non-preferred choice condition (either free or restricted, contingent on participant preference). Conversely, the inverse value was determined and assigned to the preferred choice condition (again, contingent on participant preference). For example, if the randomized three number sequence for the non-preferred choice arrangement was 4-3-4, the inverse three number sequence for the preferred choice arrangement relative to the number of points away from the median was 0-1-0, with 2 being the median single cell value

Sessions and Instructions

At the start of each session, participants were shown to the computer by a research assistant and informed that all instructions related to the study would be provided via the computer. Training consisted of a sequence of six static instructional screens in which the initial and terminal link representations of the experimental stimuli were presented, one at a time, with instructions on how the participant must interact with each stimulus. Participants were also informed at this time that for each point

Table 1

Reinforcer Value Assigned Probability Per Phase

Treatment Phase	Points									
	0	1	2	3	4	0	1	2	3	4
Baseline										
1-31	10%	20%	40%	20%	10%					
Intervention	Non-preferred choice situation					Preferred choice situation				
32-34	15%	20%	20%	20%	25%	25%	20%	20%	20%	15%
35-37	15%	15%	20%	25%	25%	25%	25%	20%	15%	15%
38-40	10%	15%	20%	25%	30%	30%	25%	20%	15%	10%
41-43	10%	10%	20%	30%	30%	30%	30%	20%	10%	10%
44-46	5%	10%	20%	30%	35%	35%	30%	20%	10%	5%
47-49	5%	5%	20%	35%	35%	35%	35%	20%	5%	5%
50-52	0%	5%	20%	35%	40%	40%	35%	20%	5%	0%
53-55	0%	0%	20%	40%	40%	40%	40%	20%	0%	0%
56-58	0%	0%	20%	40%	40%	40%	40%	20%	0%	0%
59-61	0%	0%	20%	40%	40%	40%	40%	20%	0%	0%
Withdrawal										
62-91	10%	20%	40%	20%	10%					

Note. Percentages indicate probability of corresponding number occurring when running a quasi-randomized number generating equation.

earned during the study, they would be compensated $0.01 in addition to the $5.00 they were already receiving for completing the study; as points were predetermined by the researcher for two of the three phases, all participants were compensated for the maximum number of points possible in the study, which was 700 points or an additional $7.00, making total compensation $12.00 per participant. However, participants were not made aware of this until the end of the study.

Following training, participants were given a 5-question quiz to ensure understanding of each of the requirements of the study. A score of 100% was required to move onto the next phase of the study. If a participant failed to score 100%

on his or her first try, he or she was provided a printed copy of the training to review a second time and retake the quiz. No participants failed to pass the quiz.

RESULTS

Figure 2 shows the proportion of response allocation between the choice arrangements for all participants during baseline, intervention, and withdrawal conditions. Of the 12 participants, 8 allocated more responses to the free choice arrangement during baseline, 3 to the re-

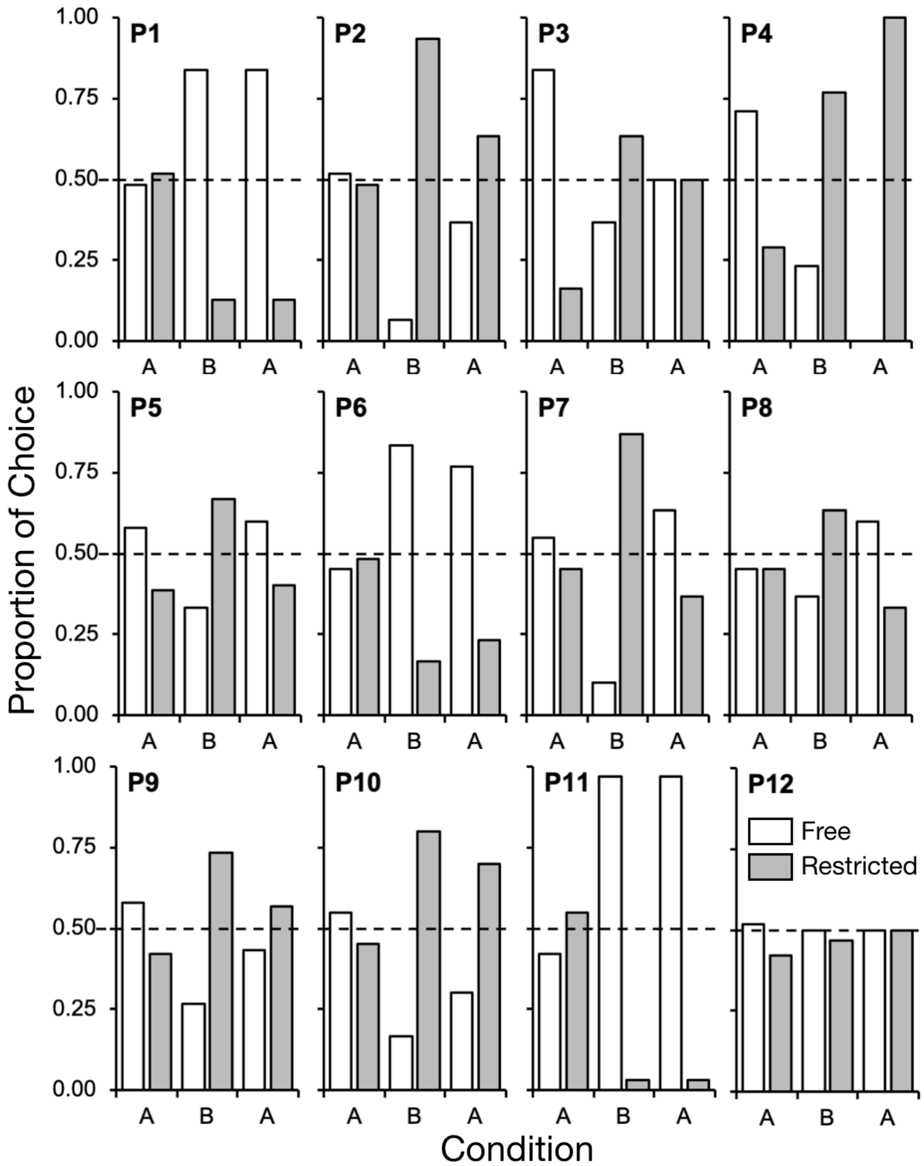


Figure 2. Proportion of responses allocated, per phase, to each choice arrangement, for all participants. White vertical bars represent free choice array selections and gray vertical bars represent restricted choice array selections

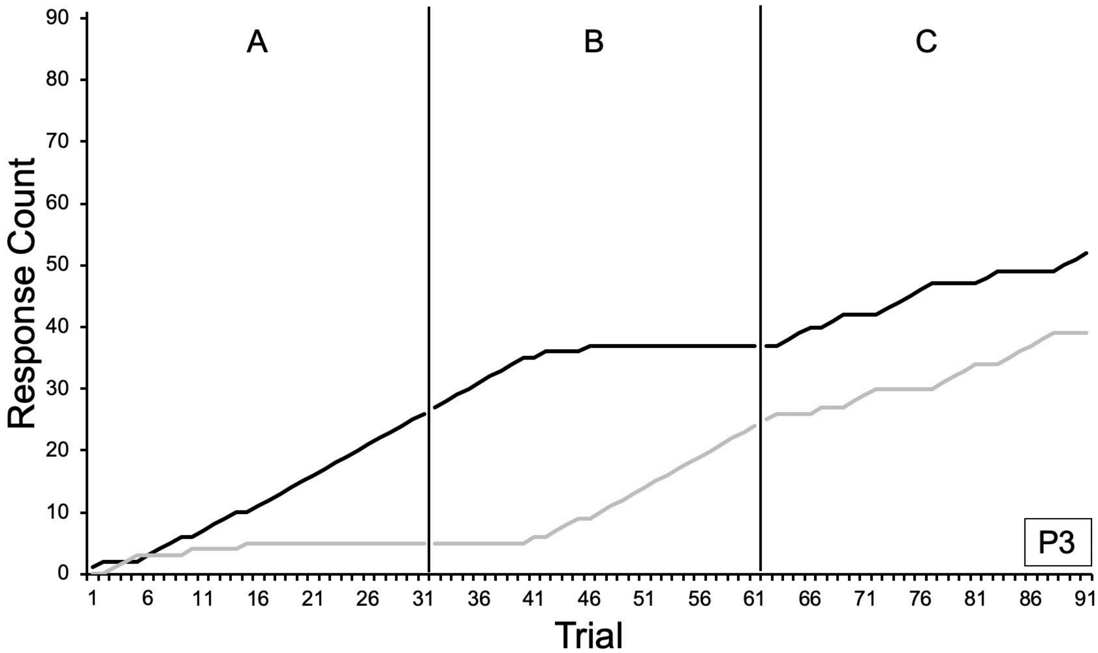


Figure 3. Cumulative response count for participant 3. Responding indicated a clear preference for the free choice array during baseline that persisted until trial 41 during the B phase. Following trial 41, participant 3 allocated all remaining B phase responses toward the restricted choice array. During the return to baseline (phase 3), an oscillation between the two choice arrangements can be seen. Free choice selections are represented by black lines and restricted choice selections by grey lines.

