

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA

**FORMAÇÃO E MANUTENÇÃO DE CLASSES DE EQUIVALÊNCIA: EFEITOS DA  
MAGNITUDE DAS CONSEQUÊNCIAS E DO USO DE UM PROCEDIMENTO  
GAMIFICADO**

Alceu Regaço dos Santos

São Carlos – SP

2021

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA

**Formação e manutenção de classes de equivalência: efeitos da magnitude das  
consequências e do uso de um procedimento gamificado**

Dissertação apresentada ao Programa de Pós-  
Graduação em Psicologia da Universidade  
Federal de São Carlos como requisito para a  
obtenção do título de mestre, sob orientação da  
Dra. Deisy das Graças de Souza.

São Carlos – SP

2021



**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

Centro de Educação e Ciências Humanas  
Programa de Pós-Graduação em Psicologia

---

**Folha de Aprovação**

---

Defesa de Dissertação de Mestrado do candidato Alceu Regaço dos Santos, realizada em 22/10/2021.

**Comissão Julgadora:**

Profa. Dra. Deisy das Graças de Souza (UFSCar)

Profa. Dra. Verônica Bender Haydu (UEL)

Profa. Dra. Maria Helena Leite Hunziker (USP)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Psicologia.

“Controle aversivo faz parte da vida;  
nem sempre é ruim, nem sempre é evitável.”

João Cláudio Todorov (2011)

### **Apoio Financeiro**

O autor contou com bolsa da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (Processo No. 2019/01248-2) durante o período de julho de 2019 a setembro de 2021. Houve também o apoio Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES/PROEX no período de março a junho de 2019. A pesquisa foi conduzida no âmbito do Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino, financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Processo No. 465686/2014-1) e pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Processo No. 2014/50909-8). As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do(s) autor(es) e não necessariamente refletem a visão da FAPESP.

## **Agradecimentos**

Eu tenho tanta gente para agradecer, que é até difícil saber por onde começar, afinal, eu contei com tanto apoio para conseguir chegar até aqui, seja acadêmico, pessoal, financeiro ou familiar, que eu acho injusto colocar uma ordem de importância ou acabar esquecendo alguém (o que certamente aconteceu). Então, espero ter conseguido descrever a devida relevância que cada uma dessas pessoas tem (e que tantas outras pessoas também têm, mas que eu não coloquei pelo limite de espaço ou pelos limites da minha memória), saibam que vocês são todos muito queridos por mim.

Às minhas professoras e professores, eu agradeço imensamente à Deisy e ao Júlio, que me orientaram durante todo o mestrado e, praticamente, toda a graduação. Eu sempre penso que tenho muita sorte por trabalhar com vocês dois e poder aprender tanto com cada um, e a minha vontade é fazer todos os meus projetos com a orientação em conjunto, para que eu sempre possa aproveitar a quantidade de coisas que eu tenho a aprender com vocês dois. Aqui, eu também agradeço à Mariéle e ao Marcelo, que me “adotaram” na graduação, me ensinaram e incentivaram a seguir na área de pesquisa, e me aguentaram quando eu achava que já sabia tanta coisa (sendo que eu não sabia nada rs). Agradeço ao Will, pelas supervisões que tivemos desde o começo do mestrado; apesar das conturbações da pandemia, suas sugestões foram fundamentais para que eu melhorasse o meu projeto. Também agradeço aos membros da banca de qualificação pelas correções e sugestões que contribuíram para a versão final dessa dissertação, e agradeço às professoras da banca de defesa, que aceitaram avaliar este trabalho.

Aos meus colegas de laboratório, eu agradeço especialmente ao Filipe, porque sem a ajuda dele eu nunca teria conseguido realizar a adaptação do procedimento e, conseqüentemente, o Experimento 3. Agradeço também ao Ramon e à Heloisa pelas parcerias em projetos, discussões, pesquisas e congressos. Agradeço muito ao Júlio Camargo, à Denise e ao Giovan, que foram companheiros cotidianos do laboratório e me ajudaram muito com a análise

estatística do Experimento 3. E agradeço aos outros membros dos grupos de pesquisa do qual eu faço parte, que me auxiliaram, seja em discussões, apresentações ou na análise dos dados: Murilo, Junior, Marlon, Renata e Djenane.

Às minhas amigas e amigos eu só tenho a agradecer por todo apoio e carinho que vocês têm comigo. À Mariana, meu benzinho, eu agradeço por todo o companheirismo e parceria, terminar essa dissertação sem você ao meu lado não seria a mesma coisa (ainda mais durante essa pandemia). E à minha família, principalmente meus pais Vanderson e Áurea, agradeço imensamente o incentivo à educação e o exemplo que vocês dão sobre a importância do estudo e do esforço. Também não posso deixar de agradecer à Leila, minha psicóloga, que esteve comigo durante o início do mestrado, mas que, fundamentalmente, me auxiliou a me entender e a lidar melhor com diversas questões da minha vida e do meu futuro.

Também agradeço às agências de fomento FAPESP, CAPES e CNPq pelos investimentos, pois sem bolsa e sem recursos não seria possível realizar essa pesquisa. E agradeço também aos políticos que já lutaram e ainda lutam por mais recursos para a educação.

Por fim, quero agradecer a você, que está lendo esse texto agora, sem você nada do que está escrito aqui faria sentido, espero que você aproveite a leitura dessa dissertação e que ela lhe seja útil.

## Sumário

Resumo.....	6
Introdução.....	7
Parte 1 – Experimentos 1 e 2.....	14
Método .....	14
Participantes .....	14
Equipamentos e Instrumentos .....	14
Procedimento.....	19
Resultados .....	25
Experimento 1 .....	27
Experimento 2 .....	30
Discussão.....	32
Parte 2 – Gamificação do Procedimento e Experimento 3.....	37
Parte 2.1 – Criação do jogo “Miner Troubles”.....	38
Software Miner Troubles .....	41
Parte 2.2 – Experimento 3: O efeito da magnitude das consequências no ensino de relações de equivalência.....	48
Método .....	49
Resultados .....	53
Discussão.....	68
Conclusão Geral .....	77
Referências .....	79



Regaço, A. (2021) Formação e manutenção de classes de equivalência: efeitos da magnitude das consequências e do uso de um procedimento gamificado [Dissertação de Mestrado, Programa de Pós-Graduação em Psicologia, Universidade Federal de São Carlos].

