ASSESSING THE LONG-TERM MAINTENANCE OF TWO, 3-MEMBER EQUIVALENCE CLASSES

Marcelo V. Silveira[1,2,6], Alceu Regaço[4], Natália M. Aggio[2], Mariéle Diniz Cortez[4], Mateus Silvestrin[1], Leonardo B. Marques[5], Diego D. M. C. Matos[5], Gabriela L. Teixeira[6], Ohana T. Rabello[5], Kamilla Tenório[5], Lucas T. V. M. Santana[5], Giovan W. Ribeiro[1], Julio C. Camargo[4], William F. Perez[6], Erik Arntzen[6], and Marcelo S. Caetano[1]

[1]UNIVERSIDADE FEDERAL DO ABC; [2]UNIVERSIDADE DO OESTE PAULISTA; [3]UNIVERSIDADE DE BRASÍLIA;

[4]UNIVERSIDADE FEDERAL DE SÃO CARLOS; [5]UNIVERSIDADE FEDERAL DE ALAGOAS; [6]INSTITUTO PAR;

[7]OSLO METROPOLITAN UNIVERSITY

Research on the maintenance of equivalence classes has shown that class-consistent responding often remains stable despite the absence of training. Several factors influence these outcomes, including stimulus characteristics, training procedures, and class size. However, only two studies to date have examined the maintenance of equivalence classes composed of stimuli with different valence (e.g., happiness and anger). The present study replicates this previous research by teaching two three-member equivalence classes and assessing the maintenance of emergent relations 30 days later. Results showed that 8 out of 23 participants maintained the equivalence classes. However, stimulus valence did not significantly affect maintenance performance, contrary to previous findings.

Keywords: maintenance of equivalence relations, stimulus equivalence, typical humans

According to the stimulus equivalence account, after conditional discrimination training involving different stimuli (e.g., AB and BC relations), the emergence of an equivalence class is assessed by testing for the properties of reflexivity (e.g., AA, BB, and CC relations), symmetry (e.g., BA and CB relations), transitivity (e.g., AC relation), and equivalence (e.g., CA; Fields & Verhave, 1987; Sidman & Tailby, 1982). The maintenance of equivalence classes can be assessed by conducting tests for emergent relations after a predetermined interval following class formation.

Several studies have shown that these equivalence relations often remain stable over time, even in the absence of intervening training (Aggio & Domeniconi, 2005; Camargo & Haydu, 2015; Eilifsen & Arntzen, 2017; Haydu & de Paula, 2008; Saunders et al., 1988; Spradlin et al., 1992; Silveira et al., 2016, 2021; Wirth & Chase, 2002).For example, in the study by Eilifsen and Arntzen (2015), 24 adults were exposed to a series of match-to-sample (MTS) trials that established AB, AC, AD, and AE conditional discriminations. Half of the participants were exposed to MTS-probe trials and demonstrated symmetry and transitivity for these conditional discriminations. All participants were then re-exposed to MTS training and probe trials after two and four weeks. Results showed that, even several days after training, all participants were able to perform the trained conditional discrimination and respond in accordance with stimulus equivalence. Interestingly, the robust maintenance reported by these researchers was independent of the participants' prior experience with equivalence testing.

According to the Network Model (Spradlin et al., 1992), the long-term retention of equivalence relations is a direct function of class size. The model proposes that equivalence classes comprise both trained and untrained relations, and that this complex network structure ensures the integrity of the class and enhances the likelihood of stability over time. Spradlin et al.'s (1992) Network Model allows the prediction that long-term maintenance effects are more likely to be observed in larger than in smaller equivalence classes. In this regard, the stability of equivalence classes would be a function of class size. Indeed, a recent systematic review of variables influencing performance on maintenance tests confirms that class size is one of the most consistent variables positively impacting the long-term maintenance of equivalence relations (Regaço et al., 2023).

