Paidéia
2018, Vol. 28, e2826. doi: http://dx.doi.org/10.1590/1982-4327e2826
ISSN 1982-4327 (online version)

Failure to Produce False Memories Through the Stimulus Equivalence Paradigm[1]

Natalia Maria Aggio[2], Julio César de Rose[2]

[2]*Universidade Federal de São Carlos, São Carlos-SP, Brazil*

Abstract: Stimulus equivalence has been adopted as a behavioral explanation for false memories. The present study aimed to test false memories using lists compound of equivalent stimuli. 10 undergraduate students learned three 4-member (Classes 1, 2, 3) and three 12-member equivalence classes (Classes 4, 5, 6). A week later these participants performed a recognition test. Participants first saw a study list comprising 10 of the 12 stimuli from Classes 4, 5 and 6. Later, they saw a list comprising all stimuli from study list (targets), the remaining stimuli from the Classes 4, 5 and 6 (critical lures) and nine stimuli from Classes 1, 2 and 3 (non-related lures). Due to the equivalence relation between targets and critical lure, it was expected that the second would be recognize as much as the first, but results indicated critical and non-related lures where equally recognized and at low levels.

Keywords: stimulus equivalence, false memory, matching to sample

Fracasso em Produzir Falsas Memórias por Meio do Paradigma da Equivalência de Estímulos

Resumo: Equivalência de estímulos tem sido adotada como uma explicação comportamental das falsas memórias. Este estudo teve por objetivo testar falsas memórias usando listas compostas por estímulos de classes de equivalência. No presente estudo, 10 estudantes universitários aprenderam três classes de equivalência com quatro estímulos (Classes 1, 2, 3) e três com 12 estímulos (Classes 4, 5 e 6). Uma semana depois esses participantes realizaram um teste de reconhecimento. Primeiro viram uma lista de estudo composta de 10 dos 12 estímulos das Classes 4, 5 e 6. Depois viram uma lista com todos os estímulos da lista de estudos (alvos), os demais estímulos das Classes 4, 5 e 6 (distratores críticos), e nove estímulos das Classes 1, 2 e 3 (distratores não-relacionados). Devido à relação de equivalência entre os alvos e os distratores críticos, era esperado que os participantes reconhecessem os segundos na mesma proporção dos primeiros, porém os resultados indicaram que os críticos e não relacionados foram reconhecidos na mesma proporção e em níveis baixos.

Palavras-chave: equivalência de estímulos, memória falsa, matching to sample

Fracaso en Producir Falsas Memorias por el Paradigma de la Equivalencia de Estímulos

Resumen: Equivalencia de estímulos ha sido adoptada como explicación conductual de las falsas memorias. El presente estudio tuvo por objetivo probar falsas memorias usando listas compuestas por estímulos de clases de equivalencia. 10 estudiantes universitarios aprendieron tres clases de equivalencia con cuatro estímulos (Clases 1, 2, 3) y tres con 12 estímulos (Clases 4, 5, 6). Una semana después, realizaron un test de reconocimiento. Primero, vieron una lista de estudio con 10 de los 12 estímulos de las Clases 4, 5 y 6. Después, vieron una lista con todos los estímulos de la lista de estudios (*targets*), los demás estímulos de las Clases 4, 5 y 6 (distractores críticos), y nueve estímulos de las Clases 1, 2 y 3 (distractores no-relacionados). Debido a la relación de equivalencia entre *targets* y distractores críticos, era esperado que fuesen reconocidos los segundos en la misma proporción de los primeros, pero los resultados indicaron que fueron reconocidos en cantidades semejantes y en niveles bajos los críticos y no-relacionados.

Palabras clave: equivalencias de estímulos, falsas memoria, matching to sample

Behavioral methods such as matching-to-sample training or stimulus pairing may establish equivalence relations between physically dissimilar stimuli (Sidman, 1994). Stimuli related by equivalence comprise equivalence classes, also called artificial stimulus categories (Bennett, Meulders, Baeyens, & Vlaeyen, 2015), so that equivalent members may substitute for each other. This has been called *transfer of functions*: functions trained directly to one

[1]Paper derived from the doctoral dissertation of the first author under supervision of the last author, defended in 2014, in the Graduate Program in Psychology at Universidade Federal de São Carlos. Support: National Council for Scientific and Technological Development (Portuguese acronym: CNPQ, Grant No. 465686/2014-1, 573972/2008-7) e São Paulo Research Foundation (FAPESP, Grant No. 08/57705-8, 10/08399-1 e 2014/50909-8).

