EXPLORING AN ALTERNATIVE FOR RUNNING EXPERIMENTS DURING SOCIAL DISTANCING: A PILOT STUDY OF STIMULUS EQUIVALENCE

Marcelo V. Silveira[1,4], Heloísa R. Zaparolli[2], Guilherme Sbrocco[2], Mateus Silvestrin[1], Ohana T. Rabelo[3], Kamilla Tenório[4], Leonardo B. Marques[3], Diego D. M. C. Matos[4], Erik Arntzen[5], Marcelo S. Caetano[1,6]

[1] CENTER FOR MATHEMATICS, COMPUTING AND COGNITION, UNIVERSIDADE FEDERAL DO ABC (UFABC)

[2] UNIVERSIDADE FEDERAL DE SÃO CARLOS;

[3] DEPARTAMENT OF EDUCATIONAL SCIENCES, UNIVERSIDADE FEDERAL DE ALAGOAS;

[4] FACULTY OF MEDICAL SCIENCES, UNIVERSIDADE FEDERAL DE ALAGOAS;

[5] DEPARTAMENT OF BEHAVIORAL SCIENCES, OSLO METROPOLITAN UNIVERSITY;

[6] INSTITUTO NACIONAL DE CIÊNCIA E TECNOLOGIA SOBRE COMPORTAMENTO, COGNIÇÃO E ENSINO (INCT-ECCE)

Web-based platforms were useful tools for running psychology experiments remotely during the pandemic. They enabled the collection of behavioral data while keeping participants and researchers safe from infection threats imposed by the coronavirus crisis. This study focused on the generality of equivalence outcomes typically observed in basic research laboratories. Participants accessed the experimental task remotely and were trained in a matching-to-sample (MTS) task. The procedure was designed to establish conditional discriminations between (A) pictures of women expressing happiness and angriness (A1 and A2), and (B, C) several abstract shapes (B1, B2, C1, and C2). Upon completion of one-to-many training, participants were tested on MTS-probe trials designed to assess the non-trained equivalence relations BC (B1C1 and B2C2) and CB (C1B1 and C2B2). Finally, two semantic judgment blocks of trials assessed equivalence-priming effects for the B and C stimulus-stimulus relations and priming effects for semantically related words. Results showed the expected patterns based on equivalence class formation, that is, emergent BC and CB relations following training of AB and AC conditional discriminations, and semantic priming effects were observed in the semantic judgment trials. In general, these results suggest that studies on stimulus equivalence may be conducted by means of web-based platforms.

Keywords: equivalence relations; web-based experimentation; matching-to-sample; semantic judgment; emergent behavior; COVID-19

In most developing countries, the number of infected people and deaths by COVID-19 was far greater than the world average (WHO, 2021). In Brazil, for example, the pandemic persisted more than it could have due to ineffective strategies adopted by the federal government to mitigate or interrupt the spread of the new coronavirus (cf., Kissler et al., 2020). Therefore, social distancing measures and other restrictions to group gatherings were kept from March 11th 2020 to April 20th 2022 according to Brazilian federal law #13.979.

Considering all those measures and restrictions, the COVID-19 pandemic considerably impacted scientific endeavors in Brazil. Between 2020 and 2022, in-person data collection was restricted and research production and publication have decreased unprecedentedly (Raynaud et al., 2021). This unfortunate situation had an impact on the ability of Brazilian institutions and researchers to

Author Note: This study was part of MVS's post-doctoral research at UFABC, which was supported by the São Paulo Research Foundation (FAPESP, postdoctoral fellowship #2017/06178-7). MVS also thanks FAPESP for the special funding Bolsa de Estágio em Pesquisa no Exterior (BEPE, Fellowship #2019/14661-5). HRZ has a doctoral scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Grant#88887.629693/2021-00). GS has a doctoral scholarship from São Paulo Research Foundation (FAPESP, Grant #2021/13212-2). MVS has a doctoral scholarship from the São Paulo Research Foundation (FAPESP; fellowship #2016/24951-2). OTR has a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, #2021SLRO1822). MSC is funded by the São Paulo Research Foundation (FAPESP, Grant #2019/25795-2).

All authors thank Dr. André M. Cravo for thoughtful comments on our data. This project is supported by Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino (INCT-ECCE), funded by the São Paulo Research Foundation (FAPESP, Grant #2014/50909-8), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Grant No. 88887.136407/2017-00), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant #465686/2014-1). We thank Deisy de Souza, chairperson of INCT-ECCE. Please address correspondence to MVS, Rua Abrahão João, 55, Casa 16, São Carlos – SP, CEP 13562-150, Brazil. e-mail: marcelopsi06@gmail.com.

https://doi.org/10.17605/OSF.IO/SXWA5

attract new graduate students, establish collaborations, and secure grants and financial support (see Oliveira et al., 2021, for a broad review and relevant discussion on this topic). In this scenario, the possibility of conducting experiments remotely proved to be a short-term solution for human behavior researchers. It allowed them to continue with data collection despite the measures for social distancing.