Another variable that has been studied is the inclusion of stimuli with positive or negative valence in the equivalence classes. Silveira et al. (2016) examined the maintenance of three four-member equivalence classes

composed of abstract forms and pictures of human faces expressing happiness, anger, or neutrality. After confirming class formation, participants completed a maintenance test 30 days later. Results showed that 69% of participants responded correctly to relations in the "happy" class, 50% in the "neutral" class, and 38% in the "angry" class. These findings suggest that the equivalence class containing happy faces was more stable over time than those containing neutral or angry faces.

Author Note: Address correspondence to: Marcelo Silveira (email:marcelopsi06@gmail.com) , https://orcid.org/0000-00203-3091-0788

This experiment was a collaboration of researchers from Brazilian and Norwegian universities sharing common interests into the variables from which complex learning and human cognition are functions. MVS had a post-doctoral scholarship from the São Paulo Research Foundation (FAPESP, Grant #2017/06178-7) and a special funding from the same institution (Bolsa de Estagio em Pesquisa no Exterior – BEPE, Grant #2019/14661-5). This research is also part of the scientific program of Instituto Nacional de Ciência e Tecnologia sobre Comportamento Cognição e Ensino (INCT- ECCE), supported by São Paulo Research Foundation (Grant #2014/50909-8), by the Brazilian Council for Scientific and Technological Development (CNPq: grant #465686/2014 1), and by Coordination for the Improvement of Higher Education Personnel (CAPES: grant #88887.136407/2017-00). We thank Deisy de Souza, chairperson of the INCT-ECCE, for the support.

Other Author Information

Alceu Regaço https://orcid.org/0002-5209-0082

Natália M. Aggio https://orcid.org/0000-0002-9719-6067

Mariéle Diniz Cortez https://orcid.org/0000-0003-3301-8330

Mateus Silvestrin https://orcid.org/0000-0002-3482-3676

Leonardo B. Marques https://orcid.org/0000-0003-4993-4587

Diego D. M. C. Matos https://orcid.org/0000-0002-8415-6955

Gabriela L. Teixeira https://orcid.org/0009-0004-0599-0477

Ohana T. Rabello https://orcid.org/0000-0003-0634-1116

Kamilla Tenório https://orcid.org/0000-0002-8088-7507

Lucas T. V. M. Santana https://orcid.org/0009-0004-1975-0503

Giovan W. Ribeiro https://orcid.org/0000-0003-0631-9381

Julio C. Camargo https://orcid.org/0000-0002-2927-9928

William F. Perez https://orcid.org/0000-0001-8365-1439

Erik Arntzen https://orcid.org/0000-0002-8471-1058

Marcelo S. Caetano https://orcid.org/0000-0002-5353-7982

Subsequent studies replicated Silveira et al. (2016), while also examining the impact of using Simultaneous Matching-to-Sample (SMTS) versus Delayed Matching-to-Sample (DMTS; see Procedure section for details) on the maintenance of equivalence classes (Silveira et al., 2021). The results reproduced the initial findings, showing higher maintenance of the "happy" class compared to the others, particularly when the DMTS procedure was used. Thus, it appears that both positive valence and class size can enhance performance on equivalence class maintenance tests. However, these two variables have not been systematically manipulated together.

The present study aims to test the combined effect of positive stimulus valence and small class size by replicating the procedures of Silveira et al. (2016; 2021), but establishing two three-member classes instead of three four-member classes (i.e., two smaller classes instead of three larger ones).

This experiment is also an extension of Silveira et al., (2025) which have conducted the equivalence procedures via a web-based platform. The experimental task was accessed remotely by participants, who were given AB and AC conditional discrimination training in a DMTS procedure. This procedure aimed to teach conditional discriminations involving (A) images of human faces portraying either happiness or anger (A1 and A2), and (B, C) distinct abstract shapes (B1, B2, C1, and C2). Following training, the participants completed MTS-probe trials that evaluated the emergence of untrained BC and CB equivalence relations. The results aligned with predictions based on equivalence class formation within laboratory settings. The present study aims to extend these findings by examining the maintenance of experimentally established equivalence classes via web-based platforms. Ultimately, the current work aims to contribute to the current knowledge about variables affecting the long-term maintenance of equivalence relations (cf., Regaço et al., 2023)

METHOD

Participants

Twenty-five volunteers (13 males and 12 females; Mage = 26.5 years, range = 18–35) were recruited via social media advertisements. Data collection occurred between May and November 2021.