Correspondence address: Natalia Maria Aggio. Universidade Federal de São Carlos. Departamento de Psicologia. Rodovia Washington Luís, Km 235, São Carlos-SP, Brazil. CEP 13.565-905. E-mail: nanaggio@hotmail.com

stimulus are also exhibited by equivalent stimuli (de Rose, McIlvane, Dube, & Stoddard, 1988; Dougher, Augustson, Markham, Greenway, & Wulfert, 1994; Fields, Arntzen, Nartey, & Eilifsen, 2012). Actually, different methods converge to indicate that equivalent stimuli are related in meaning: they are rated similarly in Semantic Differential instruments (Bortoloti & de Rose, 2011a; Bortoloti, Rodrigues, Cortez, Pimentel, & de Rose, 2013; de Almeida & de Rose, 2015; de Almeida, Bortoloti, Ferreira, Schelini, & de Rose, 2014; Perez, de Almeida, & de Rose, 2015), they prime each other in a semantic priming paradigm (Barnes-Holmes et al., 2005; Bortoloti & de Rose, 2011b), and they produce N400 effect in electrophysiological measures (Amd, Barnes-Holmes, & Ivanoff, 2013; Barnes-Holmes et al., 2005; Bortoloti, Pimentel, & de Rose, 2014; Haimson, Wilkinson, Rosenquist, Ouimet, & McIlvane, 2009; Tabullo, Yorio, Zanutto, & Wainselboim, 2015).

These properties of stimulus substitutability, relatedness of meaning, and transfer of functions have recently led researchers to adopt stimulus equivalence as a behavioral explanation for the phenomenon of false memories (Challies, Hunt, Garry, & Harper, 2011; Guinther & Dougher, 2010, 2014). False memories are defined as memories of events that never occurred or distorted memories of events that occurred (Aggio, Pedrosa, & de Rose, 2017; Roediger & McDermott, 1995).

A paradigm often used to study false memories is the DRM (Deese-Roediger-McDermott) paradigm (Deese, 1959; Roediger & McDermott, 1995). In this paradigm, participants learn a list of target words that are all semantically related to each other and to a critical word not included in the list (such as "dolls," "female," and "dress," all related to the word "girl," which does not occur in the list). Participants often report that the critical lure was present in the list. These false memories have been attributed to the semantic relatedness between the critical word and all other words in the list (Gallo, 2010).

Guinther and Dougher (2010) reproduced this effect with words that were not semantically related in the language, such as animal, brick, cigar, etc. Matching-to-sample training established each of three sets comprised of 24 of such words as a class of equivalent stimuli. Participants then viewed a list of 12 target words, all members of the same equivalence class. Those that had demonstrated equivalence classes on tests, showed significantly more false recognitions of equivalent words that were not present in the list, compared to participants that had not demonstrated equivalence classes. Based on this research, it may be hypothesized that false memories in the DRM procedure can be conceived as intrusions of equivalent stimuli in reports about stimuli that the person has actually seen. In a schematic way, a person has seen stimulus A and subsequently reports having seen stimulus B, which is different from A but is equivalent to A (Challies et al., 2011; Guinther & Dougher, 2010, 2014).

The DRM effect is attributed to the semantic relatedness of words (Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001), Guinther and Dougher (2010) were able to construct the semantic relatedness by establishing equivalence classes with previously unrelated words. The equivalence literature shows, however, that formation of equivalence classes may establish a "semantic" relation between non-verbal stimuli. For instance, Bortoloti et al. (2014) and Haimson et al. (2009) established equivalence classes comprising non-representative forms. When pairs of stimuli were later presented successively, event-related cortical potentials showed a pronounced negative peak about 400 ms after the second stimulus when this stimulus was not equivalent to the first one, reproducing the N400 effect that has been established in cognitive neuroscience as an index of semanticity (Kutas & Hillyard, 1980).

If the formation of equivalence classes establishes semantic relations between stimuli, the DRM effect may be obtained even with non-verbal stimuli. Therefore, the present study aimed to test false memories using list compound of equivalent stimuli to verify if the DRM effect would be replicated. Equivalence classes comprising nonsense words were established and in a subsequent memory test, participants viewed lists that omitted two critical members of each class and then received a recognition test that included all class-members and stimuli not related to those members in classes.