Remote data collection does not involve significant physical risks to participants or investigators. However, the platforms used to collect data might not be as effective in replicating the behavioral results as in-person procedures already established in the literature. Crump et al. (2013), for example, used an online crowdsourcing service called Amazon Mechanical Turk (AMT or MTurk) to implement a diverse body of tasks from experimental psychology. Even though some experiments were successful in replicating the data available in the literature, in their Category Learning Task (Experiment 8) participants took nearly twice the number of blocks to reach the criterion when compared to Nosofsky et al.'s (1994) laboratory study and also showed a higher probability of making errors.

A robust line of research in the Behavior Analysis tradition had begun with a study of Murray Sidman and his colleagues by the year of 1971. His pioneer work on emergent behavior in language-disabled persons bears on a behavioristic theory of meaning comprehension and other language-like behaviors. The experiments conducted by Sidman et al. (1982) with children, monkeys, and baboons provided the empirical basis for the development of the Stimulus Equivalence paradigm. This paradigm has provided ways to test some properties (reflexivity, symmetry, and transitivity) that enable the discrimination between equivalence relations from other forms of complex discriminated operants (e.g., paired-associated relations, behavioral chaining and conditional discriminations) (Pilgrim, 2016; Sidman, 1994, 2000). That is, the training of relations ArB and ArC (i.e, "r" is the relation between stimuli "A", "B" and "C") by reinforcement contingencies, is followed by the emergence of relations between each stimulus and itself (i.e., ArA, BrB and CrC), between stimuli in the opposite trained direction (i.e., BrA and CrB), and also between stimuli that were never directly trained before (i.e. BrC and CrB) (Sidman & Tailby, 1982). When one or more of those relations is demonstrated, equivalence classes are established, with stimuli that can be interchangeable in some contexts. Since that seminal work, the Stimulus Equivalence Paradigm has been used as a model to study complex behaviors such as reading and comprehension (de Rose et al., 1992; Sidman, 2018), clinically-relevant behaviors (Augustson & Dougher, 1997), racial and social prejudice (Mizael et al., 2021; Watt et al., 1991), and rule-governed behavior (Hayes et al., 1989).

Cognitive researchers (Meyer & Schvaneveldt, 1971; Neely, 1991) traditionally use a procedure called the lexical decision task when analyzing symbolic relations. With this task, it is possible to identify an effect known as the semantic priming effect. Basically, this effect is observed when a participant recognizes a semantically related stimulus more quickly than an unrelated one. For example, if the participant is presented with a prime word and then with a semantically associated target word (e.g., tiger-lion), the recognition of the target word occurs faster than if presented with a non-associated target word (e.g., tiger-house).

The semantic priming effect was broadly investigated using words in the participants' natural language (see Neely, 1991, for a review). However, the lexical decision task has also been employed to demonstrate analogous semantic priming effects evoked by pseudo-words (Barnes-Holmes et al., 2005; Hayes & Bisset, 1998) and abstract stimuli (Bortoloti & de Rose, 2011) in accordance with their class-memberships. Bortoloti and de Rose (2011), for example, trained associations between faces expressing happiness, angriness, and neutrality (A1, A2 and A3) and abstract black and white drawings, randomly assigned to three groups B, C and D (i.e., stimuli B1, B2, B3, C1, C2, C3, D1, D2 and D3) using the simultaneous matching to sample (SMTS) procedure. In this computerized task, each trial started with the presentation of a sample stimulus in the middle of the screen. A click on that stimulus produced three comparison stimuli in three different corners of the screen. The sample stimulus remained on the screen while the comparison stimuli appeared. Participants should choose one of the comparison stimuli, depending on the sample, by clicking on it. For example, in one trial, a given sample stimulus A1 and comparison stimuli B1, B2, and B3 are shown. A correct response (B1, in this example) produces positive feedback in the form of a sequence of tones and an arrangement of stars moving on the computer screen. Incorrect responses (B2 or B3) are just followed by a 0.75-s intertrial interval. In the next trial, A3 might be the sample stimulus, for which a response on B3 is correct, and so on. This is an "AB" type of training (stimuli from set A are samples, stimuli from set B are comparisons).

As Bortoloti and de Rose's (2011) procedure continued, the participants (young adults) completed AB, AC, and AD training. Upon completion of training of these relations, participants did a lexical decision task. Accuracy and reaction times were measured. Participants had to press "s" on the computer keyboard when two related stimuli were presented (e.g., B1-C1, B2-C2, D1-B1) and press "n" when unrelated stimuli were shown (e.g., B2-C1, B3-C2, D3-B1) (see also Barnes-Holmes et al., 2005). Then, participants were exposed to a series of matching to sample (MTS) probes that assessed the emergence of BD and DB associations. Bortoloti and de Rose (2011) reported robust differences in accuracy and reaction times for the related and unrelated stimulus-stimulus pairs. This semantic effect can be interpreted as evidence of the establishment of equivalence relations by the MTS procedure, suggesting that meaning comprehension of words in naturally occurring languages is based on the same basic mechanisms.