### **Resumo**

Estudos têm mostrado que a relação entre estímulos de classes de equivalência pode variar em função de diversos parâmetros experimentais. Entretanto, poucos estudos investigaram o efeito das consequências como variável independente. Este estudo está dividido em três experimentos, que têm como objetivo geral avaliar o efeito do uso de consequências reforçadoras e/ou punitivas na formação e manutenção de relações de equivalência. Os Experimentos 1 e 2 tiveram como objetivo específico comparar o uso de contingências de reforço e/ou de punição no ensino das relações de duas potenciais classes de estímulos com quatro membros cada. Para isso, foi utilizado um procedimento tradicional de Matching-to-Sample (MTS) e os participantes foram distribuídos em três grupos de acordo com as consequências utilizadas: Grupo Reforço (GR), Grupo Reforço-Punição (GR-P) e Grupo Punição (GP). Um maior número de participantes do GR atingiu o critério de formação das classes de equivalência em comparação aos outros grupos. O Experimento 3 teve como objetivo específico a avaliação do efeito da magnitude das consequências de reforçamento e de punição na formação e manutenção das classes de equivalência. Para a coleta de dados foi utilizada uma versão gamificada do MTS (o jogo Miner Troubles, desenvolvido para o estudo), adaptada para a realização de coletas online, devido às limitações decorrentes da pandemia de Covid-19. Participaram da pesquisa 56 pessoas (entre 18 e 57 anos), distribuídas aleatoriamente entre três grupos: Grupo Mais-Ganho (GmG), Grupo Balanceado e Grupo Mais-Perda (GmP). No primeiro grupo, o participante ganhava 4 pontos ao acertar e perdia 1 ponto ao errar; no segundo, o participante ganhava 1 ponto ao acertar e perdia 1 ponto ao errar; e no terceiro grupo o participante ganhava 1 ponto ao acertar e perdia 4 pontos ao errar. Os resultados mostraram que o uso de uma maior magnitude de punição prejudicou a aprendizagem de relações condicionais e a emergência das relações não treinadas: o desempenho dos participantes do GmP foi menos acurado nos Testes de Simetria e de Manutenção das relações em comparação aos outros grupos. Este experimento estende os achados de outros estudos sobre o efeito da punição na aprendizagem discriminativa e enseja discussão sobre o uso de punição de forma mais intensa na aprendizagem.

*Palavras-chave:* magnitude das consequências, classes de equivalência, manutenção de classes, gamificação, *matching-to-sample*.

## Introdução

Sidman e Tailby (1982) definiram relações de equivalência entre estímulos com base nas propriedades de reflexividade, simetria e transitividade: após o ensino de discriminações condicionais<sup>1</sup> (como AB e BC), a reflexividade seria demonstrada por relações de identidade (AA, BB, CC), a simetria seria demonstrada por relações com os termos revertidos (BA, CB) e a transitividade seria demonstrada pela relação entre dois termos relacionados a um terceiro (AC). Todas estas relações que documentam as propriedades de equivalência devem *emergir*, ou seja, ser *derivadas* das relações ensinadas, sem que elas próprias tenham sido ensinadas, desta forma, devem ocorrer na ausência de reforço diferencial. Esta definição foi posteriormente refinada por Sidman (1994; 2000), que passou a definir relações de equivalência como pares ordenados de todos os elementos positivos que participam de uma contingência. A principal mudança seria incluir, nas classes formadas pelos estímulos equivalentes, outros elementos da contingência, além dos estímulos antecedentes. Assim, no exemplo dado, teríamos que considerar que as discriminações condicionais AB e BC envolvem respostas e consequências para respostas corretas e, em determinadas condições, estes termos da contingência, principalmente as consequências, podem fazer parte das classes (Dube et al., 1987, 1989; Goyos, 2000; Silveira et al., 2017, 2018).

Uma das consequências da formação de classes de estímulos equivalentes é a transferência de função. Este fenômeno ocorre quando o controle exercido por um dos estímulos da classe se estende aos demais estímulos equivalentes a ele (de Rose et al., 1988; Dougher et al., 1994; Hayes et al., 1991; Perez et al., 2015). Grande parte dos estudos sobre transferência de funções estabelece, primeiramente, classes de estímulos arbitrárias, como no estudo clássico de Dougher et al. (1994): depois da formação de classes de estímulos abstratos, um dos

---

1 Sidman e Tailby (1982) distinguiram entre “discriminação condicional” e “*matching-to-sample*” (emparelhamento com o modelo). A distinção é complexa e não pode ser tratada aqui para não alongar demasiadamente o texto. Neste texto vamos tomar discriminação condicional e *matching-to-sample* como termos que designam um mesmo *procedimento* e podem ser usados de modo intercambiável.

membros de uma das classes foi emparelhado com choque elétrico e tornou-se estímulo condicionado eliciador de uma resposta eletro dérmica (GSR). Esta função eliciadora passou a ser exercida também pelos outros membros da classe, que nunca haviam sido emparelhados com choque.

Um número menor de estudos utilizou estímulos que possuem significado e, portanto, já apresentariam funções específicas (cf. Fields & Arntzen, 2018), para verificar se estas funções se transfeririam para estímulos equivalentes (e.g., Barnes-Holmes et al., 2000; Santos & Rose, 2018) ou para verificar o efeito destas funções sobre a formação de relações de equivalência (e.g., Arntzen et al., 2015). Estudos desenvolvidos na UFSCar, em particular, têm utilizado expressões faciais de emoções como estímulos significativos em classes de equivalência (Bortoloti & de Rose, 2009, 2012; Silveira et al., 2016).

No estudo de Bortoloti e de Rose (2009), expressões de alegria, neutralidade e raiva foram, cada uma delas, estabelecidas como equivalentes a vários estímulos abstratos. Considerando que as expressões faciais teriam um significado universal, de origem filogenética (Darwin & Prodger, 1998; Ekman & Friesen, 1971), estes autores queriam verificar se o significado das expressões seria adquirido por estímulos abstratos. Para isso, utilizaram um instrumento bastante usado para medidas quantitativas<sup>2</sup> de significado de palavras, locuções, figuras, etc., o Diferencial Semântico (Engelmann, 1978; Osgood et al., 1957; ver mais detalhes na seção do Método). Comparando os escores de diferencial semântico após a formação das classes de equivalência, Bortoloti e de Rose (2009) verificaram que os estímulos abstratos adquiriram significado, ou mais especificamente, valência semelhante à da expressão emocional à qual

---

<sup>2</sup> Mais explicitamente, o Diferencial Semântico permite quantificar dimensões do significado como, por exemplo, valência, potência e atividade, que são fatores extraídos em estudos psicométricos envolvendo análise fatorial. Com base na análise fatorial do Diferencial Semântico usado nos estudos de Bortoloti e de Rose (Almeida et al., 2014), a dimensão de significado que apresentou resultados significativos nos estudos foi a dimensão avaliativa, ou *valência*, ou seja, a atratividade ou repulsividade do estímulo (cf., Amd et al., 2013; Fields & Arntzen, 2018), de modo que quando tratamos de medida de significado neste texto, estamos tratando especificamente da valência. É possível que outras dimensões de significado também sejam transferidas, mas o instrumento só tinha precisão para medir a valência.

eram equivalentes. Ou seja, a valência da expressão emocional (conforme medida pelo diferencial semântico) se transferiu para os estímulos abstratos equivalentes.