Instructions, Application, and Stimuli

Participants were instructed to find a comfortable, well-lit, and quiet room with a chair and a table for the experiment. They were required to remain seated and use a computer mouse and keyboard during the experimental sessions (see instructions in Appendix A). The experiment was conducted remotely using AnyDesk software, which allowed participants to access the experimenter's computer. PsychoPy3 (v2020.1.1; Peirce et al., 2019) controlled instructions, stimulus presentation (visual and auditory), and data collection. The stimuli consisted of images of human faces expressing happiness or anger (A1 and A2, respectively), and black-and-white abstract stimuli (B1, B2, C1, and C2; see Figure 1).

Procedure

MTS-Training Blocks

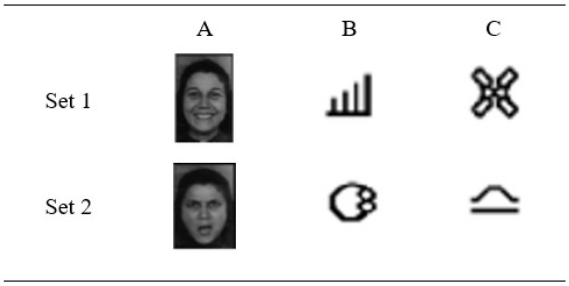
AB and AC conditional discriminations were established across three 16-trial blocks without a performance-based criterion[1] (i.e., each block ended after 16 trials, regardless of participants' accuracy). The first and second blocks involved AB and AC conditional discriminations, respectively, and the third block was a mixed training block consisting of intermixed AB and AC trials (eight trials each). The first four trials of the AB and AC training blocks used a Simultaneous Matching-to-Sample (SMTS) procedure, in which the sample stimulus remained on-screen after being clicked, and a cueing procedure was used (the incorrect comparison stimulus was faded out). The remaining 12 trials of each block used a Delayed Matching-to-Sample (DMTS) procedure without cues: clicking the sample stimulus removed it from the screen, followed by the immediate presentation of the comparison stimuli (0-s DMTS). The mixed training block used only DMTS without cues.

All trials began with the presentation of a sample stimulus (A1 or A2) at the top-center of the screen. After clicking the sample, two comparison stimuli appeared in the bottom corners of the screen (position randomized across trials). Correct selections resulted in stimulus removal and simultaneous presentation of a bell sound and the word "Correct" (in green letters) for 1.5 s. Incorrect selections resulted in stimulus removal and simultaneous presentation of a buzz sound and the word "Wrong" (in red letters) for 1.5 s. A 0.55-s intertrial interval (ITI) followed each trial.

The participants were given AB, AC and mixed training blocks only once and they progressed through these stages regardless of their performance in each block.

Figure 1

Stimuli Used in the Training and Testing Phases



Equivalence Test

After completing the training, participants were informed that their behavior would no longer produce feedback for correct or incorrect responses, but that they should continue responding according to what they had learned. Afterwards, two 12-trial blocks assessed the emergence of BC (B1C1 and B2C2) and CB (C1B1 and C2B2) relations. An SMTS procedure was used, and no consequences were presented for correct or incorrect selections. To confirm the formation of two three-member equivalence classes (A1B1C1 and A2B2C2), participants had to achieve at

[1] The use of training blocks without a performance-based criterion was due to limitations in the software and the constraints of online data collection. Although performance criteria are an important variable (Richling et al., 2023), most participants met the criterion during the equivalence test.

least 91.7% accuracy in both BC and CB test blocks.

Maintenance Test

The maintenance test was conducted 30 days after the equivalence test and was identical, except for the instructions presented before the test (see Appendix A). All participants were asked to return for the maintenance test, regardless of their performance in the prior equivalence test.