Considering the importance of maintaining research projects during scenarios of social distancing, especially in the areas presented above, the present study investigated the effectiveness of reproducing Bortoloti and de Rose's (2011) laboratory experiment in an online version. A remote access application allowed participants to access the experimental tasks remotely. A PsychoPy (Pierce et al., 2019) script running in the experimenter's computer provided instructions and stimuli and recorded all experimental events. Participants were trained on conditional discriminations between pictures of human faces expressing emotions (A1 and A2) and abstract black and white forms (B1, B2, C1, and C2), and were tested for untrained transitivity associations BC and CB (Arntzen, 2012; Sidman, 1994). Compared to Bortoloti and de Rose (2011) study, a smaller number of classes (i.e., two instead of three) and a smaller class size (i.e., with three stimuli each, not four) were used to simplify the procedure and enhance the probability of class formation in the present study. Semantic judgment trials assessed priming effects for stimuli sharing common class membership (Barnes-Holmes et al., 2005; Bortoloti & de Rose, 2011; Hayes & Bisset, 1998) and for semantically related words from the participants' natural language. The total of correct responses in each conditional discrimination training and equivalence test and the number of participants that achieved each accuracy level were used to assess the reliability of the data obtained. Response time in the semantic judgment task was also analyzed. Our focus was to simulate the laboratory conditions in which real-time data acquisition occurs, allowing participants and experimenters to interact in real-time albeit from a physical distance. Similar to the laboratory studies, we expected the successful establishment of equivalence relations, reflected by faster response times (RTs) and higher accuracies in the related than in the unrelated pairs.

METHOD

Participants

Twenty-two adults between 18-35 years old were recruited via Facebook advertisements between April and October of 2021. Participants received an e-mail containing the following instructions:

> Thank you for your interest in participating in this experiment. To ensure everyone's safety and to collaborate with the distancing measures crucial for containing the COVID-19 Pandemic from progressing, all procedures will be performed remotely. You shall use your computer or laptop in the quietest place at your home. You shall also ask your relatives to be left alone for approximately 20 to 30 minutes. You will find the consent form attached to this e-mail. You should read the document, fill it out, and return it to the experimenter BEFORE the beginning of the experiment. You can reply to this email with questions at any time before you fill out the Consent Form. However, the experiment will not take place without your full consent to participate.

All instructions were provided in Brazilian Portuguese (see Appendix 1). The procedures were approved by the University Human Research Ethics Committee.

Pre-Experimental Instructions, Applications, and Stimuli

All participants were instructed to find a comfortable, illuminated, and silent room with at least a chair and a table to participate in the experiment. They should remain seated, use headphones, a computer mouse, and a keyboard during the individual experimental session (approximately 20 min). They had to download the AnyDesk® application to access the experimental tasks remotely. The application provided a fast and high-quality image presentation and response-time precision. The latency between the participant response and the recording in the experimenter's computer was below 16 ms. The experimenter's computer ran PsychoPy3 (v2020.1.1) (Pierce et al., 2019), a software also used for in-person computerized data collection that presented all the instructions, visual and auditory stimuli, and recorded all responses.

Instructions were shown in the computer screen throughout the different phases of the experiment. Short paragraphs were presented in different frames and remained available until the participant pressed the spacebar key, after which the next paragraph was shown. The experimenter informed the participants that he would be away from the computer while the experiment was running and that they would be instructed to call the experimenter by the end of the task.

The visual stimuli used in the study were a happy and an angry facial expression, and abstract shapes presented in Figure 1. Auditory stimuli were a ring tone and a buzzer available from the PsychoPy3 libraries.

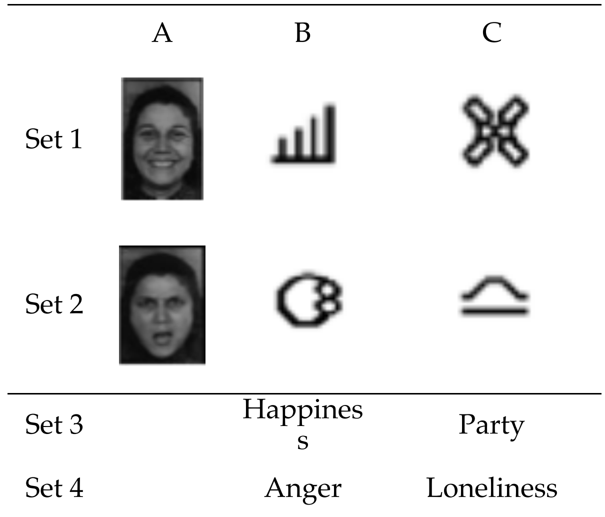

Figure 1. Stimuli Used in the Experiment. *Note.* Pictures of faces expressing happiness and angriness (A1 and A2) and the abstract forms B (B1 and B2) and C (C1 and C2) used in MTS procedures. The forms B and C and the words at the bottom portion of the figure were used in the Semantic Judgment blocks. In the experiment, the words were shown in Brazilian Portuguese.

Procedure

Conditional Discrimination Training and Formation of Equivalence Classes. Instructions were provided before the beginning of the training and test phases. The conditional discriminations (i.e., AB and AC) were trained in two blocks. The initial block consisted of four trials of the SMTS procedure, in which all stimuli remained available until the participant selected one of the comparison stimuli with a mouse click. Each of these four trials began with the presentation of the faces A1 or A2 as the sample stimulus. A click on the sample produced the comparison stimuli at the bottom corners of the screen. To enhance equivalence class formation, during the SMTS training block, grayed-out versions of the incorrect B and C comparison stimuli were used, whereas the sample stimulus and correct B or C comparison appeared in their original black and white colors (Silveira et al., 2021). Correct (class-consistent) responses resulted in the removal of all stimuli from the screen and presentation of a ring bell with the word "Correct!" in green letters for 1.5 s. The selection of the incorrect comparison stimulus resulted in the presentation of a buzz sound with the word "Wrong!" in red letters for 1.5 s. After the STMS training block, a second block of 12 trials with a delayed MTS procedure (0 s DMTS) was conducted. During this block, a click on the sample stimulus resulted in its removal from the computer screen, followed by the immediate presentation of the comparison stimuli. All trials were separated by an intertrial interval (ITI) of 550 ms. The positions of the comparison stimuli in each corner of the computer screen varied randomly.