Com base nos escores do diferencial semântico, foi possível verificar a magnitude da transferência em função de vários parâmetros. Verificou-se que a transferência diminuía em função da distância nodal (número de nódulos que separavam o estímulo abstrato da expressão facial; cf., Fields et al., 1984), e a apresentação simultânea dos estímulos modelo e comparação no procedimento de discriminação condicional (*simultaneous matching-to sample*). Subsequentemente, Bortoloti, Rodriguez, Cortez, Pimentel, e de Rose (2013) mostraram que a magnitude da transferência também é afetada pela duração do treino de discriminação condicional; também há indícios de que a magnitude da transferência é maior para expressão de alegria do que de raiva (Bortoloti et al., 2019, 2021; Bortoloti & de Rose, 2011; Silveira et al., 2016).

Fields e colegas já haviam sugerido, com base na probabilidade de transferência, que relações de equivalência variam em “relatedness”, ou seja, no “grau de relacionamento” entre os estímulos pertencentes à classe (Albright et al., 2019; Fields et al., 1995). Os dados de Bortoloti e de Rose (2009) acrescentaram novas evidências a esta proposição, na medida em que mostraram variações na magnitude da transferência e não apenas na probabilidade. Estas considerações levaram Bortoloti e de Rose (2011a) e Doran e Fields (2012) a apontar um paradoxo na definição de equivalência, que os autores destes dois artigos compararam à afirmação irônica de George Orwell em “A Revolução dos Bichos” (Orwell, 2021) de que “todos os animais são iguais, mas alguns são mais iguais do que os outros”.

Tanto Bortoloti e de Rose quanto Doran e Fields (ver também Sidman, 1994) discutiram possibilidades de resolver esta contradição teórica entre equivalência e variação em graus de relacionamento (ou seja, não equivalência). Entretanto, não é este o objetivo deste estudo.

Dois outros estudos que utilizaram expressões emocionais nos interessam particularmente porque realizaram outras medidas, além do Diferencial Semântico, para avaliar a magnitude. Bortoloti & de Rose (2012) usaram um delineamento de grupo, em que um grupo recebeu treino de *matching-to-sample* simultâneo (SMTS) e outro recebeu treino de *matching-to-sample* com atraso (DMTS). No SMTS, o modelo era apresentado na tela, o participante clicava sobre ele com o *mouse*, produzindo a apresentação dos estímulos de comparação, enquanto o modelo permanecia na tela. No DMTS, o participante clicava sobre o modelo e este era removido, seguindo-se um intervalo (atraso, neste caso de 2 s) após o que os estímulos de comparação eram apresentados, na ausência do modelo. Estes procedimentos foram usados para estabelecer, para cada um dos grupos, classes de quatro estímulos equivalentes. Um dos objetivos do experimento foi confirmar, com outro método, além do Diferencial Semântico, os resultados de Bortoloti e de Rose (2009), que haviam mostrado maior grau de relacionamento nas classes de equivalência estabelecidas por DMTS. Neste estudo de 2012, os autores confirmaram estes resultados utilizando o *Implicit Relational Assessment Procedure* (IRAP, ver detalhes sobre o procedimento no Método).

Silveira et al. (2016) recorreram à medida de manutenção das classes para investigar a diferença em graus de relacionamento entre estímulos. Neste estudo foram estabelecidas três classes de estímulos equivalentes (A1B1C1D1, A2B2C2D2 e A3B3C3D3). Em cada classe, os estímulos A eram faces expressando alegria (A1), neutralidade (A2) e raiva (A3) e os demais estímulos eram figuras abstratas. Após os testes de relações emergentes, escores de diferencial semântico evidenciaram a transferência do significado das faces para os estímulos abstratos. Os participantes retornaram após 30 dias e foram submetidos a testes de equivalência e avaliações com o diferencial semântico. As medidas de manutenção das classes e os escores de diferencial semântico mostraram que a classe de estímulos relacionada à expressão alegre resistiu mais à passagem do tempo, indicando maior grau de

relacionamento entre os estímulos desta classe. Este estudo mostrou também que a manutenção das classes após um intervalo de tempo pode fornecer indicações sobre o grau de relacionamento.

Resumindo o que foi encontrado pelos autores, algumas variáveis aumentaram o grau de relacionamento entre os estímulos, tais como: uso de DMTS (Bortoloti & de Rose, 2009, 2012), estímulos com valência “positiva” (Silveira et al., 2016) e quantidade de treino (Bortoloti et al., 2013). Outras variáveis, no entanto, geraram uma diminuição da força das relações, como o uso de SMTS, estímulos com valência “negativa” e a distância nodal das relações testadas (Fields et al., 1984). Esses resultados foram registrados a partir de diferentes procedimentos, como o IRAP, o diferencial semântico e o teste de manutenção das classes de equivalência.

Esses achados compõem uma base empírica importante para a compreensão e o refinamento conceitual da aprendizagem de relações arbitrárias entre estímulos e da equivalência de estímulos. Contudo, um aspecto pouco estudado é o efeito que consequências aversivas têm na formação de classes de equivalência e, conseqüentemente, no grau de relacionamento dos estímulos. Esta lacuna torna-se ainda mais significativa se considerarmos que o uso e o estudo da punição é um tema controverso e pouco estudado (c.f., Hunziker, 2017; Sidman, 1989; Todorov, 2001, 2011), mas que continua sendo um procedimento muito comum nos métodos tradicionais de ensino e muito presente em outras situações cotidianas (Gershoff, 2017; Klevens et al., 2019; Sidman, 1989; Skinner, 1965).

Em um estudo sobre discriminação simples, Penney e Lupton (1961) testaram três procedimentos de ensino em três grupos diferentes: no Grupo R (*Reinforcement Only*) eram reforçadas as respostas corretas; no Grupo P (*Punishment Only*) era punidas as respostas incorretas; e no Grupo P-R eram reforçadas as respostas corretas e punidas as respostas incorretas. Como estímulo punitivo foi utilizado um tom intenso aversivo e como estímulo

reforçador foram utilizados doces. Este estudo demonstrou que o uso de punição junto ao reforçamento foi mais eficaz do que a utilização exclusiva de reforçamento, confirmando os resultados de estudos realizados anteriormente (Brackbill & O'Hara, 1958). E, além disso, o uso de punição exclusivamente foi mais eficaz do que o procedimento que utilizava reforçamento e punição de forma conjunta. Este dado foi replicado não só por outros estudos deste mesmo período e com tarefas semelhantes (e.g., Costantini & Hoving, 1973; Meyer & Offenbach, 1962; Penney, 1967), mas também com outros tipos de tarefas, como o seguimento de regras em procedimentos de discriminação condicional (Cortez et al., 2020; Perez et al., 2010).