Statistical Analysis

Statistical analyses were conducted using R (Version 3.5.1; R Core Team, 2018) within the RStudio environment (RStudio Team, 2015), using the lme4, car, emmeans, and ggplot2 packages. A mixed-effects logistic regression model estimated the probability of class-consistent responding during BC and CB tests, both immediately after training (equivalence test) and 30 days later (maintenance test). The model included the following fixed effects: (1) trial outcome (class-consistent vs. class-inconsistent responding), (2) test type (equivalence test vs. maintenance test), (3) tested relation (BC vs. CB), (4) emotional valence of the equivalence classes (Happy vs. Angry), and (5) all interactions among these variables. Pairwise post-hoc comparisons were conducted using Tukey's HSD to assess differences between levels of the independent variables.

RESULTS

Of the 25 initial participants, 88% (n = 22) met the criterion for equivalence class formation. Among those who completed the maintenance test (n = 23), 35% (n = 8) maintained the equivalence classes (see Table 1). During the BC and CB equivalence tests, the probability of class-consistent responding was close to 1.00 for both classes, with no statistically significant difference between them . In the BC and CB maintenance test, the probability of class-consistent responding decreased significantly for both classes compared to performance in the equivalence test (p = .049). Individual data suggests that performance deterioration was more pronounced for the B2C2 and C2B2 than

Table 1

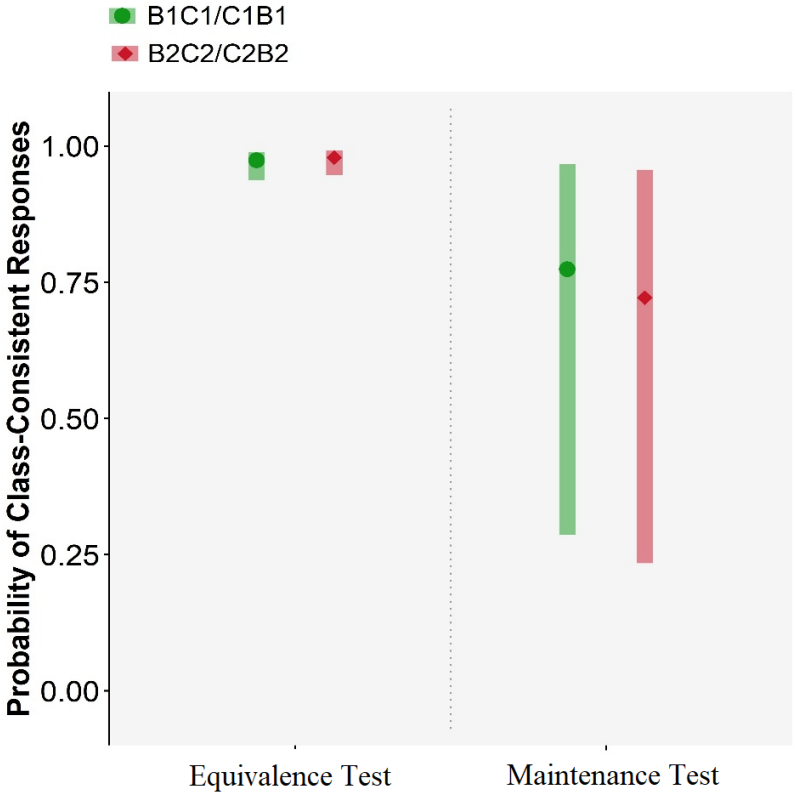
Percentage of Class-Consistent Responses Achieved by All Participants across Training, Equivalence Class Formation Tests, and Follow-Up Tests