The baseline conditional discriminations were established in three distinct stages. The AB and AC conditional discriminations were trained during the first and second stages, respectively. The third stage comprised a 16-trial DMTS block with intermixed AB and AC trials. Participants progressed through the training stages regardless of their performance in each block.

Then, in the fourth and fifth stages, a series of MTS-probe trials with no feedback assessed BC and CB equivalence relations, respectively. The sequence of MTS training and testing stages are shown in Figure 1.

Semantic Judgment Tasks. **Practice Trials**. Participants were given detailed instructions on the sequence of stimuli presentation, response keys, and consequences for correct, incorrect, and late responses. A single 24-trial priming block with black-and-white drawings of celestial bodies (star and planet Saturn) and hand tools (splitting maul and handsaw) were used in practice trials. The events in this session were presented in the following sequence: (i) fixation cross (+), (ii) the prime stimulus, (iii) the target stimulus, and (iv) a question mark. These four events were presented for 1,000 ms each. Participants had to use the A and L keys to indicate whether the prime and target stimuli were related or unrelated, respectively. Related figure-figure combinations were 'star-planet Saturn', 'planet Saturn-star', 'splitting maul-handsaw', 'handsaw-splitting maul'. Unrelated combinations were 'star-handsaw', 'planet Saturn-splitting maul', 'handsaw-planet Saturn', and 'star-splitting maul'. Each correct response was followed by a bell tone for 1,000 ms with the text "Correct!" in green letters and each incorrect response was followed by a buzz for 1,000 ms with the text "Wrong!" in red letters. If participants did not respond within 2,000 ms from the presentation of the question mark, the sentence "Respond faster!" in white letters appeared on the computer screen for 1,000 ms. A total of 12 related and 12 unrelated stimulus-stimulus pairings were randomly intermixed in this practice block. All participants proceeded to the next phase upon completion of a single 24-trial practice block and were given the same priming block sequence (see Table 1). They did not have to achieve any specific criterion of correct responses to proceed.

Semantic Judgment Task with Arbitrary Bs and Cs Stimuli. After the practice block, participants were given the following instructions: "Now that you know how to perform the task, do the same with the abstract figures as you have previously." Participants were given one, 40-trial semantic judgment task block comprising 20 presentations of related (i.e., B1-C1, C1-B1, B2-C2 and C2-B2) and unrelated (i.e., B1-C2, C1-B2, B2-C1 and C2-B1) pairs. Each of the related and unrelated pairs was presented an equal number of times. The stimuli were presented in the following sequence: (i) fixation cross in the center of the screen for 500 ms, (ii) the prime stimulus in the center of the screen for 1,250 ms, (iii) the target stimulus for 1,250 ms, and (iv) a question mark for 1,500 ms.

Procedure		Expected Keyboard Response
Practice block	**Related** Saturn-Star Star-Saturn Handsaw- Maul Maul-Handsaw	Press "A"
	Unrelated Saturn-Handsaw Maul-Saturn Star-Maul Handsaw-Star	Press "L"
Semantic judgment task with arbitrary stimuli	**Related** B1-C1 C1-B1 B2-C2 C2-B2	Press "A"
	Unrelated B1-C2 C2-B1 B2-C1 C1-B2	Press "L"
Semantic judgment task with words	**Related** Happiness-Party Party-Happiness Anger-Loneliness Loneliness-Anger	Press "A"
	Unrelated Happiness-Anger Anger-Happiness Party-Loneliness Loneliness-Party	Press "L"

Table 1. Sequence of semantic judgment block, expected responses, and number of trials.

Participants were required to press the A or L keys on the computer keyboard to indicate whether the prime and target were related or unrelated based on stimulus class memberships. Pressing the A or L keys was considered correct following presentations of the class-consistent relations B1-C1, B2-C2, C1-B1 and C2-B2 (consistent trials) or the presentation of the class-inconsistent relations B2-C1, B1-C2, C2-B1 and C1-B2 (inconsistent trials), respectively. Pressing the A or L keys on the inconsistent or consistent trials respectively was considered as incorrect responses. During this block, there were no programmed consequences for correct and incorrect responses, and no consequences for slow responses to the question mark.

Semantic Judgment Task with Words. To compare the priming effect of intra-experimental derived equivalence relations with the priming effect produced by naturally occurring semantically related words, another stimulus type (i.e., words instead of pictures) was used in an additional word-based semantic judgment block of trials. The following instructions were provided: "We are almost done. Before finishing, repeat the previous task replacing the abstract figures with everyday language words: happiness, party, loneliness, and anger. Evaluate whether the word pairs are related (by pressing the A key), or whether they are unrelated (by pressing the L key). Press the spacebar to start". Positive- (happiness and party) and negative-valence (anger and loneliness) words appeared as prime and target stimuli during this block, as the positive- and negative-valence abstract stimuli pertaining to Sets 1 and 2, respectively, have been used in the semantic judgment task with arbitrary stimuli Bs and Cs. All other parameters were identical to those used in the previous test block.