De forma geral, os estudos demonstraram que a utilização de punição, seja exclusivamente ou em conjunto com o reforçamento, facilita a aprendizagem discriminativa quando comparada ao uso exclusivo de reforçamento. Entretanto, essa variável ainda não foi estudada no ensino de relações condicionais, em que é prevista a emergência de novas relações não treinadas e nas quais a dificuldade da tarefa é maior do que em tarefas de discriminação simples. Verificar o efeito das consequências nessa tarefa pode auxiliar a desenvolver procedimentos eficazes para o ensino de comportamentos socialmente relevantes, como a leitura e escrita (de Rose et al., 1996).

Além disso, o estudo dos tipos de consequências empregadas nos procedimentos de ensino, incluindo punição, pode trazer algumas implicações também para a teoria sobre relações de equivalência. Considerando que as consequências podem fazer parte das classes de equivalência (Dube et al., 1987; Sidman, 1994, 2000), é inevitável questionar se a utilização de estímulos aversivos punitivos alteraria a função dos estímulos em uma classe de estímulos equivalentes. Sidman (2000) definiu que apenas os pares de elementos positivos fariam parte das classes de equivalência, desta forma, estímulos não relacionados à contingência de reforçamento (como as sinalizações de erro) não fariam parte da classe formada. Neste caso,

ocorreria um processo de reforçamento negativo: o participante se comporta de forma a evitar a ocorrência da punição. E se as classes de equivalência abrangem as consequências de reforço, no caso do reforçamento negativo, é necessário verificar se os estímulos punitivos poderiam se tornar parte das classes. Uma forma de se verificar isso é avaliando se o uso de punição (e, conseqüentemente, de reforçamento negativo) pode alterar a avaliação dos estímulos da classe, assim como ocorre quando é utilizado reforçamento positivo (Silveira et al., 2018).

Desta forma, o objetivo do presente estudo foi verificar o efeito do uso de reforço positivo e punição na formação e manutenção de classes de equivalência. Nos Experimentos 1 e 2, as consequências reforçadoras positivas consistiram em *feedback* de acerto para respostas corretas, que tendem a funcionar como reforçador positivo para virtualmente todos os participantes. As consequências punitivas consistiram na apresentação de um *feedback* de erro, acompanhado de um som presumivelmente aversivo, após respostas incorretas; estas consequências aversivas poderão, portanto, ser evitadas pela emissão de respostas corretas. Estes dois experimentos serão apresentados de forma conjunta, e as diferenças entre eles foram sinalizadas ao longo do método descrito na Parte 1.

O Experimento 3 tem o mesmo objetivo e foi feito utilizando o jogo Miner Troubles, uma versão gamificada do *Matching-to-Sample* (ver Parte 2.1). Devido à pandemia de Covid-19 o procedimento e as consequências tiveram que ser adaptados para serem realizados de maneira remota, online. Desta forma, a variável independente manipulada neste caso foi a magnitude das consequências para acerto e erro (ver mais detalhes na Introdução e Método do Experimento 3). Essas manipulações foram realizadas para que se mantivesse parte das condições estabelecidas nos Experimento 1 e 2 (ver método – Parte 2.2) e de forma que a pergunta de pesquisa continuasse a mesma.



## **Parte 1 – Experimentos 1 e 2**

### **Método**

#### ***Participantes***













Participaram da pesquisa 15 estudantes universitários no Experimento 1 e 12 no Experimento 2, todos sem experiência em pesquisas envolvendo discriminações condicionais e que não estavam cursando psicologia. Antes do início do experimento, os participantes leram e assinaram um Termo de Consentimento Livre e Esclarecido (TCLE). O projeto e o TCLE foram aprovados pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de São Carlos (Processo nº. 10923019.7.0000.5504).

#### ***Equipamentos e Instrumentos***

Todas as etapas do procedimento foram realizadas em uma sala do Laboratório de Estudos do Comportamento Humano, no Edifício Carolina Bori, da Universidade Federal de São Carlos. Um computador com o software *PsychoPy3* (Peirce, 2007) foi utilizado para apresentação de estímulos e registro de respostas dos participantes. Foram utilizadas, como estímulos, 12 figuras abstratas em preto e branco, apresentadas na Figura 1. Oito delas foram usadas para constituir duas potenciais classes de estímulos equivalentes (A1/B1/C1/D1 e A2/D2/C2/D2). As outras figuras (A3, B3, C3 e D3) foram usadas como estímulos “nulos”, cuja seleção nunca era reforçada. Elas tinham como função diminuir a probabilidade de controle por rejeição sobre a formação das classes equivalência (Dougher et al., 1994; Sidman, 1987; como exemplo de uso de estímulos nulos). Assim, em cada tentativa foram apresentados três estímulos de comparação, um da Classe 1, um da Classe 2, e um nulo (Classe 3).

**Figura 1**

*Representação esquemática das imagens para cada classe*

	A	B	C	D
Classe 1				
Classe 2				
Classe 3				

Foram utilizados dois tipos de consequência neste experimento. A consequência para acerto, presumivelmente reforçadora (SR+), foi a apresentação da palavra escrita “Correto!” na cor verde, por 1 s. As consequências para erros, presumivelmente punitivas (Pun+), foram a apresentação de um “X” na tela do computador (Experimento 1) ou da palavra escrita “Errado!” (Experimento 2) em vermelho e no centro da tela, simultaneamente à apresentação de um som potencialmente aversivo, por 1 s (consequência composta). O Intervalo entre tentativas (IET) foi sinalizado pela apresentação de uma tela preta sem qualquer outro *feedback*. O som aversivo, nomeado como “Buzzer” (som nº712), foi selecionado do *International Affective Digital Sounds* (Bradley & Lang, 2007). O IET quando apresentado após SR+ ou Pun+ teve duração de 500 ms, mas quando apresentado após respostas para as quais não havia consequências programadas, teve duração de 1500 ms para manter a mesma duração entre uma resposta e o início da próxima tentativa (ver Tabela 1). Para todos os participantes, foram utilizados o mesmo computador e o mesmo volume do som. Além disso, durante o experimento, o teclado era retirado da sala para que o participante não tivesse controle sobre o volume do estímulo auditivo.