stricted choice arrangement, and 1 participant allocated an even number of responses to each (due to the selection of the control arrangement). However, only two participants, P3 and P4, showed a clearly differentiated preference between the choice arrangements, with both allocating at least 70% of responses to the free choice arrangement (see Figure 3 for a representative example). In general, most participants' responding during baseline indicated a general indifference between the free and restricted choice arrangements, often choosing to alternate responding between each choice arrangement in a relatively patterned manner (see Figure 4 for a representative example).

Following baseline, eight participants were assigned to the DRRC condition and four to the DRFC condition. All but one participant's responding came under the control of the differentiated reinforcement procedure, with P12 continuing to allocate a higher proportion of responding to her preferred choice arrangement. At the end of the intervention condition, eight participants were allocating responding in a manner consistent with a restricted choice preference and four with a free choice preference, which was the opposite of their baseline results. Upon return to baseline conditions, seven participants (P1, P2, P4, P6, P9, P10, and P11) continued to allocate a higher proportion of responding to the recently conditioned choice arrangement preference, indicating maintenance of the recent conditioning procedure, with three allocating a higher proportion to their baseline choice arrangement preferences (P5, P7, and P8), and two showing indifference (P3 and P12).

To assess for the possibility of unintended differentiated reinforcement occurring during the baseline condition, therefore causing a preference and conditioning effect prior to the presentation of the parametric differentiated reinforcement procedure, an analysis of the average total trial point value awarded following each choice condition was conducted. As can be seen in Figure 5, 7 of the 12 participants actually earned more points, on average, during baseline under their non-preferred choice arrangement.

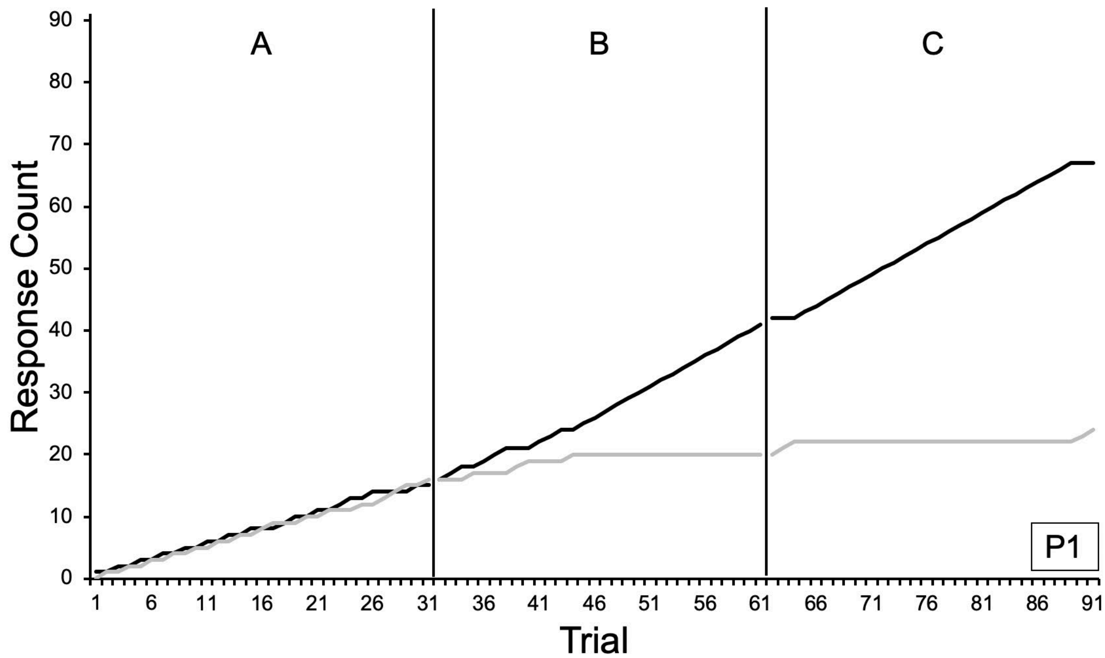


Figure 4. Cumulative response count for participant 1. Responding indicated a general indifference between the free and restricted choice arrays during baseline, with a clear preference emerging during the B phase that maintained following a return to the baseline procedure. Free choice selections are represented by black lines and restricted choice selections by grey lines.

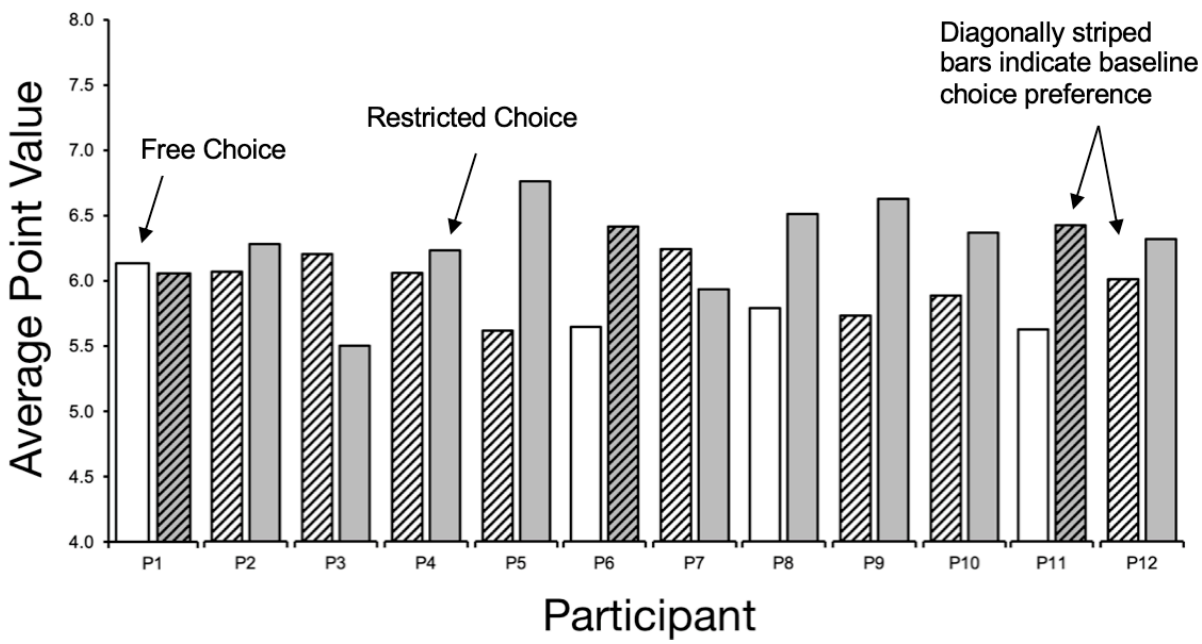


Figure 5. Average point totals earned per choice arrangement per participant during baseline. Free choice averages are displayed as white bars and restricted choice by gray bars. Bars with upward diagonal lines represent the choice arrangement preference for each participant excluding participant 8 who did not show a preference between the choice arrangements during baseline.

DISCUSSION

The current study investigated the effect of a differentiated reinforcement procedure on the choice arrangement preferences of human participants using automated computer software that allowed participants of any choice preference during baseline to participate. While previous studies have successfully altered free and restricted choice preferences of humans and non-humans using differential reinforcement procedures, few have been able to simultaneously investigate both free and restricted choice arrangement preferences and compare and contrast intervention effects using the same experimental procedure.