Method

Participants

Participants were 10 undergraduate students, eight women and two men, aged between 18 and 21 years old, from a Brazilian University. Their native language was Portuguese. Informed consent was obtained from all individual participants included in the study. Participants received coupons that could be redeemed at the university cafeteria.

Equipment, Setting, and Stimuli

Individual sessions were conducted in a 2m × 3m quiet laboratory room, with one desk and one chair. Two computers were used in the study, a notebook with Windows® operational system and Apple Macintosh Performa 6360.

The experimental procedure was divided in two phases. In Phase 1, the software used to present stimuli and record responses was MTS v. 10.32 (Dube & Hiris, 1997); in Phase 2, another software was programmed to present a go/no-go task. The Portuguese version of the subtest Symbol Search of the Wechsler Intelligence Scale for Children (Wechsler, 2002) was used as a distracter task. In Phase 1, stimuli on Pre-training were three familiar pictures and three Portuguese words. In Classes 1, 2 and 3, established in Phase 1A, geometrical shapes were used as nodes (a rectangle; a circle and a six-point star). Other stimuli were nonsense words. In Classes 4, 5 and 6, established in Phase 1B, 36 nonsense words were used. Six lists of nonsense words were used in Phase 2.

Procedure

Data collection. In Phase 1 three 4-member and three 12-member equivalence classes were established using a 2 seconds delayed matching-to-sample (MTS) training. Phase 1 was completed in three or four one-hour-sessions, depending on participants' performance. Each session occurred in a different day. One week after the last session of Phase 1 participants came back to the laboratory to perform a memory test on Phase 2.

Phase 1: DMTS training. DMTS trials always began with the presentation of the sample on the center of the screen. A mouse click on the sample window was followed by removal of the sample and presentation, two seconds later, of the comparisons on the corners of the screen. A mouse click on the stimulus designated as correct was followed by a screen with stars, whereas a black screen followed the clicks on an incorrect stimulus. Comparison stimuli were gradually introduced in the initial trials of DMTS. On the first trial, only the correct comparison was presented (e.g., A1-B1); on the second trial, two comparisons were presented (e.g., A2-B2, B1); on the third trial, three comparisons were presented (e.g., A3-B3, B2, B1). Beginning at the fourth trial, three consecutive trials of each sample-comparison relation were presented consecutively, with three comparisons each and, finally, each relation was presented three times more, in a randomized sequence. Each block was repeated up to four times if the participant did not reach criterion. Criterion for ending the pre-training and the equivalence training and test were 90% correct responses in the block and in each relation.

Six Portuguese words were used as stimuli in eighteen DMTS trials in pre-training. After reaching the learning criterion, the participant began the next phase; otherwise, the block was repeated.

The simple to complex protocol was used to establish three 4-member equivalence classes using only nonsense words. In training blocks WX, WY, WZ baseline relations were trained to establish the three 4-member equivalence. Each relation was trained in a separate block, which consisted of 21 trials, all followed by differential feedback. After meeting criteria in each baseline training block, a symmetry test was conducted in a block comprising baseline trials and symmetry-test trials. Differential feedback was presented after baseline trials but not after test trials. Equivalence relations were tested in additional blocks (according to the simple-to-complex protocol).

In training blocks, when criteria were not met, the same block was repeated up to eight times. In test blocks, when criteria were not achieved, a full training block containing all relations trained up to that moment was presented and participants could repeat this block once. If a participant did not reach criteria even after repeating the full training block, a new training block was conducted containing the relation in which the participant did not reach criteria in the full training block. This sequence could be repeated up to four times. If a participant still did not meet criterion in the equivalence test, a Portuguese verbal prompt equivalent to "Pay attention on blocks in which the computer tells you whether your response is correct. You will find a clue to make correct choices without feedback" was presented. The participant had four more blocks to meet criterion. The minimum number of presentations of each stimulus was similar and all stimuli involved in equivalence tests were presented the same number of times on training and test blocks.