Os participantes foram distribuídos em três grupos de acordo com as consequências utilizadas: Grupo Reforço (GR), Grupo Reforço-Punição (GR-P) e Grupo Punição (GP). Para o Grupo Reforço, respostas corretas de seleção do estímulo de comparação foram seguidas por SR+/IET e respostas incorretas apenas pelo IET. Para o Grupo Reforço-Punição, respostas corretas foram seguidas por SR+/IET e respostas incorretas por Pun+/IET. Para o Grupo Punição, respostas corretas foram seguidas apenas pelo IET e respostas incorretas por Pun+/IET (ver Tabela 1).

**Tabela 1**

*Representação esquemática das contingências de treino para os três grupos.*

	Resposta	Consequência
Grupo Reforço	Acerto	SR+ (1s) e IET (0,5s)
	Erro	IET (1,5s)
Grupo Reforço-Punição	Acerto	SR+ (1s) e IET (0,5s)
	Erro	Pun+ (1s) e IET (0,5s)
Grupo Punição	Acerto	IET (1,5s)
	Erro	Pun+ (1s) e IET (0,5s)

**Diferencial Semântico (Experimento 1 e 2).** Foi utilizado um instrumento de Diferencial Semântico em ambos os experimentos (Almeida et al., 2014). O instrumento é composto por 13 escalas Likert ancoradas por adjetivos opostos: Alegre/Triste, Relaxado/Tenso, Liso/Áspero; Bonito/Feio; Leve/Pesado; Macio/Duro; Bom/Mau; Agradável/Desagradável; Rápido/Lento; Ativo/Passivo; Rico/Pobre; Dominante/Submisso.

**IRAP (Experimento 1).** O outro instrumento de medida utilizado no Experimento 1 foi o *Implicit Relational Assessment Procedure* (IRAP; cf. Barnes-Holmes et al., 2018). O IRAP é frequentemente utilizado como instrumento psicométrico para avaliar atitudes “implícitas”, particularmente as relacionadas com preconceitos. Neste caso, o pesquisador parte do que pressupõe ser a atitude que prevalece no grupo estudado. Por exemplo, Rabelo, Bortoloti e

Souza (2014) aplicaram o IRAP com crianças pressupondo que elas consideram que bonecas são brinquedos de meninas e que carrinhos são brinquedos de meninos. Cada tentativa do IRAP apresenta um modelo (neste exemplo, nome de menina ou menino), um “alvo” (figura de boneca ou carrinho) e duas opções de resposta (neste exemplo, “combina” ou “não combina”). O instrumento compara a latência do responder em blocos “consistentes”, em que o participante tem que responder de acordo com o que se presume ser a atitude do grupo estudado (menina-boneca -> combina; menina-carrinho -> não combina; menino-boneca -> não combina; menino-carrinho-> combina) com a latência em blocos “inconsistentes”, em que o participante tem que responder de modo contrário (menina-boneca -> não combina, etc.). As latências médias dos blocos consistentes são subtraídas das latências dos blocos inconsistentes, obtendo-se o D-IRAP. Se o valor do D-IRAP é positivo, confirma-se a atitude presumida<sup>3</sup>.

No presente estudo o IRAP foi usado para avaliar o grau de relacionamento entre estímulos abstratos equivalentes entre si, como feito por Bortoloti e de Rose (2012). Assim, neste como naquele estudo, depois de B1 ter se tornado equivalente a D1, e B2 equivalente a D2, as tentativas consistentes do IRAP foram: B1-D1 -> verdadeiro; B1-D2 -> falso; B2-D1 -> falso; e B2-D2 -> verdadeiro. Nas tentativas inconsistentes, as respostas foram revertidas. Os valores de D-IRAP foram usados para estimar o grau de relacionamento entre os estímulos.

**FAST (Experimento 2).** O FAST foi utilizado no Experimento 2 no lugar do IRAP pois os participantes relataram muita dificuldade com o IRAP e poucos participantes atingiram o critério para se utilizar os dados na análise. O FAST é um instrumento computadorizado que tem como objetivo ensinar classes de estímulos funcionais (O’Reilly et al., 2012, 2013). Os participantes realizam um bloco de treino (16 tentativas) e dois de teste (50 tentativas cada). Em cada tentativa é apresentado no centro da tela um de quatro estímulos (A1, A2, D1 e D2)

---

<sup>3</sup> Esta descrição do IRAP é bastante sucinta e deixa de lado várias complexidades do instrumento e das medidas obtidas, mas é suficiente para os propósitos deste texto.

de forma aleatória, e é esperado que o participante clique na tecla “Z” ou “M” diante dos estímulos. No bloco consistente, se ele apertar “Z” para os estímulos A1 e D1, ou apertar “M” para os estímulos A2 e D2, será apresentada a palavra “CORRETO” escrita em vermelho no centro da tela e passará para a próxima tentativa. Caso o participante aperte “Z” para os estímulos A2 e D2 ou “M” para os estímulos A1 e D1, será apresentada a palavra “ERRADO” escrita em vermelho e passará para a próxima tentativa. No bloco inconsistente se os participantes apertarem “Z” para A1 e D2 e “M” para A2 e D1 é apresentado o *feedback* de acerto e, se fizerem o contrário, é apresentado o *feedback* de erro. Cada tentativa tinha um limite de 3000 ms, a partir da apresentação do estímulo no centro da tela; caso o participante não respondesse nesse intervalo de tempo, era apresentado o *feedback* de erro seguido pela próxima tentativa. A ordem dos blocos consistentes e inconsistentes foi feita de forma semi-aleatória entre os participantes, e o bloco de treino foi realizado com palavras aleatórias, fora do contexto da pesquisa (e.g., MAÇÃ, BRAÇO, VACA e CALÇA). O Fast registra os acertos dos participantes em cada um dos blocos (consistente e inconsistente) e a taxa de aprendizagem em cada um dos blocos. É esperado que a taxa de aprendizagem seja maior nos blocos consistentes do que nos blocos inconsistentes.

**Sorting (Experimento 2).** O Sorting foi utilizado no Experimento 2 para testar outras relações emergentes além das testadas no teste de equivalência. O procedimento do Sorting consiste na apresentação de um baralho de cartas, cada uma contendo um estímulo dos membros das classes de equivalência esperadas, a partir do ensino da linha de base de discriminações condicionais ensinadas. O baralho é entregue ao participante, com as cartas embaralhadas de forma aleatória, e é apresentada uma instrução para ele dividir as cartas em grupos. Essa é uma forma alternativa de se testar a formação de classes de equivalência (Arntzen et al., 2017), mais rápida e de fácil aplicação em comparação aos testes computadorizados de *Matching-to-Sample* (Arntzen et al., 2015).