The results of the current study were consistent with previous research in that differentiated reinforcement, like differential reinforcement, altered the choice arrangement preferences of human participants. Of the 12 college-aged participants, all but one showed a preference for the non-preferred choice arrangement during the differentiated reinforcement procedure. Additionally, for seven of these participants, conditioned choice preferences persisted when reinforcement was returned to baseline levels during a withdrawal condition (maintenance). These findings are consistent with Karsina et al. (2011).

The current study extends some aspects of previous choice research in several potentially important ways. First, the use of a computer algorithm to assign participants to one of two intervention phases, depending on baseline responding, allowed all participants to be included. In previous research, a priori exclusion and inclusion criteria were necessary or deemed desirable, depending on the purpose of the study. However, by including all participants, it was possible to analyze intervention effects for participants that preferred both free and restricted choice arrangements as well as for those that showed little to no pre-intervention preferences.

Second, the current study used monetary compensation that was designed to appear as if it was corollary with a participant's performance. We hypothesized that by establishing motivation for higher point totals, participants would be more sensitive the prevailing contingencies in place during the intervention phase. However, the motivational effect of monetary compensation was not experimentally demonstrated and therefore, cannot be said to have contributed to the data in any meaningful way. Future research might want to investigate this further as it could be found that the use of a potent conditioned reinforcer such as money could more closely capture the motivation of a participant in a naturalistic choice situation.

However, while the design of this study—specifically the use of an algorithm—did demonstrate the potential utility of using computer-assisted interfaces for investigating behavioral phenomena, there are several limitations. First, the data do not indicate why some participant's responding was undifferentiated during baseline, but others was not. It is possible the baseline condition was too brief and, therefore, inadequate to establish a preference for some participants. Alternatively, the contingency arrangement during baseline might have inadvertently reinforced an undifferentiated pattern of behavior in some participants. Lastly, some people might not have a preference between free and restricted choice arrangements. However, since most participants' behavior did come under control of the differentiated procedure, with conditioning effects persisting during the withdrawal phase, inclusion and further analysis of these data are warranted.

A second limitation is related to the limited number of trials presented during each condition. While 91 trials were presented to each participant, it might be that additional trials are required to allow changes in response patterns to be fully recognized. This seems to be especially true during the withdrawal phase. For example, several participants' responding during the withdrawal phase appear to be returning to response allocation more consistent with baseline levels (see Figure 6 for a representative example). It might be found that with a withdrawal phase extended another 30 to 60 trials, more participants' preferences would correspond to baseline preferences as the transitory effects of the intervention procedure loses control. Extending phases would also allow for a better assessment of stability, which could be programmed into the computer algorithm as a means of determining when to switch conditions.

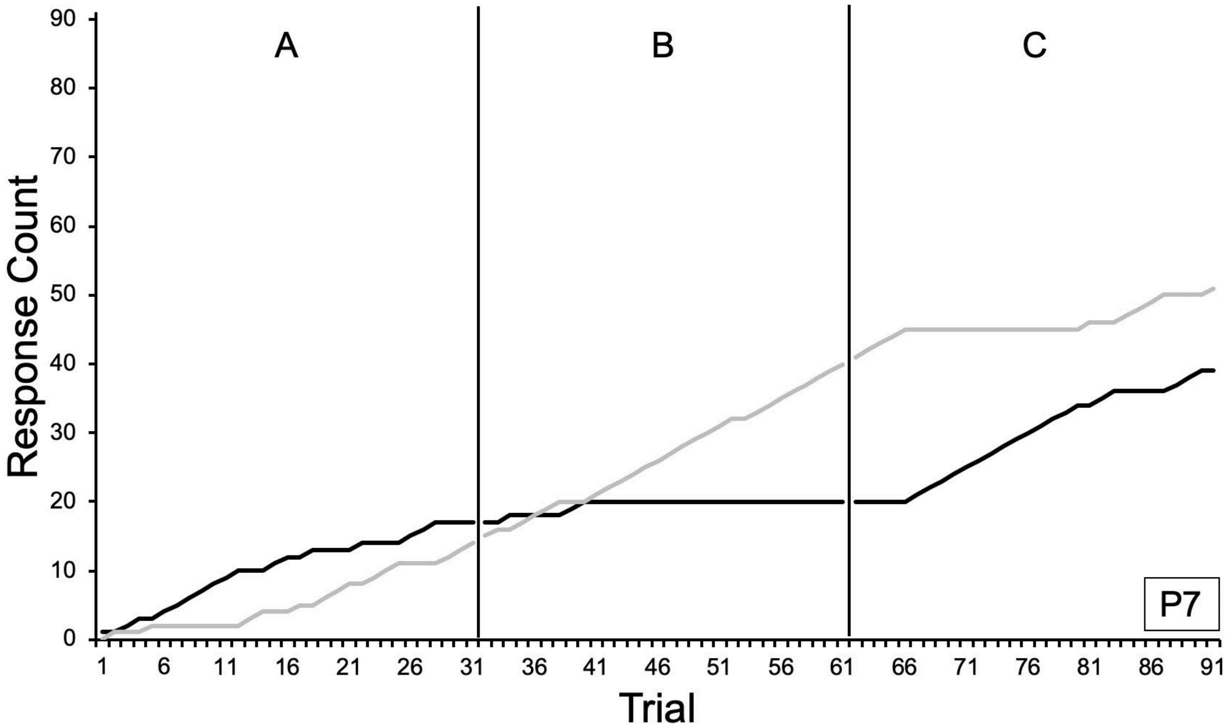


Figure 6. Cumulative response count for participant 7. Responding indicated a slight preference for the free choice situation during baseline, with a clear preference emerging during the B phase that maintained following a return to the baseline procedure. Free choice selections are represented by black lines and restricted choice selections by grey lines.

Thirdly, since the experimental interface relied on the use of colors (blue, red, and green) to function as discriminative stimuli for the different choice arrangements, it is possible that participants who experience color-blindness, would not be able to adequately discriminate between the experimental stimuli. This was not assessed during the current study, but future researchers using similar experimental interfaces would be wise to assess for this prior to implementation of the study. A fourth potential limitation, also related to the use of color, is that the color assignment between each choice type was not randomized between or across participants, meaning a color-based bias could have inadvertently affected participant preference.

The current study sought to investigate choice arrangement preferences in humans that preferred either free or restricted choice situations and was able to show, at least preliminarily, that when individuals with free or restricted choice preferences are exposed to the same experimental procedures, little difference can be found in each group's sensitivity to differentiated contingencies. Consistent with previous research, the majority participants in the current study did show a preference for the free choice arrangement over the restricted choice arrangement. However, the baseline preference was unremarkable.

By conducting additional research in this area, it seems plausible that certain response patterns will be identified that share similarities in the histories of the individual. For example, some participants that showed no preference during baseline ended up showing clear preferences during and after intervention—but others did not. Lastly, for some participants, there appeared to be a "cancelling out effect" in that they showed a preference during baseline, allocated responding to the more favorable choice option during intervention, and then showed no preference during withdrawal, however, additional research is needed to assess the significance of this finding.

REFERENCES

Catania, A. C. (1975). Freedom and knowledge: An experimental analysis of preference in pigeons. Journal of the Experimental Analysis of Behavior, *24*, 89-106.

Catania, A. C. (1980). Free of choice: A behavioral analysis. The Psychology of Learning and Motivation, *14*, 97-145.

Catania, A. C. & Sagvolden, T. (1980). Preference for FC over forced choice in pigeons. Journal of the Experimental Analysis of Behavior, *34*, 77-86.

Fisher, W. W. & Mazur, J. E. (1997). Basic and applied research on choice responding. Journal of Applied Behavior Analysis, *30*, 387-410.

Fisher, W. W., Thompson, R. H., Piazza, C. C., Crosland, K., & Gotjen, D. (1997). On the relative reinforcing effects of choice and differential consequences. Journal of Applied Behavior Analysis, *30*, 423-438.

Hayes, S. C., Kapust, J., Leonard, S. R., & Rosenfarb, I. (1981). Escape from freedom: Choosing not to choose in pigeons. Journal of the Experimental Analysis of Behavior, *36*, 1-7.

Karsina, A., Thompson, R. H., & Rodríguez, N. M. (2011). Effects of a history of differential reinforcement on preference for choice. Journal of the Experimental Analysis of Behavior, *95*, 189-202.