Post-Experimental Procedures

Upon completion of the semantic judgment tasks, the computer informed the participant that the experiment was over and that the participant was allowed to chat with the experimenter and receive explanations about the experimental analysis of language-like behaviors.

Statistical Analysis

Reaction Time (RT) and Accuracy (ACC) in the semantic judgment tasks were modeled with a two-stage multilevel procedure to accommodate the repeated-measures nature of the data (Gelman, 2017). In the first step, we fitted a generalized linear model (GLM) to the data from each participant, given the non-normal distributions of the dependent variables. In the second step, the coefficients from the fits were tested with one-sample t-tests to identify significant effects. For instance, the effect of congruence on RTs was first estimated for each participant, then the values of the 22 coefficients (one for each participant) were tested to ascertain if, on average, they were different from zero. The link function for the RT model was selected by comparing the Bayesian Information Criterion (BIC) of fits with gamma and inverse Gaussian links using a paired t-test. In principle, both links are adequate to model RT data. However, by selecting the link with better fit to the data, we ensure superior accuracy for the model and subsequent comparisons. For both groups, the gamma function showed a better fit (both p's < 0.02) (cf., Lo & Andrews, 2015).

Two models assessed the effects of RT and ACC: 1) the predictors were congruence (i.e., related vs. unrelated with the previous stimulus) and stimulus type (i.e., abstract forms or words) (all trials, 80 per participant); 2) the predictors were valence (i.e., positive vs. negative) and stimuli relatedness (related pairs, 40 trials per participant). A contrast matrix allowed interpretation of the expected effects in a repeated-measures two-way ANOVA. All analyses were performed in R (version 4.0.3), and effect sizes were obtained using the effectsize package (Ben-Shachar et al., 2020, version 0.8.1). The same approach and models were used for the analysis of accuracy, with the only differences being the use of a log link function (because of the binomial distribution of accuracy data) and the use of the correct/incorrect vector as the dependent variable.

RESULTS

Conditional Discrimination Training and Formation of Equivalence Classes

The participants demonstrated high levels of accuracy on the conditional discrimination training blocks. Correct responses ranged between 14/16 and 16/16 in the AB training and between 12/16 and 16/16 in the AC training. In the AB and AC intermixed training block, most of participants scored 16/16 or 15/16, and just four of them scored 10/16 or fewer. Similarly, most of the participants demonstrated high levels of accuracy on the equivalence-probe trials. In the BC test block, 14 participants scored between 11/12 and 12/12, and just eight of them scored 10/12 or fewer. In the CB test block, 19 participants scored between 11/12 and 12/12 trials, while the remaining three participants scored 10/12 or fewer.

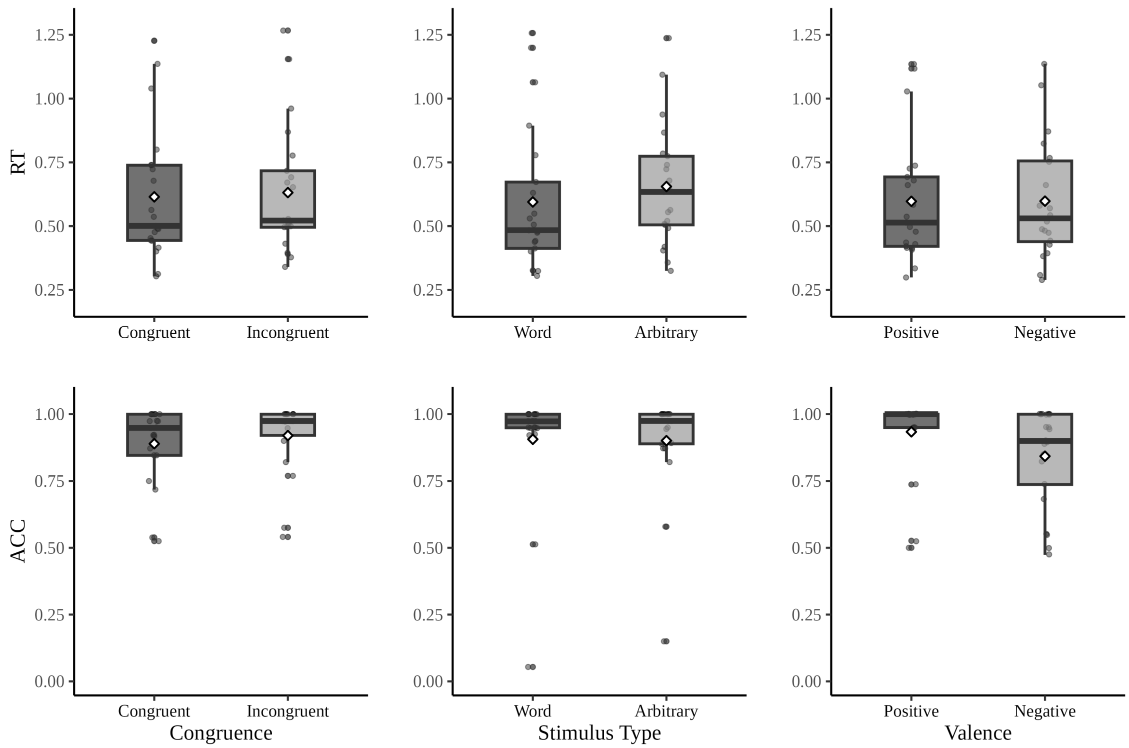

Figure 2. Reaction Times (RT) and Accuracies (ACC) in the Semantic Judgment Blocks with Arbitrary Stimuli and Words as a Function of Congruence, Stimulus Type, and Valence. *Note*: White diamonds represent the group means.