Participants	1st session					2nd session	
	AB	AC	Review	BC test	CB test	BC test	CB test
P1	100	100	100	**91.7**	**100**	0	0
P2	100	100	100	**100**	**100**	0	33.3
P3	62.5	100	93.8	**100**	**91.7**	**100**	**100**
P4	100	100	100	**100**	**100**	8.3	**91.7**
P5	100	93.8	100	**100**	**91.7**	**100**	**100**
P6	93.8	100	100	83.3	**91.7**	75	**100**
P7	81.3	93.8	100	**100**	**100**	-	-
P8	93.8	100	100	**91.7**	**100**	**100**	**100**
P9	93.8	100	100	**100**	**100**	**91.7**	0
P10	100	93.8	100	**100**	**100**	75	**100**
P11	100	100	100	**91.7**	**100**	0	0
P12	100	100	93.8	**91.7**	**100**	**100**	**91.7**
P13	100	100	100	**91.7**	**100**	8.3	0
P14	62.5	100	81.2	83.3	**100**	**91.7**	**100**
P15	100	100	100	66.7	**91.7**	83.3	**100**
P16	93.8	81.3	100	**100**	**91.7**	-	-
P17	100	100	100	**100**	**100**	0	0
P18	100	100	100	**100**	**100**	0	0
P19	100	93.8	100	**100**	**100**	**100**	**100**
P20	100	100	100	**100**	**91.7**	8.3	0
P21	100	100	100	**100**	**100**	66.7	50
P22	87.5	93.8	93.8	**91.7**	**100**	0	0
P23	100	100	100	**100**	**100**	**100**	**100**
P24	100	100	100	**100**	**100**	**100**	**100**
P25	100	100	100	**100**	**100**	83.3	**100**

Note. Scores in bold represent the criterion achievement. Participants with a dash in the 2nd session represent that the participants did not do the maintenance test.

Figure 2

Probability of Class-Consistent Responses on Equivalence and Maintenance Tests



Note. Points represent mean and shaded areas represent 95% confidence interval.

for the B1C1 and C1B1 relations. Figure 2 shows the responses predicted from the fitted model's fixed (populational) effects for each relation and test type. The plot shows predicted average responses and variability in the response scale, allowing the direct inspection of the predominant behavioral pattern extracted from the data by the statistical model.

DISCUSSION

This experiment aimed to replicate Silveira et al. (2016, 2021) by assessing the maintenance of two three-member equivalence classes composed of pictures of faces expressing happiness or anger. Both these studies have confirmed that smaller "happy classes" are more likely to be maintained than the "angry class" and "neutral class". However, the results did not replicate Silveira et al.'s (2016, 2021) findings, as no significant difference was observed in the number of participants who maintained the two classes. The only observed difference between classes was a greater decline in class-consistent responding for the class composed of angry faces compared to the class composed of happy faces, although this difference was not statistically significant.

Even though the current study have confirmed the viability of using a web-based platform for conducting stimulus equivalence experiments with typical adults (see Silveira et al., 2025), some limitations should be considered: (1) the limited number of training trials; (2) the absence of a mastery criterion for the training blocks; (3) the absence of follow-up measurements of the baseline relations; and 4) the use of only two classes, which may have attenuated the "happy superiority effect" (Bortoloti et al., 2019, 2024) by allowing participants to respond both by selection and by rejection (Plazas, 2024; Sidman, 1987). That is, participants who maintained B1C1 and C1B1 relations could have responded correctly to B2C2 and C2B2 relations by rejecting the

incorrect comparison stimulus. In this sense, when equivalence classes are based on selection control, the transfer of function is consistent with the experimentally defined classes (A1B1C1 - “happy class”; A2B2C2 - “angry class”). However, when classes are formed through rejection control, the derived transfer effect may be opposite to that specified by the experimenter (A1B2C1 - “angry class”; A2B1C2 - “happy class”) (cf., Perez et al., 2015). This hypothesis is reinforced by the higher maintenance of the "angry" class relations observed in the current study compared to maintenance of the same classes in Silveira et al. (2021). Future studies should control for this effect by introducing three comparison stimuli per trial or by teaching the relations of three stimulus classes. In addition, further research is needed to clarify the relationship between class size and stimulus valence on maintenance performance.

REFERENCES

Aggio, N. M., & Domeniconi, C. (2012). Formation and maintenance of equivalence classes: A study with elderly participants. *Acta Comportamentalia*, *20*(1), 29–43.