Martin, T. L., Yu, C. T., Martin, G. L., & Fazzio, D. (2006). On choice, preference, and preference for choice. The Behavior Analyst Today, *7*, 234-241.

Schmidt, A. C., Hanley, G. P., & Layer, S. A. (2009). A further analysis of the value of choice: Controlling for illusory discriminative stimuli and evaluating the effects of less preferred items. Journal of Applied Behavior Analysis, *42*, 711–716.

Sellers, T. P., Bloom, S. E., Samaha, A. L., Dayton, E., Lambert, J. M., & Keyl-Austin, A. A. (2013). Evaluation of some components of choice making. Journal of Applied Behavior Analysis, *46*, 455-464.

Skowronski, J.J., & Carlston, D.E. (1982). The effects of previously experienced outcomes on the desire for choice. Journal of Personality and Social Psychology, *43*, 689-701.

Tiger, J. H., Toussaint, K, A., & Roath, C. T. (2010). An evaluation of the value of choice making opportunities in single-operant arrangements: Simple fixed- and progressive-ratio schedules. Journal of Applied Behavior Analysis, *43*, 519-524.

van Haaren, F. (2017). Differential versus differentiated reinforcement. Behavior Analysis: Research and Practice, *17*, 98-100.

BRIEF REPORT

FAILURE TO SYSTEMATICALLY REPLICATE THE FACILITATIVE EFFECTS OF PROGRESSIVE MUSCLE RELAXATION ON DERIVED STIMULUS RELATIONS

Adam H. Doughty and Jenna A. Soydan

COLLEGE OF CHARLESTON

Behavior analysts, historically (e.g., Lattal & Harzem, 1984) and recently (e.g., Vyse, 2013), have discussed strategies and tactics to not only promote the survival of the experimental analysis of behavior as a field of inquiry but to strengthen it. One strategy is to continue to refine our understanding of the environmental control over derived, or emergent, stimulus relations (e.g., Critchfield, Barnes-Holmes, & Dougher, 2018). This improved understanding helps behavior analysts contribute to analyses of language, cognition, and other inter-disciplinary topics (e.g., Barnes-Holmes, Finn, McEnteggart, & Barnes-Holmes, 2018). Tyndall, Howe, and Roche (2016) adopted the tactic of examining derived relations in a manner appealing to other psychologists by investigating derived relations as a function of brief progressive muscle relaxation (PMR). From a behavioral perspective (Tyndall et al., 2016), PMR may exert its facilitative effects by sharpening stimulus control over emergent responding (i.e., by reducing the likelihood of extraneous forms of control such as when participants report "cognitive intrusions" during learning tasks).

Tyndall et al. (2016) exposed 35 adult participants to five phases. Across the initial three phases, participants learned arbitrary-matching-to-sample (AMTS) discriminations that could have developed into two 4-member stimulus-equivalence relations (A1/B1/C1/D1, A2/B2/C2/D2). In Phase 4, participants were assigned to one of three groups: PMR condition, Nonrelaxation Condition 1, and Nonrelaxation Condition 2. Participants in the PMR condition listened to an 11-min recording to induce deep relaxation. Participants in the two control groups completed either a simple- or conditional-discrimination task for a similar period of time. In the final phase, all participants received equivalence testing (C1/A1, C2/A2, D1/A1, D2/A2) in the absence of differential consequences. Approximately half of the participants in the PMR group responded successfully in equivalence testing, whereas only one participant did so across the control groups. These findings are important in demonstrating the facilitative effects of mindfulness-related techniques on human cognition (e.g., Cahn & Polich, 2006; Chambers, Chuen Yee Lo, & Allen, 2008), using even abbreviated (i.e., 10 to 12 min) techniques (e.g., Hudetz, Hudetz, & Klayman, 2000; Nava, Landau, Brody, Linder, & Schachinger, 2004). Given the novelty of these findings in the context of derived stimulus relations, it is critical to explore their generality.

One means of exploring the generality of Tyndall et al. (2016) is to examine the impact of brief PMR on derived relations using an assessment other than typical AMTS probe trials. Adduction is the emergence of a novel and complex composite skill after its simpler, component skills have been learned (e.g., Andronis, Layng, & Goldiamond, 1997; Chase, 2003; Epstein, 1987). Three studies have investigated the relation between derived relations and adduction (Arntzen, Petursson, Sadeghi, & Eilifsen, 2015; Bucklin, Dickinson, & Brethower, 2000; Rippy & Doughty, 2017). Bucklin et al. first taught participants AB and BC relations where the A, B, and C stimuli were, respectively, (previously learned) Hebrew symbols, nonsense syllables, and (already learned) Arabic numbers. In adduction testing, arithmetic questions were posed using the Hebrew symbols. For example, participants had to add A1 and A2 such that success-

Address correspondence to Adam Doughty, Department of Psychology, College of Charleston, 57 Coming Street, Charleston, SC 29424 (doughtya@cofc.edu).

ful adduction required them to derive the untested, transitive AC relations and combine them with their extant math skills. Arntzen et al. extended these findings by demonstrating adduction involving relations other than transitive (i.e., symmetrical and equivalence); however, their participants already had derived these relations in probe-trial testing before adduction. Rippy and Doughty extended these two studies by measuring adduction involving untested equivalence relations. Group CA learned AB and BC relations (Arabic numbers [A], nonsense syllables [B], and nonrepresentational stimuli [C]), whereas Group EA learned AB, BC, CD, and DE relations (Arabic numbers [A], nonsense syllables [B, C, and D], and nonrepresentational stimuli [E]). This training could have established four, 3-member classes for Group CA and four, 5-member classes for Group EA, but neither group received derived-relations testing. The C and E stimuli for Groups CA and EA, respectively, were presented in adduction testing such that participants had to combine simple math skills with untested equivalence relations separated by one (Group CA) or three (Group EA) nodes. Successful adduction occurred for each CA participant but in only one EA participant. Whereas the positive CA results further confirmed the utility of adduction to measure derived relational learning, the negative EA results suggest a tactic to measure the impact of variables that potentially can facilitate difficult-to-derive relations.

The present research assessed the impact of brief PMR on derived relations by synthesizing the work of Tyndall et al. (2016) and the procedures applied to Group EA in Rippy and Doughty (2017). In Phases 1 and 3, participants were treated similarly to the Group-EA participants in Rippy and Doughty. They learned AB, BC, CD, and DE relations across Phase 1, and derived EA relations were measured in Phase 3 using arithmetic adduction testing. Critically, participants received 10 min of either the presence or absence of guided meditation to induce relaxation in Phase 2. Although Rippy and Doughty did not examine the effects of brief PMR, their results suggest that adduction involving three-node EA relations was possible but unlikely without additional intervention. Thus, at issue in the present research was whether the facilitative effects of brief PMR would be observed in adduction with untested, three-node equivalence relations. If demonstrated, the generality of brief PMR exposure would be revealed.

METHOD

Participants

Eight College of Charleston students (six female and two male) between the ages of 18 and 22 participated. The sample was collected by displaying flyers across campus and emailing first-year students with information about the study. Participants were told the research would involve one 3-hour laboratory visit, earning them approximately $30.00.

Apparatus

Phase 1 occurred in a smaller room with four workstations separated by dividers. Each workstation had a desk and chair. Each desk had an iMac or eMac, keyboard (which was not used by the participants), and mouse. The contingencies were programmed and responses were recorded using MTS version 11.6.7 (Dube, 1991). Phases 2 and 3 occurred in a nearby and larger conference room with one table and 12 chairs. In Phase 2, participants receiving PMR (see below) were provided with Parrot Zik 2.0 wireless headphones connected to an iPhone 7, and guided mediation was played using the iPhone application, Voice Memos. In Phase 3, participants completed, with a pen, 12 x 7 cm flashcards with arbitrary visual stimuli, numbers, and mathematical operations (see below).

Procedure

Table 1 outlines the three phases. There were five conditions in Phase 1, and all participants were treated identically. They first read these instructions:

> *Welcome to our study! In this part of the study, you will work alone on the computer for several sessions. In each session, the computer will present you with many trials. On each trial, you will be presented with one item, click on that item and additional items will appear. Click the mouse over any one of the surrounding items that you think "goes with" the one in the center,*

and one of two events will occur: (1) a star will appear on the screen or (2) the screen will darken. If a star appears, then you were correct and earned money. If the screen darkens, then you were incorrect and did not earn money. Your task is to earn as much money as possible. The computer will tell you when each session is over. When the session ends, you should find me next door. Good luck!