Immediately after formation of Classes 1, 2 and 3, participants initiated the procedure to form Classes, 4, 5 and 6. In training blocks AB, AC, AD, AE, AF, AG, AH, AI, AJ, AK, and AL, baseline relations were trained to establish the three 12-member equivalence classes using only nonsense words. Differently from Classes 1, 2 and 3, blocks comprising baseline trials and test-trials for symmetrical relations were not conducted. Equivalence relations were tested in different blocks. No differential consequences occurred in equivalence test blocks. Due to the large number of potential emergent relations, only 40 percent of them were tested. All stimuli appeared an equal number of times in test blocks. Table 1 describes this training and testing sequence.

Table 1
Sequence of Training and Test Blocks in Phase 1B, Relations in each Block and Numbers of Trials in each Block

Blocks	Relations	Trials
Training	AB	21
Training	AC	21
Equivalence Test 1	BC/CB	18
Training	AD	21
Equivalence Test 2	DC/CD	18
Training	AE	21
Equivalence Test 3	BE/EB/DE/ED	36
Training	AF	21
Equivalence Test 4	FC/CF/FE/EF	36
Training	AG	21
Equivalence Test 5	GD/DG/FG/GF	36
Training	AH	21
Equivalence Test 6	HE/EH/HG/GH	36
Training	AI	21
Equivalence Test 7	DI/ID/FI/IF/HI/IH	54
Training	AJ	21
Equivalence Test 8	JC/CJ/ GJ/JG	36
Training	AK	21
Equivalence Test 9	KB/BK/IK/KI/JK/KJ	54
Training	AL	21
Equivalence Test 10	LB/BL/LH/HL/LJ/JL/LK/KL	72

Phase 2: False memories test. Phase 2 was conducted one week after the last session of Phase 1. The session lasted about 15 minutes. Initially three lists of nonsense words were presented on a computer screen. Each list was comprised of 10 of the 12 stimuli of each equivalence class formed in Phase 1B (Classes 4, 5, and 6). Stimuli of each list were presented in sequence, in a randomized order. List 1A was composed of stimuli B5, C5, D5, E5, F5, G5, I5, J5, K5, L5; List 2A was composed of stimuli B4, C4, D4, F4, G4, H4, I4, J4, K4, L4; and List 3A of stimuli B6, C6, D6, E6, F6, G6, H6, I6, J6, L6. Therefore, the node and one more stimulus of each class did not appear in the lists. All stimuli were presented on the center of the screen, one at a time, for 2 seconds, so that the participant had enough time to read all stimuli. There was no break between lists, so the participant just saw a sequence of thirty stimuli, which was called study list. On the beginning of this task the participant was told s/he would see some words on the computer and all s/he had to do was to look at the words and try to memorize them.

After presentation of the study list finished, participants made a distracter task for three minutes. The task was the Symbol Search of the Wechsler Intelligence Scale for Children (Wechsler, 2002).

At the conclusion of the distracter task, the participant came back to the computer for a recognition test. On this test, three new lists were presented on the computer screen. The stimulus presentation was similar to the memorization task. These three new lists were comprised of all members of the equivalence classes that were in the study lists (targets), the stimuli of the equivalence classes that were not presented in the memorization study list (critical lures) and the three abstract stimuli on each list from Classes 1, 2 and 3 (non-related lures). Therefore, List 1B comprised stimuli A5, B5, C5, D5, E5, F5, G5, H5, I5, J5, K5, L5, X2, Y2, Z2; List 2B comprised stimuli A4, B4, C4, D4, E4, F4, G4, H4, I4, J4, K4, L4, X1, Y1, Z1; and List 3B, comprised stimuli A6, B6, C6, D6, E6, F6, G6, H6, I6, J6, K6, L6, X3, Y3, Z3. These lists taken together were called recognition list. The participant was instructed to identify, by pressing a button, on a go/no-go task, which stimuli of the recognition list had been presented on the study list.

Data analysis. Percentage of recognition in the false memories test was calculated for targets, critical distracters and non-related distracters for all participants. Comparison of recognition in each type of stimuli was calculated using Wilcoxon Matched Pairs test.

Ethical Considerations

This study was approved by the Ethics Committee on Human Research at the Universidade Federal de São Carlos (No. 273/2010).

Results

Six participants formed all equivalence classes and concluded the experiment. Two did not meet criteria in the last test, and two did not finish Phase 1 within the maximum time allowed.