Bortoloti, R., de Almeida, R. V., de Almeida, J. H., & de Rose, J. C. (2019). Emotional faces in symbolic relations: A happiness superiority effect involving the equivalence paradigm. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00954

Bortoloti, R., Perez, W. F., de Almeida, J. H., & de Rose, J. C. (2024). Research on human language and cognition in Brazil based on Stimulus Equivalence and Relational Frame Theory. *The Psychological Record*. https://doi.org/10.1007/s40732-023-00581-x

Camargo, J., & Haydu, V. B. (2015). Class size effects on the establishing, maintenance, and recovery of equivalence relations. *Behavior Analysis: Research and Practice*, *15*(1), 36–49. https://doi.org/10.1037/h0101068

Eilifsen, C., Arntzen, E. (2015). Effects of training structure and the passage of time on trained and derived performance. *The Psychological Record*, *65*, 1–12. https://doi.org/10.1007/s40732-014-0067-2

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior*, *48*(2), 317–332. https://doi.org/10.1901/jeab.1987.48-317

Haydu, V. B., & de Paula, J. B. C. (2008). Effects of class size on the maintenance of equivalent relations. *Psicologia: Reflexão e Crítica*, *21*, 233–251. https://doi.org/10.1590/S0102-79722008000200009

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

Perez, W. F., Tomanari, G. Y., & Vaidya, M. (2015). Effects of select and reject control on equivalence class formation and transfer of function. *Journal of the experimental analysis of behavior*, *104*(2), 146–166. https://doi.org/10.1002/jeab.164

Plazas, E. A. (2024). Exploratory study of equivalence under conjoint select and reject control. *The Psychological Record*, *74*(1), 17–33. https://doi.org/10.1007/s40732-023-00579-5

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

Richling, S.M., Fienup, D.M., Wong, K. (2023). Establishing performance criteria for skill mastery. In J. L. Matson, *Handbook of Applied Behavior Analysis. Autism and child psychopathology series* (pp. 393–410). Springer, Cham. https://doi.org/10.1007/978-3-031-19964-6_22

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *49*(1), 95–115. https://doi.org/10.1901/jeab.1988.49-95

Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes, & L. J. Hayes (Org.). *Understanding verbal relations: The second and third International Institute on Verbal Relations* (pp. 29–42). Nevada: Context.

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis*, 22(I), 11–18.

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

Silveira, M. V., Aggio, N. M., Cortez, M. D., Bortoloti, R., Rico, V. V., & de Rose, J. C. (2016). Maintenance of equivalence classes and transfer of functions: The role of the nature of stimuli. *The Psychological Record*, *66*(1), 65–74. https://doi.org/10.1007/s40732-015-0152-1

Silveira, M. V., Camargo, J. C., Aggio, N. M., Ribeiro, G. W., Cortez, M. D., Young, M. E., & de Rose, J. C. (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

Silveira, M. V., Zaparolli, H. R., Sbrocco, G., Silvestrin, M., Rabelo, O. T., Tenório, K., Marques, L. B., Matos, D. D. M. C., Arntzen, E., & Caetano, M. S. (2025). Exploring an alternative for running experiments during social distancing: A pilot study of stimulus equivalence. *Experimental Analysis of Human Behavior Bulletin*. (advance online publication). https://doi.org/10.17605/OSF.IO/SXWA5

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

Wirth, O., & Chase, P. N. (2002). Stability of functional equivalence and stimulus equivalence: Effects of baseline reversals. *Journal of the Experimental Analysis of Behavior*, *77*(1), 29–47. https://doi.org/10.1901/jeab.2002.77-29

Appendix A

First day instruction

"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 must use your computer or laptop in the quietest place at your home. You must also ask your relatives to be left alone for approximately 15 minutes. The experiment consists in two stages. In the first stage you will learn how to select some figures. The second stage of the experiment will occur 30 days later, in which you will be re-exposed to a very similar task. 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."

Maintenance instruction

"You will be exposed to the same tests you did previously. You will not be given feedback for any responses. But you should try to respond in accordance with what you have learned during the previous session. Press the spacebar when you think you are ready"