### ***Procedimento***

**Fase 1. Estabelecimento das classes de equivalência.** O procedimento foi semelhante para todos os participantes, com exceção das consequências para respostas corretas e incorretas, que variaram para os diferentes grupos (ver Tabela 1) e para os dois experimentos (ver Tabela 3). O participante sentava-se em frente ao computador, que apresentava na tela as instruções sobre a tarefa experimental:

“Nesta sessão você vai aprender a relacionar figuras. Primeiro você deve clicar com o *mouse* na figura que será apresentada no centro da tela. Depois que você clicar na figura central, outras três figuras serão apresentadas na parte de baixo da tela. Você deverá escolher uma dessas figuras, clicando nela com o *mouse*. É muito importante que você preste atenção às figuras que você estiver relacionando. Obrigado pela sua atenção e bom trabalho! Aperte ‘espaço’ para continuar”.

O experimentador dizia ao participante: “Essa é uma tarefa computadorizada. Eu não vou dar nenhuma instrução, vou só pedir que você leia o que está escrito na tela do computador e preste muita atenção no que está fazendo. Em alguns momentos vai aparecer uma tela escrito 'clique na tela para continuar'. Esta é uma tela de pausa: caso você queira ir ao banheiro, beber água, fique à vontade. No final, vai aparecer uma tela falando para me chamar; basta abrir a porta que eu estarei ali fora". O experimentador perguntava se o participante tinha alguma dúvida e, caso não tivesse, ele saía, fechava a porta e o experimento iniciava. A sequência do procedimento é apresentada na Tabela 2.

**Tabela 2**

*Sequência de treino, número de tentativas em cada bloco e critério para finalização do bloco, do Experimento 1 e Experimento 2 (informações sublinhadas).*

<b>Procedimento</b>	<b>Nº tentativas por bloco (máx. 5 blocos)</b>	<b>Critério</b>
Treino AB	12	100%
Treino AC	12	100%
Treino CD	12	100%
Treino Misto AB/AC/CD	18	100% ou <u>94%</u>
Teste Misto AB/AC/CD	18	100% ou <u>94%</u>
Teste BD/DB	24	91,70%

O procedimento começava com o ensino da relação AB, em um procedimento de DMTS 2s. Todas as tentativas iniciavam com a apresentação de um estímulo modelo no centro da tela (A1 ou A2). Um clique com o *mouse* sobre o modelo era seguido pelo desaparecimento deste e pela apresentação, 2 s depois, de três estímulos de comparação, B1, B2 e B3. As seleções designadas como corretas foram B1 para o modelo A1 e B2 para o modelo A2. A seleção dos estímulos da Classe 3 nunca era seguida pela apresentação da consequência de acerto. As respostas do participante diante do estímulo de comparação eram seguidas pela remoção de todos os estímulos e pelas consequências programadas para o grupo do qual ele fazia parte. O treino AB era composto por um bloco de 12 tentativas, seis com A1 e seis com A2 como modelo (os estímulos da Classe 3 apareciam somente como comparações). Para o Experimento 1, o critério deste e de todos os blocos, com exceção do teste BD/DB foi de 100% de acertos ou a apresentação do bloco por até o máximo de cinco vezes. Para o Experimento 2, foi alterado o critério do Treino Misto e do Teste Misto para 94,4% de acerto (um erro). Caso o participante não atingisse o critério em cinco blocos, ele passava para o próximo bloco de ensino. O procedimento foi programado dessa forma para que os

participantes não repetissem muitas vezes as mesmas condições de ensino e para que o procedimento não fosse muito aversivo para aqueles que estavam entrando em contato com Pun+ (estímulo potencialmente aversivo usado como consequência para respostas incorretas). A relação AC foi ensinada em seguida, com procedimento semelhante, e era seguida pelo ensino da relação CD, também com procedimento semelhante.

Após o ensino das relações AB, AC e CD, era conduzido um bloco de tentativas de linha de base cumulativa (Treino Misto), com apresentação, em sequência randomizada, de tentativas das relações AB, AC e CD. Este bloco teve um total de 18 tentativas, sendo três de cada uma das relações AB, AC e CD. Em seguida, este mesmo bloco era apresentado sem consequências diferenciais (Teste Misto), em preparação para os testes de formação de classes. Antes do início deste bloco, o computador apresentava a seguinte mensagem: “A partir de agora o computador não vai mais sinalizar se suas escolhas estão corretas ou erradas, mas continuará registrando suas respostas”.

Após atingido o critério na linha de base cumulativa sem consequências diferenciais, era conduzido o teste de relações emergentes que tinha como objetivo documentar a formação das classes de estímulos equivalentes. Para isto, eram testadas, na ausência de consequências diferenciais, as relações BD e DB. Este é um tipo de teste “abreviado” de equivalência: considerando a estrutura das relações ensinadas, AB, AC e CD, a emergência das relações BD e DB comprova logicamente a simetria e transitividade<sup>4</sup> de todas as relações treinadas (Bortoloti & de Rose, 2009, 2012). O bloco tinha um total de 24 tentativas, sendo seis para cada uma das relações (B1D1, D1BD1, B2D2 e D2B2). O critério para finalização do bloco era pelo menos 91,6% de acertos (máximo de dois erros) ou a repetição de no máximo cinco

---

<sup>4</sup> Embora a reflexividade seja também uma propriedade definidora da equivalência, e Sidman (Sidman, 1994) tenha insistido na necessidade de que seja também testada, os testes de reflexividade foram gradualmente desaparecendo da pesquisa sobre equivalência, inclusive por algumas ambiguidades na interpretação dos seus resultados, como apontado por Saunders e Green (1992).



vezes. As fases seguintes ao estabelecimento das classes de equivalência foram diferentes entre os experimentos (ver Tabela 3).

**Tabela 3**

*Sequência das Fases do Experimento 1 e do Experimento 2.*

	<b>Experimento 1</b>	<b>Experimento 2</b>
<b>Fase 1</b>	Estabelecimento das classes de equivalência	Estabelecimento das classes de equivalência
<b>Fase 2</b>	Diferencial Semântico	Diferencial Semântico
<b>Fase 3</b>	IRAP	FAST
<b>Fase 4</b>	X	Sorting
<b>Fase 5</b>	Teste de Manutenção	Teste de Manutenção

**Fase 2. Preenchimento do Diferencial Semântico.** Todos os participantes avaliaram os estímulos D utilizando uma folha impressa para cada estímulo, com a figura a ser avaliada disposta na parte central superior, e as 13 escalas Likert dispostas abaixo. Eles receberam uma folha com instruções para o preenchimento do instrumento. Após a leitura das instruções e eventuais esclarecimentos de dúvidas, o participante recebia três folhas, cada uma contendo um dos estímulos D1, D2 e D3, em uma sequência aleatória, para preencher de acordo com as instruções. Esperava-se que os escores de diferencial semântico mostrassem alterações na valência de D1 e de D2. Entretanto, não foram coletados dados o suficiente para se analisar o diferencial semântico.