Figure 1 shows the performance in the recognition test of only the six participants who formed the classes. Recognition of targets averaged about 75% and 85%. Only two participants recognized critical lures, and group average was 14%. Three participants recognized non-related lures and group average was 17%. Wilcoxon Matched Pairs test indicated significant difference in recognition between targets and critical lures ($p = 0.03$) and no significant difference between critical and non-related lures ($p = 0.87$). Results do not indicate typical performance on DRM lists.

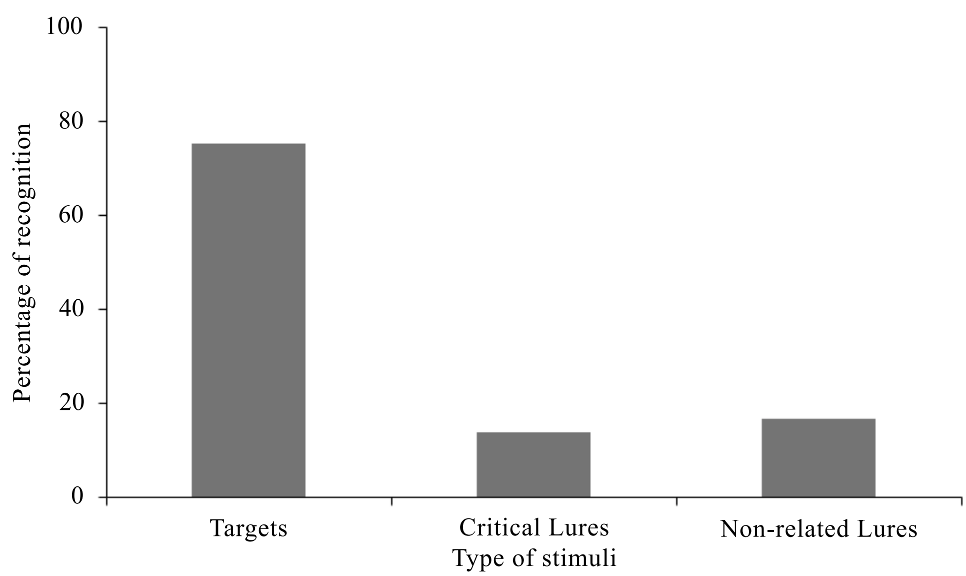

Figure 1. Average recognition of target, critical lures and non-related lures in recognition test.

Discussion

The procedure used in this study did not produce typical DRM results: participants did not show higher recognition of critical lures than non-related ones. These results can be seen as quite unexpected, considering that Guinther and Dougher (2010) showed robust effects of false memories in the kind of memory test used in this study, and these effects were replicated in a subsequent study (Guinther & Dougher, 2014). Further research is necessary to determine why false memories were not obtained, that is, what variables are responsible for the difference in results of both experiments.

This replication failure may be discussed in connection with the recent concerns about reproducibility of findings in Psychology, raised by a study (Open Science Collaboration, 2015) that attempted to replicate 100 studies published in psychological journals in 2008. The authors concluded that "a large portion of replications did not reproduce evidence supporting the original results". The controversy surrounding this study (Anderson et al., 2016; Gilbert, King, Pettigrew, & Wilson, 2016) brought reproducibility to the forefront of scientific discussion in Psychology. The problem about reproducibility may be compounded by *publication bias,* a tendency of journals to favor articles reporting studies with "positive data". The ensuing "file drawer" effect (Rosenthal, 1979) may result in an overestimation of the generality of reported effects.

On the other hand, replication failures may lead to additional research uncovering the effect of variables not considered in the original studies, which may limit the generality of the data. This study attempted to replicate the original study of Guinther and Dougher (2010), with several methodological changes. This was, therefore, a systematic replication, rather than a direct replication (Sidman, 1960). If the original results are reproduced in a systematic replication, the generality of the results is increased. If the results are not reproduced, scientific advances may result from the identification of variables responsible for the differences in the results. Another reason to pursue systematic replications is that direct replications, in which a study attempts to confirm previous findings by reproducing exactly the same method, may reproduce methodological faults of the original study, with the risk of confirming biased or flawed findings (Munafò & Smith, 2018).