Semantic Judgment Tasks

Practice Trials. All participants were given a single 24-trials practice block before the semantic judgment tasks with arbitrary stimuli and words, respectively. Despite the absence of learning criteria to advance to the next phases of the experiment, it is worth mentioning that all participants showed less than perfect performances on the figure-figure combinations (a mean of about 90% correct).

Semantic Judgment Tasks with Arbitrary Stimuli and Words. Results from the Semantic Judgment Task are illustrated in Figure 2. The two-step analysis of response time (RT; leftmost graphs) with congruence (related vs. unrelated) and stimulus type (abstract forms vs. words) as factors showed an effect only of stimulus type (t[20] = -2.12, p = .046, Cohen's D = .46; see Tables 3, 4, 5, and 6 in the "Supplementary Materials" for a complete description of the effects). That is, only stimulus type predicted reaction time. Words produced faster responses than the arbitrary stimuli in the semantic judgment task (words: mean = 0.59 s, SD = 0.20 s; arbitrary stimuli: mean = 0.65 s, SD = 0.25 s).

The Figure 2 depicts the valence (positive vs. negative) and stimulus type effects on RT. The model factors these variables indicated significant effects for both (valence: t[20] = 2.1, p = .048, Cohen's D = .45; stimulus type: t[20] = -2.3, p = .031, Cohen's D = .5). Participants' responses were faster for all types of stimulus-stimulus relations of positive valence (positive valence: mean = 0.60 s, SD = 0.23 s; negative valence: mean = 0.63 s, SD = 0.23 s). These faster responses for the word-based trials with positive valence words indicate the prominence of word-priming effect. On the other hand, words showed faster responses than symbols, as would be expected from the previous model results.

Accuracy was similar in both semantic judgment blocks (see the central graph depicted in Figure 2). No significant effects for congruence or stimulus type were found (p > .053) (left and center graphs in Figure 2). However, correct judgments were more frequent for the positive valence stimuli than for the stimuli with negative valence (t[20] = -2.9, p < .01, Cohen's D = .45).

DISCUSSION

This study describes our efforts to overcome the challenges of social distancing imposed by the recent COVID-19 pandemic on data collection, specifically in stimulus equivalence and semantic judgment tasks. In the remote experiment herein described, in accordance with our expectations, participants' performances indicated the successful establishment of equivalence relations. Furthermore, participants showed a fast and virtually errorless learning on conditional discrimination training as well as positive outcomes on equivalence tests. These results are especially relevant given that the experiment was not conducted in well-controlled laboratory conditions.

This is possibly the first demonstration of faster learning rates among participants given stimulus class formation procedures via online platforms. Crump et al., (2013, Experiment 8), for example, have used the Amazon Mechanical Turk (MTurk) web-based platform to train 234 participants on a successive discrimination procedure with squares as stimuli. The squares varied in terms of border color (yellow vs. white), fill color (blue vs. purple), texture (smooth vs. rough) and stripes (striped vs. no-stripes). The participants had to press two keyboard buttons to classify the squares as "Category A" and "Category B" in accord with instructions provided before the task began. According to the authors, the participants had to be retrained many times to learn the perceptual categories. This overall tendency of slower learning rates was not observed in the current experiment. Obviously, the performances we observed in the MTS training and testing procedures may have occurred due to the particular arrangement of variables that enhance equivalence class formation. Remember that our procedures included: (1) cues that guarantee the selection of the correct comparison (e.g., Silveira et al., 2021); (2) smaller class-size and single-node equivalence relations; and (3) familiarity with the facial expressions used as sample stimuli (e.g., Arntzen, 2004; 2012; Arntzen & Mensah, 2020; Bortoloti & de Rose, 2011; Fields et al., 2012). But the positive outcomes in the MTS procedures allow the conclusion that our approach for online experimentation may be a likely alternative to the platforms such as the MTurk (Crump et al. (2013) and other web-based platforms (e.g., OpenSesame and Pavlovia) for its applicability to program suitable experimental conditions for the study of learning and cognition. In addition, it offers feasible alternatives for simulations of laboratory contingencies. For example, as the participant and the experimenter interact in real time, the experimenter can provide occasional instructions about possible intervening variables in the participant's context and partially monitor if the participant follows the instructions before the beginning of the experimental tasks.

We were also able to collect response-time data in the semantic judgment blocks, which suggests that even remote access tools may be suitable for reaction-time experiments (cf., Bridges et al., 2020). Accuracy and reaction-time data from the semantic judgment task with arbitrary stimuli reproduced the semantic priming effects for intra- and extra-experimental relations (e.g., Barnes-Holmes et al., 2005; Bortoloti & de Rose, 2011; Hayes & Bisset, 1998; Silveira et al., 2021) with a notable enhanced priming effect size for the semantic judgment task with words as stimuli.