There were 96 trials in each session of Condition 1 wherein participants learned AB relations (see Table 2). The four AB (i.e., number to nonsense syllable) relations were presented such that each A sample stimulus occurred on 24 trials in each session. Across these 24 trials (e.g., A1), the correct comparison stimulus (e.g., B1) occurred in each screen corner six times. Every trial began with only a sample stimulus in the middle of the screen. After a click over it (observing response), the comparison stimuli immediately appeared with the sample. The stimuli were pseudorandomly organized such that each sample could not occur on more than three consecutive trials, and the correct comparison (S+) could not occur in the same location on more than three consecutive trials. Clicking the correct comparison immediately resulted in stars on the screen for 1 s. Clicking an incorrect comparison resulted in a 1.5-s dark screen. A resetting intertrial interval (ITI) of 1.5 s was used wherein the screen was blank. Condition 1 continued for at least two sessions and until there were no more than two errors per discrimination in the last session. The construction and execution of Conditions 2, 3, and 4 were identical to Condition 1 except that the participants learned the BC, CD, and DE relations in these conditions, respectively.

Each session in Condition 5 consisted of 192 trials. There were 48 trials each of the AB, BC, CD, and DE relations (12 trials with each sample). Across these 12 trials, the correct comparison occurred three times in each corner. Other procedural details (e.g., consequence delivery) remained unchanged from the initial conditions. The condition continued for at least one session and until there were no more than two errors per discrimination in a session.

Phase 2 commenced immediately after Condition 5. Participants were randomly assigned to the PMR or control group. Participants in the control group received these instructions:

Preparing your next session will take ten minutes. Please wait for your researcher to return to present you with the final session.

Participants in the PMR group received these instructions:

Your next session will require you to listen to a recording using headphones for ten minutes. It is recommended that you close your eyes while you listen. When the recording has finished, your researcher will return to collect you for your final session.

After participants read the instructions, the experimenter handed them the headphones, began the recording, dimmed the lights, and closed the door after leaving the room. A transcription of the recording is in the Appendix (the passage was identical to the one used by Tyndall et al. and was presented similarly).

Phase 3 occurred immediately after Phase 2, and the participants were treated identically. They first read these instructions:

Your next session will consist of me presenting you with 96 flashcards. Please calculate the answer and write it down. After you finish one card, I will hand you the next one. No feedback will be given during the session. However, after the session your answers will be assessed and money provided for each correct answer. Good luck!"

Table 3 shows examples of the flashcards (answers were not presented to participants such that each card was blank under the black line). The experimenter provided a pen and presented one flashcard at a time. The experimenter placed each completed flashcard to the side such that participants could not respond to previous flashcards. A limited hold for responding was set to 10 s such that the card was removed if 10 s elapsed without a response, which rarely occurred. The time between flashcards was only as long as it took for the experimenter

Table 1

Outline for training and testing for both groups.

		Discriminations	
Condition	Type	Trained	Tested
Phase 1			
1	Baseline training	AB	—
2	Baseline training	BC	—
3	Baseline training	CD	—
4	Baseline training	DE	—
5	Baseline training	AB, BC, CD, DE	—
Phase 2	PMR manipulation	—	—
Phase 3	Adduction Testing	—	EA

Note. Participants were treated differently only in Phase 2 receiving either the presence or absence of guided meditation for 10 min (see Appendix).

Table 2

Stimuli comprising the discriminations learned in Phase 1.

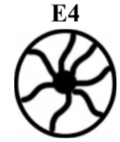
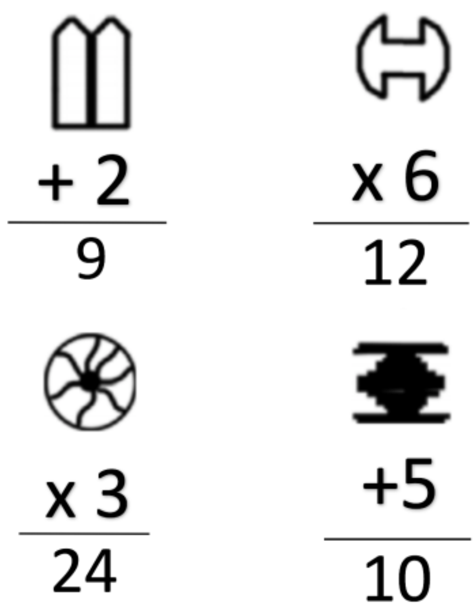
A1	B1	C1	D1	E1
2	yok	fic	het	
A2	B2	C2	D2	E2
5	mof	tep	sul	
A3	B3	C3	D3	E3
7	bal	loz	gix	
A4	B4	C4	D4	E4
8	kad	dut	roj	

Table 3

Examples of the flashcards from Phase 3 (the answers were not presented to the participants).



to place the previous flashcard to the side and present the next one. The experimenter did not provide any feedback following completion of each flashcard. As shown in Table 3, the flashcards presented participants with arithmetic questions involving the E stimuli from Phase 1. Each flashcard required participants to multiply (48 cards) or add (48 cards) an E stimulus with an Arabic number ranging from 1-12 (e.g., E1 + 9). Each E stimulus appeared on 24 flashcards (12 multiplication and 12 addition), and each Arabic number appeared on eight flashcards (four multiplication and four addition). The flashcards were organized pseudorandomly such that there were no more than three consecutive addition or multiplication problems, and no more than three consecutive flashcards with the same E stimulus.

RESULTS

Table 4 displays session-by-session accuracy scores for all participants. With only one exception, each participant in the control group completed the experiment in the minimum number of sessions. The exception was that Participant JR required a second session in Condition 5. There were three exceptions in the PMR group. Participants EL and OD required four sessions in Condition 1, and Participant EL required a second session in Condition 5.

Accuracy scores in the adduction assessment were similar across groups. Only one participant in each group achieved greater than 90% (Participants BB and BM in the PMR and control groups, respectively). Mean accuracies were 28.90% and 23.18% in the control and PMR groups, respectively. These accuracies were not significantly different: $t(6) = 0.1728$, $p = 0.8685$.

DISCUSSION

The present results do not extend the findings of Tyndall et al. (2016). Despite rapid and robust learning of the baseline relations, successful adduction occurred in only one of four participants in each group. These results confirm the challenge in adducing untested, three-node equivalence relations (Rippy & Doughty, 2017). Providing participants with brief PMR exposure did not overcome this challenge. As such, the present results limit the generality of brief PMR exposure at enhancing derived relational learning.

The discrepant results between the present research and Tyndall et al. (2016) might be attributed to three procedural differences. First, the number of possible equivalence classes was greater in the present research (i.e., four versus two). Second, the present research involved equivalence relations separated by three nodes, whereas Tyndall et al. examined one- and two-node equivalence relations. Third, Tyndall et al. assessed derived relations using AMTS probe trials, whereas adduction testing was utilized in the present work. Although numerous studies have established derived relations similar to four, 5-member equivalence classes (e.g., Arntzen, 2012; Fields & Moss, 2007; Hayes et al., 2001), it certainly is plausible that these two factors (class number and nodal distance) contributed to the findings. The results of Doughty and Soydan (2019) suggest that the third factor (testing method) may have contributed to the present findings. Two groups of college students received initial training identical to Phase 1 of

Table 4

Accuracy scores (i.e., percent correct) for each participant in each session.