One important difference between the experiment of Guinther and Dougher (2010) and the present study was the nature of stimuli that comprised equivalence classes. Guinther and Dougher used English words, with training conducted to build equivalence classes non-related to the semantic meaning of the words, that is, the semantic relations between the words were constructed in the study, by virtue of their acquired equivalence relations. When meaningless stimuli such as nonrepresentational pictures or nonsense words are used in equivalence studies, these stimuli become semantically related, as documented by several different measures (Barnes-Holmes et al., 2005; Bortoloti & de Rose, 2011b, 2012; Bortoloti et al., 2014; Haimson et al., 2009). Equivalent stimuli, even when they are meaningless before the study, may substitute one another and acquire a similar meaning. Behavioral functions acquired by the stimuli are transferred to the equivalent stimuli, so that if the equivalent stimuli B1, C1, D1, etc., acquired "remembering functions" (Guinther & Dougher, 2010) because they were in a list of stimuli to be remembered, these functions should transfer to a stimulus equivalent to them, such as E1, not present in the list, increasing the probability that E1 would be falsely remembered. This effect should not depend on the nature of the stimuli.

However, several studies have found a paradoxical effect in classes of equivalent stimuli, which is a variation in the *degree of relatedness* between stimuli (Bortoloti & de Rose, 2011a; Doran & Fields, 2012). There are some indications that the nature of stimuli may affect the degree of relatedness between stimuli in equivalence classes (Fields et al., 2012; Mensah & Arntzen, 2017; Silveira et al., 2016; Travis, Fields, & Arntzen, 2014). If degree of relatedness translates into degree of semantic relation, then it is possible that the nature of stimuli may affect the degree of semantic relation, a variable that predicts the probability of false memories (Roediger et al., 2001).

A second difference between this study and Guinther and Dougher (2010) is that we conducted memory tests one week after the training and testing for equivalence class formation. If equivalence between critical lures and targets were the relevant variable, demonstration of false memories would depend on the stability of classes during this one-week period. Many studies show that equivalence classes tend to be stable for relatively long periods (Camargo & Haydu, 2015; Silveira et al., 2016). However, it is conceivable that stability of classes may be influenced by the size and number of classes trained. This may also interact with the nature of the stimuli. Silveira et al. (2016) evaluated the stability of three equivalence classes comprising arbitrary stimuli and one meaningful stimulus, with positive, neutral, or negative valence. The class that comprised the positive stimulus was found to be more stable in a test conducted after 30 days. The interval between equivalence formation and memory tests in this study was relatively short compared to the intervals investigated in studies of stability of classes. Future researches, however, should conduct equivalence tests after the memory test, to assess maintenance of the classes.

The present experiment also differed from Guinther and Dougher (2010) in several aspects of the training and testing protocol. Guinther and Dougher established three equivalence classes, each comprising 24 words. Their study list was comprised of 12 words from one of the classes. The remaining 12 words of this class were the critical lures, whereas the words from the other classes were unrelated lures. A free recall test was conducted first, followed by a recognition test, in which words from the three classes were presented in a random order. The present study established initially Equivalence Classes 1, 2, and 3, each with four words, and then Equivalence Classes 4, 5, and 6, each with 12 words. The memory test presented three lists, each comprising 10 of the 12 stimuli from Classes 4, 5, and 6. Because stimuli were nonsense words, a recall test was not conducted. The recognition test presented the remaining two stimuli from Classes 4, 5, and 6 as critical lures, and stimuli from Classes 1, 2, and 3 as non-related lures. It is not clear how these differences may have affected the results.

Other variables should be addressed in future experiments, such as (1) the nature of the stimuli (stimuli from Classes 1, 2 and 3 were related to geometrical shapes, whereas stimuli from Classes 4, 5 and 6 were all nonsense words); (2) non-critical lures been part of 4-member equivalence classes, whereas critical lures were part of 12-member equivalence classes; (3) formation of Classes 1, 2 and 3 was completed before Classes 4, 5 and 6; (4) although the experimental approach in behavior analysis tends to use small numbers of participants (Sidman, 1960), replications with larger groups should be attempted.