The differences in the priming effect sizes can be accounted for in terms of the nature of the stimuli employed in the semantic judgment tasks. Sperber et al., (1979), for example, observed that the enhanced semantic priming effect is more likely to occur based on the semantic relatedness between words appearing in "prime-target" pairs (see also, Irwing & Lupker, 1983; Kutas & Donchin, 1980). In addition, Bortoloti and de Rose (2011) found faster responses in the semantic judgment tasks for those participants who reached highest levels of class-consistent responses in equivalence-probe trials. In this study, no mastery criteria were used for either the training or test phases. Future research could apply more stringent mastery criteria to both the training and testing phases in order to assess the influence of performance in the phases prior to the semantic judgment task on priming effect size. Manipulations in the nature of stimuli appearing as prime and target in the semantic judgment task may also be helpful to clarify this issue. These investigations are extremely relevant to determine whether these differences are due to the nature of the stimuli or any other parametric factors.

Furthermore, the semantic judgment data also showed a prevalence of word-priming effects on the response speed and accuracy over the priming for arbitrary stimuli. A possible explanation is that word pronunciation in natural speech contexts is richly connected with reinforcement provided by the verbal community (Skinner, 1957). In contrast, the abstract stimulus-stimulus relations result of differential reinforcement that was provided for a few conditional discriminations. Therefore, it is possible that the differences in accuracy and reaction times found in both blocks are related to these differences in the reinforcement histories for word usage and relational responding under these artificial contingencies (Galizio & Buskists, 1988; St. Peter Pipkin & Volmer, 2009; Silveira et al., 2021).

The remote data collection herein employed offered an immediate solution for behavioral scientists who did find themselves unable to conduct regular experimentation in their laboratories. It also has interesting benefits in comparison with in-person data collection. For example, participation in an experiment may become less time-consuming if done remotely since all participants can access the programmed contingencies from their office or home. Remote-based experiments will be especially suitable for laboratories with scarce resources, as the tools for online data collection offer an excellent cost-benefit trade-off. Moreover, remote data collection allows researchers to recruit larger and more diverse participant samples and run experiments with people from many different cities, states, or countries. Additionally, online experiments may facilitate data collection in longitudinal investigations, in which guaranteeing continuing participation is always a challenge.

Notwithstanding the possible contribution of the current research for web-experimentation in Behavior Analysis, there are many methodological issues that need clarification. For example: 1) To which extent are participants susceptible to distractors when the experiment is conducted outside the lab? 2) how can training and testing conditions be arranged to maximize stimulus class formation? 3) are equivalence classes formed by web-based MTS platforms analogous to classes formed in the experimental analysis of human behavior laboratories? 4) how might Internet connection speeds affect the participants' behavior as a whole (e.g., latency data, adventitious delayed reinforcement, unexpected multiple responses to the sample-stimulus and observing behavior)? These are just some of the issues that justify systematic follow-up work into this topic.

The population employed in this research is possibly more diverse than that of the majority of studies that have recruited undergraduate students [see Arntzen (2012) and Regaço et al., (2023), for examples]. However, the current research lacks information regarding the participants' demographic characteristics. Such surveys could be implemented for future research as ways for determining whether performances shaped by the equivalence procedures are influenced by demographic variables.

In summary, we were able to successfully collect data from all participants in the MTS and semantic judgment procedures, suggesting that remote access apps can be used to approximate the optimal laboratory conditions for human behavior research. This approach seems suitable for cases in which data collection requires constant monitoring by the experimenter such as rule-governed behavioral tasks, verbal behavior learning, or equivalence-based instruction protocols (e.g., de Almeida et al., 2020; Augland et al., 2020; Cortez et al., 2020; Hill et al., 2020). This successful demonstration may encourage other laboratories to adopt online experiments, especially during times in which social distancing measures are in effect, but also as an interesting addition or alternative to the usual in-person data collection.

REFERENCES

Augland, H., Lian, T., & Arntzen, E. (2020). Comparing a student active learning format to equivalence-based instruction. *European Journal of Behavior Analysis, 21*(2), 328-347. https://doi.org/10.1080/15021149.2020.1752513

Augustson, E. M., & Dougher, M. J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behavior Therapy and Experimental Psychiatry, 28*(3), 181–191. https://doi.org/10.1016/S0005-7916(97)00008-6

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. & 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. (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

Barnes-Holmes, D., Staunton, C., Whelan, R., Barnes-Holmes, Y., Commins, S., Walsh, D., Stewart, I., Smeets, P., & Dymond, S. (2005). Derived stimulus relations, semantic priming, and event-related potentials: testing a behavioral theory of semantic networks. *Journal of the Experimental Analysis of Behavior*, *84*(3), 417-430. https://doi.org/10.1901/jeab.2005.78-04

Ben-Shachar M, Lüdecke D, Makowski D (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Bortoloti, R., & de Rose, J. C. (2011). Avaliação do efeito de dica semântica e da indução de significado entre estímulos abstratos equivalentes [Evaluation of semantic priming effect and induction of meaning among abstract equivalent stimuli]. *Psicologia: Reflexão e Crítica*, *24*(2), 381-393. https://doi.org/10.1590/S0102-79722011000200020

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

de Rose, J. C., de Souza, D. G., Rossito, A. L., & de Rose, T. M. S. (1992). Stimulus equivalence and generalization in reading after matching to sample by exclusion. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding Verbal Relations* (pp. 69–82). Context Press.