Condition	BM	JR	MD	TH
Control Group				
AB	90	82	98	79
	98	100	98	100
BC	89	81	97	84
	100	100	100	100
CD	95	77	86	97
	99	100	100	100
DE	85	74	95	98
	100	98	100	100
AB – DE	98	96	98	99
		100		
Adduction	99	1	16	0
	BB	EL	GB	OD
PMR Group				
AB	82	40	98	25
	100	86	99	25
		97		92
		98		100
BC	96	86	95	90
	100	100	98	100
CD	95	96	95	91
	100	100	98	100
DE	97	92	94	94
	99	100	98	100
AB – DE	99	98	98	100
		99		
Adduction	93	0	0	0

the current experiment (i.e., AB, BC, CD, and DE relations were established). In Phase 2, derived EA relations were tested across groups using either probe-trial testing or a modified adduction assessment. The modified adduction assessment was identical to the adduction testing in the present experiment with one exception. Four response options (i.e., possible answers) were presented on each arithmetic flashcard surrounding the EA question. As such, both groups in Doughty and Soydan received an assessment in which sample and comparison stimuli were present. Despite this inclusion of comparison stimuli in the adduction assessment, no participant performed successfully in adduction testing, whereas each participant derived the EA relations in probe-trial testing. These findings attest to the relative difficulty inherent in deriving multi-nodal relations in an adduction assessment.

Future research examining the effects of PMR on derived relational responding should explore the aforementioned differences between the present research and Tyndall et al. (2016) as well as address the following limitations. Both Tyndall et al. and the present experiment omitted independent measures of relaxation during and after PMR exposure. The effects of prolonged PMR exposure were not assessed. It may be noteworthy to examine differential levels of PMR exposure across participants with and without a history of engaging in PMR to induce relaxation. An additional factor that may be useful to assess is the treatment of the control groups. Tyndall et al. utilized a simple learning task, whereas our participants were untreated. Finally, our small sample sizes should be noted. It is our hope that investigators continue to follow the lead of Tyndall et al., and others, and examine derived stimulus relations in the context of variables that garner attention in the broader scientific community.

REFERENCES

Andronis, P. T., Layng, T. J., & Goldiamond, I. (1997). Contingency adduction of "symbolic aggression" by pigeons. *The Analysis of Verbal Behavior, 14,* 5-17.

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*, 123-135.

Arntzen, E., Petursson, P. I., Sadeghi, P., & Eilifsen, C. (2015). Equivalence class formation in accuracy or speed conditions: Immediate emergence, adduction, and retention. *The Psychological Record, 65*, 141-159. doi: 10.1007/s40732-014-0097-9.

Barnes-Holmes, D., Finn, M., McEnteggart, C., & Barnes-Holmes, Y. (2018). Derived stimulus relations and their role in a behavior-analytic account of human language and cognition. *Perspectives on Behavior Science, 41,* 155-173. doi: 10.1007/s40614-017-0124-7.

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13,* 140-163.

Cahn, B. R., & Polich, J. (2006). Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychological Bulletin, 132*, 180-211. doi:10.1037/0033-2909.132.2.180.

Chambers, R., Lo, B. Y., & Allen, N. B. (2008). The impact of intensive mindfulness training on attentional control, cognitive style, and affect. *Cognitive Therapy and Research, 32*, 303-322. doi: 10.1007/s10608-007-9119-0.

Chase, P. N. (2003). Behavioral education: Pragmatic answers to questions of novelty and efficiency. In K. A. Lattal & P. N. Chase (Eds.), *Behavior Theory and Philosophy* (pp. 347-367). New York, NY: Plenum/Kluwer.

Critchfield, T. S., Barnes-Holmes, B., & Dougher, M. J. (2018). What Sidman did: Historical and contemporary significance of research on derived stimulus relations. *Perspectives on Behavior Science, 41*, 9-32. doi: 10.1007/s40614-018-0154-9.

Critchfield, T. S., & Twyman, J. S. (2014). Prospective instructional design: Establishing conditions for emergent learning. *Journal of Cognitive Education and Psychology, 13,* 201-217. doi: 10.1897/1945-8959.13.2.201.

Doughty, A. H., & Soydan, J. A. (2019). Differential derived stimulus relations across probe-trial versus adduction testing are not a function of comparison-stimulus presentation. *Behavioural Processes, 103903.*

Dube W. V. (1991). Computer software for stimulus control research with Macintosh computers. *Experimental Analysis of Human Behavior, 9*, 28-30.

Epstein, R., 1987. The spontaneous interconnection of four repertoires of behavior in a pigeon (Columba livia). *Journal of Comparative Psychology, 101,* 197-201. doi:10.1037/0735-7036.101.2.197.

Fields, L., & Moss, P. (2007). Stimulus relatedness in equivalence classes: Interaction of nodality and contingency. *European Journal of Behavior Analysis, 8,* 141-159.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A postSkinnerian account of human language and cognition.* New York: Kluwer Academic.

Hudetz, J. A., Hudetz, A. G., & Klayman, J. (2000). Relationship between relaxation by guided imagery and performance of working memory. *Psychological Reports, 86,* 15-20. doi:10.2466/PR0.86.1.15-20.

Lattal, K. A., & Harzem, P. (1984). Present trends and directions for the future (introduction to special issue). *Journal of the Experimental Analysis of Behavior, 42,* 349-351.

Nava, E., Landau, D., Brody, S., Linder, L., & Schächinger, H. (2004). Mental relaxation improves long-term incidental visual memory. *Neurobiology of Learning and Memory, 81,* 167-171. doi: 10.1016/j.nlm.2004.02.001.

Rippy, S. M., & Doughty, A. H. (2017). Adduction of untested derived stimulus relations depends on environmental complexity. *Behavioural Processes, 143*, 1-3. doi: 10.1016/j.beproc.2017.07.008.

Tyndall, I. T., Howe, B. E., & Roche, B. T. (2016). Exposure to progressive muscle relaxation leads to enhanced performance on derived relational responding tasks. *The Psychological Record, 66*, 213-222. doi: 10.1007/s40732-016-0163-6.

Vyse, S. (2013). Changing course. *The Behavior Analyst, 36,* 123-135. doi: 10.1007/s40614-014-0013-2.

APPENDIX

Below is a transcription of the guided-meditation recording presented to participants in the PMR group in Phase 2:

> Hello. Make yourself comfortable. Sit back and close your eyes. I am going to read out some instructions and I would like for you to follow. Become aware of your breathing. Slowly breathe in and out through your nose. Now, I would like for you to consciously begin inhaling for one, two, three, four. Now hold your breath for one, two, three, four, five, six, seven. And release for one, two, three, four, five, six, seven, eight. We are going to do this two more times. Begin to inhale for one, two, three, four. Now hold your breath for one, two, three, four, five, six, seven. And release for one, two, three, four, five, six, seven, eight. Inhale one, two, three, four. Hold. One, two, three, four, five, six, seven. Exhale. One, two, three, four, five, six, seven, eight. Now on each exhale, I would like for you to say the word "one," to yourself. It is natural for thoughts to come into mind. This does not mean that you are not following the procedure. When this happens, simply deal with the thought, do not dwell on it, but return your focus back to your breathing. Breathing in through your nose and exhaling on one. So now, deeply relax all of your muscles. Starting with your toes, feel them relaxing. All tension easing away. Next, your ankles; completely relaxing, no tension at all. Relax the muscles in your calves. No strain. And your knees. Feel them relaxing. And all of the while, you are breathing in through your nose and exhaling on "one." The muscles in your thighs are completely relaxed. The tension is easing away. Your lower back is totally and completely at ease. Completely comfortable. Feel your stomach muscles relaxing. Everything is easing away and your chest muscles. The tension is leaving them now. You are totally and completely at ease. You are totally and completely at ease. Your hands are completely relaxed. Just resting there. There is no tension in your arms. Completely relaxed. Your shoulders. There is no tension in them at all. Totally at ease. Your shoulder blades- feel them relaxing. Letting everything go. Letting it all go. And all the while, you are breathing in and exhaling on "one." Now focus upon your neck. All strains are now leaving your neck. Completely relaxed. Now, notice your mouth. It is loosening up. Your tongue drops from the roof of your mouth. Your jaw relaxes softly. And your cheeks are relaxing. All easing out. And all of the while, you are breathing in and exhaling on "one." Completely and totally at peace. The lines of your forehead are now disappearing. The tension being held in the top of your head is now being rubbed away and you are feeling completely at ease. The top of your head is totally relaxing. No tension at all. Breathing in and exhaling on one. With each breath, imagining the tension releasing from your body. Exhaling any form of stress and tension that you may have built up during the day. Your whole body is now completely relaxed. You are now totally at ease and continue to relax. Open your eyes whenever you are ready. Someone will be with you in a few moments. Thank you.