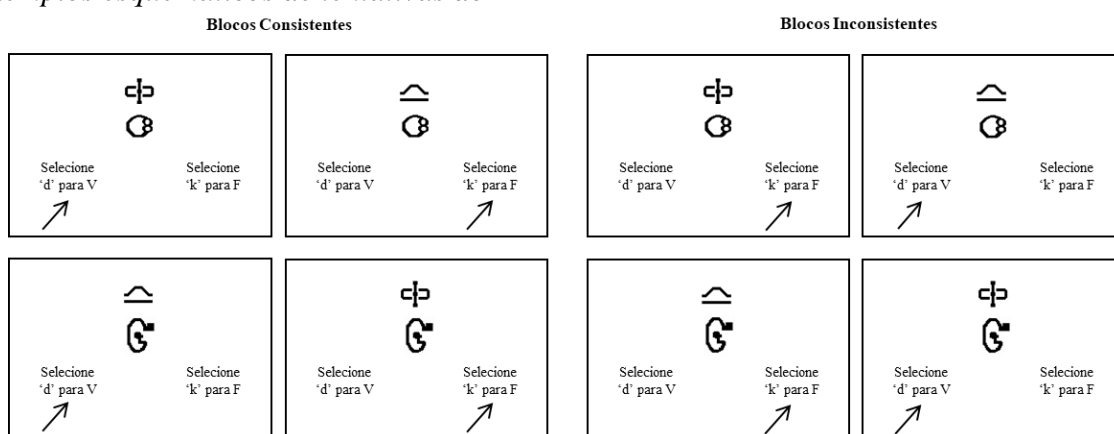
**Fase 3 (Experimento 1). IRAP.** A avaliação com o IRAP foi similar à utilizada por Bortoloti e de Rose (2012), como descrito acima, na seção *Equipamentos e Instrumentos*.

Os participantes realizaram blocos de treino e de teste, cada bloco com 32 tentativas. Em cada tentativa do IRAP eram apresentados, simultaneamente, um estímulo do conjunto A (A1 ou A2) no topo superior da tela (“modelo”), um estímulo do conjunto D (D1 ou D2) no centro

da tela (“alvo”) e duas opções de respostas (tecla “d” para verdadeiro e tecla “k” para falso) nos cantos inferiores esquerdo e direito, respectivamente. A sequência dos modelos e alvos e a posição das opções de respostas eram determinadas randomicamente pelo *software* do IRAP.

Ao longo da tarefa, os participantes eram submetidos a blocos alternados formados por tentativas ora consistentes com as classes formadas, ora inconsistentes. Para obter consequência de acertos, nos blocos consistentes os participantes deveriam escolher “d” (referente a “verdadeiro”) para D1 diante de A1 e para D2 diante de A2, e escolher “k” (referente a “falso”) para D1 diante de A2 e de D2 diante de A1; nos blocos inconsistentes os participantes deveriam escolher “d” para D2 diante de A1 e para D1 diante de A2, e escolher “k” para D1 diante de A1 e de D2 diante de A2 (ver Figura 2). Respostas corretas eram consequenciadas com a retirada de todos os estímulos e um intervalo entre tentativas de 400 ms. Caso o participante emitisse uma resposta incorreta, um “X” vermelho era apresentado logo abaixo do estímulo alvo. O intervalo entre tentativas iniciava apenas quando o participante selecionasse a resposta correta.

**Figura 2**  
*Exemplos esquemáticos de tentativas do IRAP*



Os blocos de treino no IRAP eram apresentados até que fossem emitidas no mínimo 80% de respostas corretas nos blocos consistente/inconsistente. Após atingir o critério o

participante era exposto a outro bloco consistente/inconsistente com adição do critério de latência de resposta de no máximo 2000 ms. Caso o participante não acertasse, os pares de blocos de treino eram repetidos no máximo três vezes. Participantes que atingiram o critério foram expostos aos blocos de teste, que consistiram em três pares de blocos consistente/inconsistente. Nestes blocos não havia critério.

**Fase 3 (Experimento 2). FAST.** O experimentador apresentou as seguintes instruções aos participantes, antes do experimento iniciar: “É um programa bem simples. Não aparecer palavras ou imagens na tela, uma palavra ou imagem de cada vez, e seu objetivo é descobrir qual tecla você precisa apertar diante de cada uma delas, se é a letra ‘Z’ ou a letra ‘M’. Não tem nada a ver com as letras em si, é só para ficar uma na mão esquerda e uma na mão direita. Você não vai receber dica prévia sobre qual tecla deve apertar, mas conforme você for acertando ou errando, vai descobrindo. O seu objetivo é acertar o máximo possível, mas não se preocupe se você estiver errando. Tente trabalhar de maneira rápida, pois se você demorar muito para responder o programa irá entender como erro. Nesse caso, aparecerá o *feedback* de “errado” e irá para a próxima palavra. Serão três blocos. O primeiro é somente para você se familiarizar com a tarefa, então se tiver alguma dúvida pode perguntar. Eu vou pedir para você ler o texto e quando quiser, pode apertar alguma tecla para iniciar.”

O computador apresentava a seguinte mensagem: “Na seção seguinte, sua tarefa é aprender qual botão pressionar quando uma imagem aparecer na tela. **IMPORTANTE:** Durante essa fase você deve pressionar somente as letras Z e M. Por favor, localize-as no teclado agora. Essa seção inicial é dada para que você possa praticar a tarefa. Para ajudar você a aprender, o programa irá indicar se você está acertando ou errando. Se você tiver alguma pergunta, por favor, pergunte a(o) pesquisador(a) agora. Pressione qualquer tecla”.

O experimentador ficava na mesma sala que o participante durante o bloco de treino. Ao fim deste, o experimentador perguntava se o participante tinha alguma dúvida. Caso não tivesse, o experimentador deixava o participante sozinho para finalizar essa fase.

**Fase 4 (Experimento 2). *Sorting*.** Após a Fase 3, os participantes do Experimento 2 recebiam um bloco de cartas. O experimentador dava as seguintes instruções: “Eu vou te entregar esse bloco de cartas e pedir para que você as divida em grupos, assim como você acha que elas devem ser divididas” e saía da sala, deixando o participante sozinho.

**Fase 5. Teste de manutenção.** Ao fim do experimento, os participantes foram dispensados e informados de que deveriam retornar entre 30 e 35 dias depois. No retorno, eles foram submetidos novamente ao teste de equivalência BD/DB, fizeram nova avaliação dos estímulos D com o Diferencial Semântico e os participantes do Experimento 2 repetiram o *Sorting*. O IRAP e o FAST não foram repetidos, pois poucos participantes conseguiram chegar aos blocos de teste do IRAP e os dados do FAST foram inconclusivos. Todos os participantes retornaram, e o tempo médio de retorno foi de 35,8 dias, período mais longo do que o planejado devido a problemas na remarcação do dia da pesquisa um mês depois.