False memories in the DRM paradigm are an effect of the semantic relations between words. As Guinther and Dougher (2010) noted, the stimulus equivalence paradigm permits to construct semantic relations between stimuli. In equivalence studies, such stimuli are usually meaningless pictures or nonsense words, although meaningful stimuli are sometimes used in equivalence studies. Guinther and Dougher used common words, but they are semantically non-related in the language, so that the semantic relations were constructed during the study. They showed a strong effect of false memories, which is coherent with the expectation that equivalent stimuli will substitute for one another and share stimulus functions. Therefore, remembering functions were transferred to class-members that did not appear in the memory lists. As Guinther and Dougher (2010) noted, the possibility of experimentally constructing semantic relations may provide an important contribution to the understanding of false memories in the DRM paradigm. For this, it is vital to establish the conditions in which the effect can be reproduced, and to report conditions in which it cannot, as done in the present report.

References

Aggio, N. M., Pedrosa, S. C. D., & de Rose, J. C. (2017). Falsas memórias na perspectiva da análise do comportamento. *Acta Comportamentalia, 25*(3), 379-393. Retrieved from http://revistas.unam.mx/index.php/acom/article/view/61633/54292

Amd, M., Barnes-Holmes, D., & Ivanoff, J. (2013). A derived transfer of eliciting emotional functions using differences among electroencephalograms as a dependent measure. *Journal of the Experimental Analysis of Behavior, 99*(3), 318-334. doi:10.1002/jeab.19

Anderson, C. J., Bahnik, S., Barnett-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., … Zuni, K. (2016). Response to comment on Estimating the reproducibility of psychological science. *Science, 351*(6277), 1037. doi:10.1126/science.aad9163

Barnes-Holmes, D., Staunton, C., Whelan, R., Barnes-Holmes, Y., Commins, S., Walsh, D., … 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), 417-433. doi:10.1901/jeab.2005.78-04

Bennett, M. P., Meulders, A., Baeyens, F., & Vlaeyen, J. W. (2015). Words putting pain in motion: The generalization of pain-related fear within an artificial stimulus category. *Frontiers in Psychology, 6*, 520. doi: 10.3389/fpsyg.2015.00520

Bortoloti, R., & de Rose, J. C. (2011a). An Orwellian account of stimulus equivalence. Are some stimuli "more equivalent" than others? *European Journal of Behavior Analysis, 12*(1), 121-134. doi: 10.1080/15021149.2011.11434359

Bortoloti, R., & de Rose, J. C. (2011b). Avaliação do efeito de dica semântica e da indução de significado entre estímulos abstratos equivalentes. *Psicologia: Reflexão e Crítica, 24*(2), 381-393. doi: 10.1590/S0102-79722011000200020

Bortoloti, R., & de Rose, J. C. (2012). Equivalent stimuli are more strongly related after training with delayed matching than after simultaneous matching: A study using the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record, 62*(1), 41-54. doi: 10.1007/BF03395785

Bortoloti, R., Pimentel, N., & de Rose, J. C. (2014). Electrophysiological investigation of the functional overlap between semantic and equivalence relations. *Psychology & Neuroscience, 7*(2), 183-191. doi: 10.3922/j.psns.2014.015

Bortoloti, R., Rodrigues, N. C., Cortez, M. D., Pimentel, N., & de Rose, J. C. (2013). Overtraining increases the strength of equivalence relations. *Psychology & Neuroscience, 6*(3), 357-364. doi: 10.3922/j.psns.2013.3.13

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. doi: 10.1037/h0101068

Challies, D. M., Hunt, M., Garry, M., & Harper, D. N. (2011). Whatever gave you that idea? False memories following equivalence training: A behavioral account of the misinformation effect. *Journal of the Experimental Analysis of Behavior, 96*(3), 343-362. doi: 10.1901/jeab.2011.96-343 de Almeida, J. H., Bortoloti, R., Ferreira, P. R. S., Schelini, P. W., & de Rose, J. C. (2014). Análise da validade e precisão de instrumento de diferencial semântico. *Psicologia: Reflexão e Crítica, 27*(2), 272-281.doi: 10.1590/1678-7153.201427207

de Almeida, J. H., & de Rose, J. C. (2015). Changing the meaningfulness of abstract stimuli by the reorganization of equivalence classes: Effects of delayed matching. *The Psychological Record, 65*(3), 451-461. doi: 10.1007/s40732-015-0120-9

de Rose, J. C., McIlvane, W. J., Dube, W. V., & Stoddard, L. T. (1988). Stimulus class formation and functional equivalence in moderately retarded individual's conditional discrimination. *Behavioural Processes, 17*(2), 167-175. doi: 10.1016/0376-6357(88)90033-2