EXPERIMENTAL ANALYSIS OF HUMAN BEHAVIOR BULLETIN 2020, 32, 60-65

RESEARCH IN PROGRESS

BEHAVIORAL TECHNOLOGIES OF TEACHING AT 50: NEW OPPORTUNITIES AND NEW CHALLENGES

William J. McIlvane, Christophe J. Gerard, and David A. Smelson

UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL

This paper concerns an ongoing effort at the University of Massachusetts Medical School to develop computer-supported instructional procedures that may contribute to national efforts to address high-priority public health challenges. Recent efforts have yielded publications aimed at supporting practitioners responding to the increasing prevalence of autism spectrum disorders (e.g., McIlvane et al., 2018). Very recently, this effort has taken a new direction – supporting practitioners who provide services to people with co-occurring mental health and substance use disorders (CODs). In describing this new initiative, we hope to call attention to currently neglected research opportunities for behavior analysts.

As background on behavior analytic instructional science, two seminal works were published in 1968. The first was Skinner's *The Technology of Teaching* in which he summarized an emerging behavioral science relevant to teaching. He contrasted its methods and findings with longstanding educational methods used from the primary grades through graduate education. The second seminal work was Keller's Goodbye Teacher, in which he outlined the Personalized System of Instruction (PSI). Keller summarized its main features as follows: (1) learners go at their own pace; (2) learners master one unit's information before advancing to the next; (3) lectures and demonstrations motivate rather than provide critical information; (4) written word is emphasized in teacher-student communication; (5) proctors conduct repeated testing, immediate scoring, and tutoring.

Via these two works, Skinner and Keller helped launch the field of programmed instruction. One revolutionary perspective was foundational: the teacher's responsibility is to arrange conditions under which students truly learn the material. Instead of merely grading students against whatever standard they preferred, teachers would become responsible for "grading" themselves against the standard of student achievement. If teaching proved ineffective or inefficient, then it was the teachers' responsibility to improve it. In doing that, Skinner and Keller expected that teachers would apply the best information available to refine their teaching, to incorporate new methods that promised improved outcomes, and over time perfect their teaching to the extent possible.

Behavior analysis has many successful, sustained applications in education and training of normally capable children and people with developmental disabilities. For example, The Journal of Behavioral Education publishes many such reports. In a review of its offerings overall, we found relatively few projects with normally capable adults. Concerning behavior analytic programmed instruction as a whole, many studies with normally capable adults were published prior to 1990. In this century, however, examples are also relatively few. Literature relating to PSI implementation is similar.

One is led to ask why methods with a substantial basis in empirical evidence have made so little recent impact on education and training programs for college students and normally capable adults in general. This situation seems

Work reported here was supported by NIAAA Grant AA026751 and an Intergovernmental Personnel Agreement with the United States Veterans Administration. Neither agency was involved in formulation, design, or execution of project. We thank Ayorkor Gaba, Jennifer Harter, and Brooks Thompson for their contributions in its design and executionCorrespondence to: William J. McIlvane, UMass Medical School, 55 Lake Ave. North, Worcester, MA 01566. E-mail: william.mcilvane@umassmed.edu.

especially puzzling given the explosion of distance learning programs that are excellent vehicles for programmed instruction. Concerning PSI specifically, social medial platforms could be used to implement its cooperative learning features, but such projects seem virtually nonexistent (cf. Eyre, 2007; see Svenningsen, Bottomley & Pear [2018] for an example of computer-assisted instruction with PSI features).

To account for neglect of such opportunities, systems inertia and intellectual and/or emotional opposition come to mind. However, there is the highly plausible alternative that programmed instruction and PSI are perceived as dated approaches. Most foundational work on these topics was accomplished decades ago. Moreover, much of the technology employed back then was primitive by today's technological standards. Developers today might be excused for evaluating such technology as obsolete.

In one effort to correct this misperception, our project on CODs was inspired by the Sidman and Sidman (1965) Neuroanatomy programmed text, the result of a partnership between a neurologist and a behavior analyst. Our project partners an expert on health care services with a behavior analyst and a computer scientist. As background, CODs are common. They challenge providers to provide effective treatments, especially for individuals who are homeless and/or involved in the criminal justice system. Research has shown that integrated, coordinated community-based services have the potential to improve client outcomes. However, there is a clear gap in training the health care workforce on such treatment techniques – a gap we hope to begin closing.

Our project is focused on developing a programmed distance-learning course called MISSION U. Its purpose is to teach practitioners about an evidence-based, transdisciplinary case management model called Maintaining Independence and Sobriety through Systems Integration, Outreach and Networking (MISSION). Listed in the Substance Abuse and Mental Health Service Administration-Registry for Evidence Based Practices (NREPP), MISSION has had substantial impact, particularly with vulnerable populations of veterans and individuals who are homeless and/or involved in the criminal justice system. While MISSION is extensively documented in treatment manuals, a recent multisite implementation study identified the need for more comprehensive training tools (Smelson et al. 2015).

The first phase of the MISSION U project aimed to: (1) demonstrate implementation of a prototype MISSION U software platform with meaningful MISSION content and (2) create a MISSION overview that can be freely distributed to exemplify the instructional technology. We programmed two prototype modules, one a general introduction to the MISSION model and the other on MISSION home visit procedures. Both modules were broken up into smaller units that allowed for frequent learning evaluations and remedial programming.

COURSE DESIGN

Instructional Design

The Individualized Instruction Model (IIM) (or "personalized instruction") specifies that learners should be supported in completing work autonomously and accurately, focusing on their specific capabilities and need areas (Pappas, 2014). This model is directly traceable to behavior analysis research with normally capable adult learners in Skinner (1968) and Keller (1968). In this model, didactic teaching contains frequent assessments to gauge ongoing progress. Tests assess not only acquisition but also application of new material. Our course design draws also from behavior analytic analyses of acquisition and generalization of new behavior (e.g. McIlvane & Dube, 2003).

MISSION U is designed primarily for Case Managers and Peer Support Specialists who deliver and/or coordinate services for persons with COD. These learners are heterogeneous in age, ethnicity, income, SES, level of education, job responsibilities, career objectives, and first language – heterogeneity that virtually demands use of the Individualized Instruction Model.

The design has been influenced also by current teaching principles that posit that some individuals learn best visually whereas others do best when the material is presented in the auditory mode. There is evidence that (1) learners may have distinct preferences for the manner in which information is consumed and (2) multimodal (e.g., combined auditory-visual) presentation has demonstrated advantages (cf. Kharb et al, 2013). Notably, these principles comport well with those of quasi-basic behavioral research with other populations (Green, 1990; Soraci et al., 1991).

Figure 1

Figure 1 shows a simple technique for multimodal guidance of attending. Its bottom portion shows a full frame from the module on home visit procedures. As the frames above it indicate, the constituent information is introduced gradually and coordinated with instructor voice-over narration that conveys the same information simultaneously. The full frame is presented only at the end when the learner is invited to review it before going forward in the program.

Good management of attending also aims to reduce the scanning burden and thus to increase the likelihood of attending to relevant content. To encourage generalization of skills learned in the didactic components of MISSION U, we also incorporate e-simulations in which a Case Manager and a Peer Support Specialist simulate work with a client. Images like those shown in Figure 2 are interspersed throughout the audio-visual e-simulation. Periodic challenge questions require learners to make operational and clinical decisions in which they must apply what they have learned. These questions contrast specific MISSION techniques with clinically plausible choices that do not comport with the MISSION model. Immediate feedback is given, and learners are given an opportunity to respond correctly afterward if errors occurred.

Figure 2

Pretest/Posttest Design

Given the target audience, we have been especially careful to verify that MISSION U instructional procedures are sufficient to teach its learners about the MISSION program specifically (i.e., as opposed to merely recruiting general knowledge and/or opinion about how clients with CODs should be treated). The example in Table 1 shows that pre/post testing contrasts the MISSION model with other plausible clinical options. If learners respond correctly to such posttest questions, they show that they have understood and remembered what was taught during the teaching units. Such questions do not allow them to respond consistently correctly by guessing or by applying previous knowledge about other therapeutic approaches. Challenge questions in the e-simulations have a similar purpose.

Technical Assistance

Consistent with certain aspects of the Keller Plan, our design incorporates an optional 1-hour videoconference that allows learners to interact with instructors who are fully conversant with the MISSION model. These videoconferences allow learners to discuss the e-simulation case and also related topics from their own case management or peer support experience. We plan to expand such supports by including proctor/peer tutoring by Case Managers and Peer Support Specialists in subsequent versions of MISSION U.