Cortez, M. D., Ferreira, M. S., Lima, P. M. P., Silva, M. G. A., & Reis, M. J. D. (2020). Efeitos de reforçadores de magnitude aumentada e do controle aversivo na manutenção do seguimento de regras em estudantes universitários [Effects of the increased reinforcer magnitude and aversive control on the maintenance of rule-following in undergraduate students]. *Revista Brasileira de Terapia Comportamental e Cognitiva*, *22*(1). https://doi.org/10.31505/rbtcc.v22i1.1409

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

de Almeida, J. H., Cortez, M. D., & de Rose, J. C. (2020) The effects of monitoring on children's rule-following in a computerized procedure. *The Analysis of Verbal Behavior*, *37*, 1-1. https://doi.org/10.1007/s40616-020-00130-5

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*(2), 163–181. https://doi.org/10.1901/jeab.2012.97-163

Galizio, M., & Buskist, W. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: Selecting reinforcers and arranging contingencies. *The Behavior Analyst*, *11*(1), 65–69. https://doi.org/10.1007/BF03392457

Gelman, A. (2017). Two-stage regression and multilevel modeling: A commentary. *Political Analysis*, *13*(4), 459-461. https://doi.org/10.1093/pan/mpi032

Gentili, C. & Cristea I. A. (2020). Challenges and opportunities for human behavior research in the coronavirus disease (COVID-19) pandemic. *Frontiers in Psychology*, *11*, 1786. https://doi.org/10.3389/fpsyg.2020.01786

Hayes, S. C. (1989). *Rule-Governed Behavior: Cognition, Contingencies, and Instructional Control*. New York, NY: Plenum.

Hayes, S. C., & Bisset, R. T. (1998). Derived stimulus relations produce mediated and episodic priming. *The Psychological Record*, *48*, 617-630. https://doi.org/10.1007/BF03395293

Hill, K. E., Griffith, K. R., & Miguel, C. F. (2020). Using equivalence-based instruction to teach piano skills to children. *Journal of Applied Behavior Analysis*, *53*(1), 188–208. https://doi.org/10.1002/jaba.547

Irwin, D. I., & Lupker, S. J. (1983). Semantic priming of pictures and words: A levels of processing approach. *Journal of Verbal Learning & Verbal Behavior*, *22*(1), 45–60. https://doi.org/10.1016/S0022-5371(83)80005-X

Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., and Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period. *Science*, *368*(6493), 860–868. https://doi.org/10.1126/science.abb5793

Kutas, M., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potentials. *Brain Research*, *202*(1), 95–115. https://doi.org/10.1016/0006-8993(80)90646-0

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Mizael, T. M., de Almeida, J. H., Roche, B., & de Rose, J. C. (2021). Effectiveness of different training and testing parameters on the formation and maintenance of equivalence classes: Investigating prejudiced racial attitudes. *The Psychological Record*, *71*(1), 265-277. https://doi.org/10.1007/s40732-020-00435-w

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic Processing In Reading: Visual Word Recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*(3), 352–369. https://doi.org/10.3758/BF03200862

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pilgrim, C. (2016). Considering definitions of stimulus equivalence. *European Journal of Behavior Analysis*, *17*, 105-114. https://doi.org/10.1080/15021149.2016.1156312

Regaço, A., Zapparoli, H.R., Aggio, N.M., Silveira, M. V. & Arntzen, E. (2023). Maintenance of stimulus equivalence classes: A bibliographic review. *Psychological Record*, *73*, 1–11. https://doi.org/10.1007/s40732-023-00535-3

Silveira, M. V., Camargo, J. C., Aggio, N, M., Ribeiro, G. W., Cortez, M. D., Young, M. E., & de Rose (2021). The influence of training procedure and stimulus valence on the long-term maintenance of equivalence relations. *Behavioural Processes*, *185*, 104343. https://doi.org/10.1016/j.beproc.2021.104343

Shepard, R.N., Hovland, C.I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*(13), 1–41.

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

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

Sidman, M. (1994). *Equivalence Relations and Behavior: A Research Story*. Boston: Authors Cooperative.

Sidman M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*(1), 127–146. https://doi.org/10.1901/jeab.2000.74-127

Sidman, M. (2018). What is interesting about equivalence relations and behavior? *Perspectives on Behavior Science*, *41*(1), 33–43. https://doi.org/10.1007/s40614-018-0147-8

Sperber, R.D., McCauley, C., Ragain, R.D., Weil, C. M. (1979). Semantic priming effects on picture and word processing. *Memory & Cognition* *7*(5), 339–345. https://doi.org/10.3758/BF03196937

St. Peter Pipkin, C., & Vollmer, T. R. (2009). Applied implications of reinforcement history effects. *Journal of Applied Behavior Analysis*, *42*(1), 83–103. https://doi.org/10.1901/jaba.2009.42-83

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record*, *41*(1), 33–50. https://doi.org/10.1007/BF03395092

World Health Organization (2021). General guidance on Coronavirus disease (Covid19). Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019