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17-22. Retrieved from http://psycnet.apa.org/fulltext/1960-03918-001.pdf

Doran, E., & Fields, L. (2012). All stimuli are equal, but some are more equal than others: Measuring relational preferences within an equivalence class. *Journal of the Experimental Analysis of Behavior, 98*(3), 243-256. doi: 10.1901/jeab.2012.98-243

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62*(3), 331-351. doi: 10.1901/jeab.1994.62-331

Dube, W. V., & Hiris, J. (1997). *Match to Sample Program (Version 11.1.3)* [Computer software]. Waltham, MA: E. K. Shriver Center for Mental Retardation.

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. doi:10.1901/jeab.2012.97-163

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the illusion. *Memory & Cognition, 38*(7), 833-848. doi: 10.3758/MC.38.7.833

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on Estimating the reproducibility of psychological science. *Science, 351*(6277), 1037. doi: 10.1126/science.aad7243

Guinther, P. M., & Dougher, M. J. (2010). Semantic false memories in the form of derived relational intrusions following training. *Journal of the Experimental Analysis of Behavior, 93*(3), 329-347. doi: 10.1901/jeab.2010.93-329

Guinther, P. M., & Dougher, M. J. (2014). Partial contextual control of semantic false memories in the form of derived relational intrusions following training. *The Psychological Record, 64*(3),457-473. doi: 10.1007/s40732-014-0012-4

Haimson, B., Wilkinson, K. M., Rosenquist, C., Ouimet, C., & McIlvane, W. J. (2009). Electrophysiological correlates of stimulus equivalence processes. *Journal of the Experimental Analysis of Behavior, 92*(2), 245-256. doi: 10.1901/jeab.2009.92-245.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Mensah, J., & Arntzen, E. (2017). Effects of meaningful stimuli contained in different numbers of classes on equivalence class formation. *The Psychological Record, 67*(3), 325-336. doi: 10.1007/s40732-016-0215-y

Munafò, M. R., & Smith, G. D. (2018). Repeating experiments is not enough. *Nature, 553*(7689), 339-401. doi: 10.1038/d41586-018-01023-3

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science.*Science, 349*(6251), aac4716. doi: 10.1126/science.aac4716

Perez, W. F., de Almeida, J. H., & de Rose, J. C. (2015). Transformation of meaning through relations of sameness and opposition. *The Psychological Record, 65*(4), 679-689. doi: 10.1007/s40732-015-0138-z

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814. doi: 10.1037/0278-7393.21.4.803

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-407. doi: 10.3758/BF03196177

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. doi: 10.1037/0033-2909.86.3.638

Sidman, M. (1960). *Tactics of scientific research*. Oxford, United Kingdom: Basic Books.

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

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. doi: 10.1007/s40732-015-0152-1

Tabullo, A., Yorio, A., Zanutto, S., & Wainselboim, A. (2015). ERP correlates of priming in language and stimulus equivalence: Evidence of similar N400 effects in absence of semantic content. *International Journal of Psychophysiology, 96*(2), 74-83. doi: 10.1016/j.ijpsycho.2015.03.004

Travis, R. W., Fields, L., & Arntzen, E. (2014). Discriminative functions and over-training as class-enhancing determinants of meaningful stimuli. *Journal of the Experimental Analysis of Behavior, 102*(1), 47-65. doi: 10.1002/jeab.91

Wechsler, D. (2002). *WISC-III: Escala de inteligência Wechsler para Crianças: Manual* (V. L. M. Figueiredo, Trad., 3aed.). São Paulo, SP: Casa do Psicólogo.

Natalia Maria Aggio holds a Ph.D. from the Universidade Federal de São Carlos, São Carlos-SP, Brazil.

Julio César de Rose is a Full Professor of the Universidade Federal de São Carlos, São Carlos-SP, Brazil.

Received: Oct.18, 2016
1st Revision: Sep.05, 2017
2nd Revision: Mar.05, 2018
Approved: Mar.23, 2018

How to cite this article:
Aggio, N. M., & de Rose, J. C. (2018). Failure to produce false memories through the stimulus equivalence paradigm. *Paidéia (Ribeirão Preto), 28,* e2826. doi:http://dx.doi.org/10.1590/1982-4327e2826