Software Architecture

MISSION U is a hybrid of a commercially available e-learning package (iSpring, n. d.) and in-house programming that substantially expands the package's capabilities to meet the needs of our instructional design. In our in-house programming to date, we have emphasized off-the-shelf Open Source components that provided powerful, transparent, non-proprietary resources to give us maximum power and flexibility. In addition, we also developed new software that allows these components to interact functionally.

Consistent with our Open Source approach, our course is currently implemented in a widely used Learning Management System called Moodle. Its features allow for housekeeping such as learner registration, enrollment, and data reporting. Moodle also allows for limited program branching based on learner responses to challenge questions. However, this feature is not sufficiently flexible to optimize the instructional flow for individual learners, especially those who struggle with the material. Thus, our technical development group has developed proprietary software that permits virtually unlimited branching within the Moodle or other Learning Management System environments.

PILOT STUDY OF MISSION U

Nineteen participants were recruited from programs serving clients in 5 states. The racially diverse sample had an average of 8.8 years in the field. All participants received pre/posttests for all units and all completed the e-simulation. The technical assistance opportunity was piloted with nine participants. Participants then completed software satisfaction/content ratings and a final overall post-test. Four results were clear:

(1) Instructional technology delivered highly effective instruction. Even given our demanding question design, pretest/posttest score distribution differences were virtually nonoverlapping visually and thus highly significant ($p < .0000000000001$) based on statistical analysis with the Excel t-test function (paired sample, one-tailed test).

(2) In the e-simulation, there was significant evidence of application of MISSION principles. Accuracy scores on challenge questions averaged 86% (range: 100%-71%), significantly greater than chance scores (~33%).

(3) Technical Assistance was clearly beneficial – and very positively received. On a final overall post-test, scores of those who received technical assistance were substantially higher than those who did not. The Cohen d effect size was large (.833).

(4) Learner satisfaction with MISSION U was high. One satisfaction measure concerned usability of the MISSION U human interface and the other whether MISSION U course content was appropriate to learner needs. Percent satisfaction with usability and content was 94% and 93%, respectively. Disagreements were spread across the range of questions posed. Thus, no

---

[1] For readers interested in technical details, the program infrastructure is based on: (1) MEAN, an open-source JavaScript software stack for building dynamic web applications. MEAN is an acronym abbreviating its components: (a) MongodB – a free, open source, cross-platform document-oriented database program, (b) ExpressJS – a web framework for NodeJS; see d, (c) AngularJS – a client-side JavaScript framework that extends HTML with new attributes, and (d) NodeJS – an open-source, cross-platform JavaScript run-time environment that executes server-side JavaScript code, and (2) a Nginx Web Server that provides static content for Learning Modules and static resources (html, css and JavaScript files).

specific feature of the human interface design or course content was deemed faulty by users as a group.

CURRENT STATUS OF THE PROJECT

As impressive as these data might seem, we think these findings merely show promise that we are on a path towards a programmed course that would fully satisfy criteria articulated by Skinner, Keller, and Sidman. While most pretest scores were at or near chance levels as we had planned, the final posttest score distribution ranged from intermediate to high accuracy. We could likely have produced high accuracy posttest scores overall by making the posttest questions less demanding, but that would have defeated one important purpose of the pilot – fair witnessing that our procedures for managing attending and other aspects of instructional technology were responsible for posttest gains.

Even though the participants in our pilot were normally capable adults and we gave them an unusual amount of instructional support for training such as this, we were not surprised that posttest scores fell short of perfection. We recalled Holland's Forward II to The Technology of Teaching (2003). His students (Harvard undergraduates) apparently performed similarly on the first efforts of his course development group. We recall also the lessons from the Sidman and Stoddard (1966) description of program development for testing patients with neurological and neurodevelopmental disorders in which many revisions of an initially well-designed teaching program proved necessary before a satisfactory version was produced. Work of this nature virtually demands cycles of program development, testing, and refinement. We have already completed a substantial revision of one of our learning modules that we plan to release when we optimize learning outcomes.

CONCLUSION

We prepared this preliminary report mainly to highlight a research and development opportunity that we hope will attract the attention of and perhaps challenge our colleagues and students. With today's hardware and software, development of carefully programmed courses for normally capable adults has become a feasible and affordable proposition. For our work, we chose iSpring based on program features that we needed for our project. However, there are many such systems to choose from with different features, strengths, weaknesses and price points (cf. Capterra Course Authoring Software, n. d.). If custom programming is needed for certain applications, we and our colleagues have had some success in recruiting low-cost student help from university computer science and engineering programs. Involvement of their faculty in such collaborations has been of particular help.

Whereas the technologies described by Holland and by Sidman and Stoddard led to slow, painstaking work to produce a single program, current technology presents the opportunity to develop much more capable instructional technology in much less time. Although we had resources to do custom programming, current authoring programs have powerful features and built-in tools to aid developers. It is increasingly possible to realize the early vision of Skinner, Keller, Sidman, and others who foresaw a powerful behavioral instructional technology to improve education and training outcomes for a broad range of students. Indeed, we think that a new Golden Age of programmed instruction is within reach.

LITERATURE CITED

Capterra Course Authoring Software (n.d.). Retrieved from https://www.capterra.com/course-authoring-software/.

Eyre, H. L. (2007). Keller's Personalized System of Instruction: Was it a fleeting fancy or is there a revival on the horizon? *The Behavior Analyst Today*, *8*(3), 317-324. http://dx.doi.org/10.1037/h0100623.

Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation*, *95*, 260-270

Holland, J. G. (2003). Forward II to The Technology of Teaching. Retrieved from www.bfskinner.org/wp-content/uploads/2016/04/ToT.pdf

iSpring Suite 9.7 (n. d.) Retrieved from www.ispringsolutions.com/ispring-suite.

Keller, F. S. (1968). Goodbye, teacher. *Journal of Applied Behavior Analysis*, *1*, 78-89.

Kharb, P., Praina, P. S., Jindal, M., & Singh, V. (2013). The learning styles and preferred teaching-learning strategies of first year medical students. *Journal of Clinical and Diagnostic Research*, *7*, 1089-1092.

McIlvane, W. J. & Dube, W. V. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst*, *26*, 195-213.

McIlvane, W. J., Kledaras, J. B, Gerard, C. J., Wilde, L., & Smelson, D. (2018). Algorithmic analysis of relational learning processes in instructional technology: Some implications for basic, translational, and applied research. *Behavioral Processes*, *152*, 18-25.

Pappas, C. (2014, Oct 30). Instructional design models and theories: Individualized instruction model. [online]. https://elearningindustry.com/individualized-instruction-model. Downloaded Nov. 23, 2018.

Sidman, R. L. & Sidman, M. (1965). Neuroanatomy: A Programmed Text, *Vol. 1*. Lippincott Williams & Wilkins: Baltimore.

Sidman, M., & Stoddard, L. T. (1966). Programming perception and learning for retarded children. In N. R. Ellis (Ed.), International Review of Research in Mental Retardation, *Vol. 2* (pp. 151-208). New York: Academic Press.

Skinner, B. F. (1968/2003). The technology of teaching. Appleton–Century-Croft; New York. Retrieved from www.bfskinner.org/wp-content/uploads/2016/04/ToT.pdf.

Smelson, D. A., Chinman, M., McCarthy, S., Hannah, G., Sawh, L, Glickman, M. (2015). A cluster randomized Hybrid Type III trial testing an implementation support strategy to facilitate the use of an evidence-based practice in VA homeless programs. *Implementation Science*, *10*, 1-10.

Soraci, S. A., Jr., Deckner, C. W., Baumeister, A. A., Bryant, J. T., Mackay, H. A., Stoddard, L. T., & McIlvane, W. J. (1991). Generalized oddity performance in preschool children: A bimodal training procedure. *Journal of Experimental Child Psychology*, *51*, 280-295.

Svenningsen, L., Bottomley, S., & Pear, J. J. (2018). Personalized learning and online instruction. In Robert Zheng (Ed.), Digital technologies and instructional design for personalized learning (pp. 164-190). Hershey, PA: IGI Global. doi: 10.4018/978-1-5225- 3940-7.ch008.