## **Resultados**

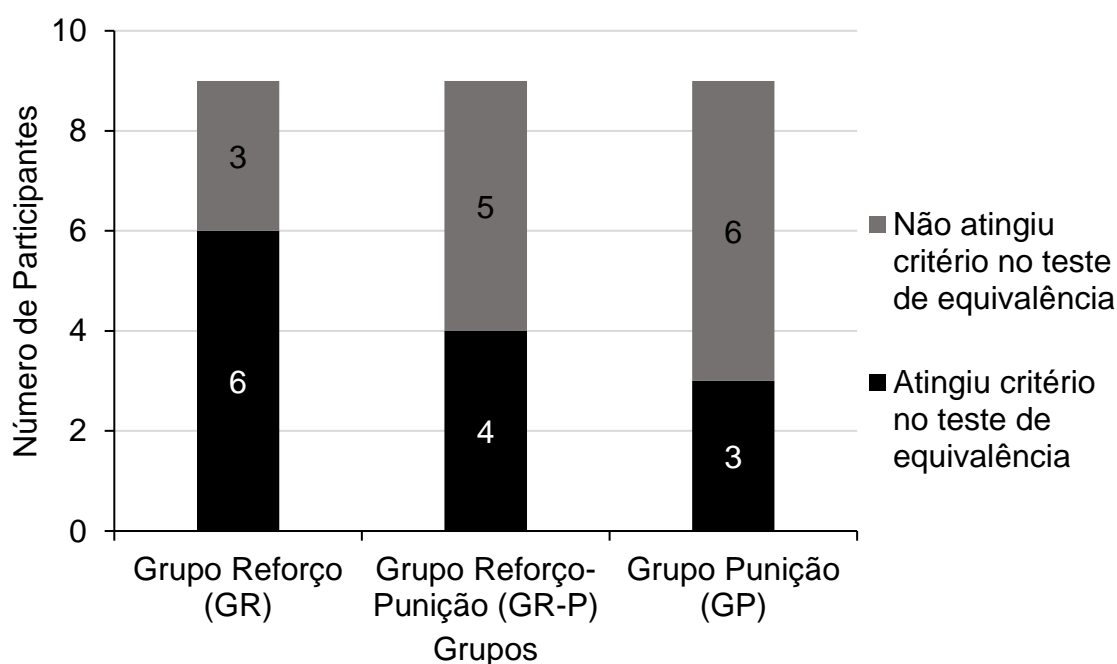
Os resultados gerais do Experimento 1 e 2 serão apresentados de forma conjunta e detalhes mais específicos de cada experimento serão apresentados individualmente. Não serão apresentados os dados relativos ao *Sorting*, ao IRAP e ao FAST pois, considerando o número pequeno de participantes, não foi possível realizar uma análise adequada pela quantidade de dados disponíveis. As limitações desses procedimentos nos Experimento 1 e 2 foram comentadas na discussão.

Foram considerados os dados de todos os participantes, independente do critério ter sido atingido nas fases de treino ou de teste, para se observar o efeito das consequências nos

desempenhos de aprendizagem das relações. Considerando os Experimentos 1 e 2, o número de participantes dos grupos Reforço, Reforço-Punição e Punição que atingiram o critério no Teste de Equivalência foram: 6, 4 e 3, respectivamente (ver Figura 3)

### Figura 3

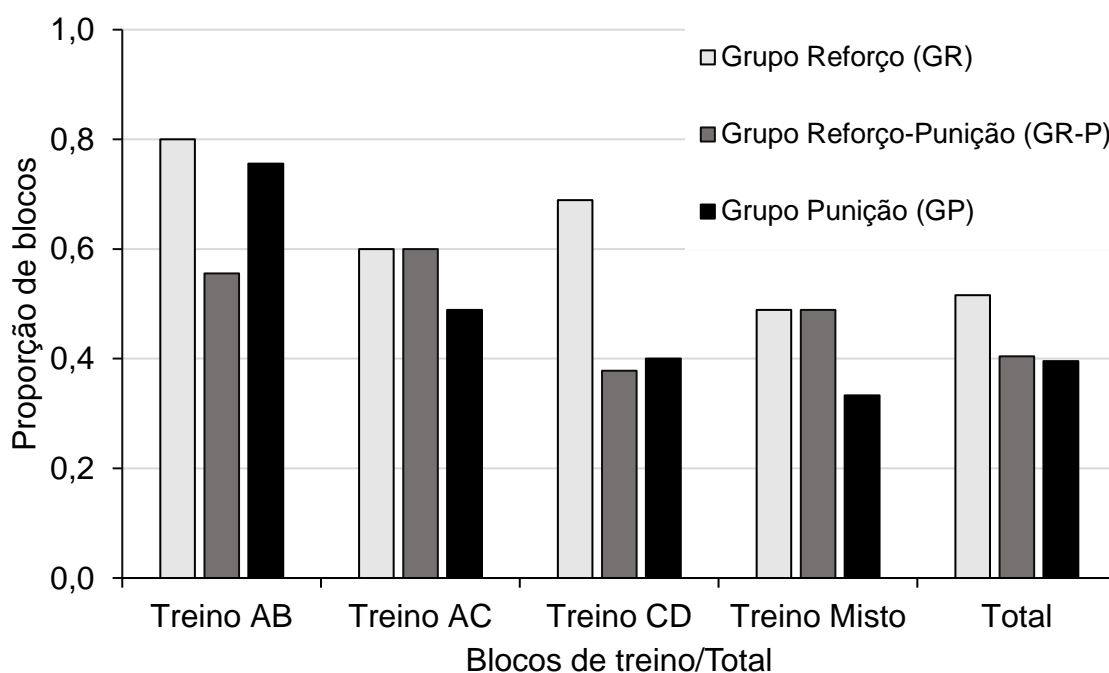
*Número de participantes que atingiram (barras pretas) e dos que não atingiram (barras cinzas) critério no Teste de Equivalência considerando os participantes dos Experimentos 1 e 2.*



Os dados dos treinos demonstram pequenas diferenças entre os grupos. O número total de blocos necessários para que o participante chegasse ao Teste Misto foi maior para o GR em comparação ao GR-P e ao GP. A Figura 4 apresenta a média da proporção de blocos (número total de blocos que os participantes precisaram realizar até atingir o critério/número total de blocos possíveis) necessários para que os participantes de cada grupo atingissem o critério e a média da proporção total dos treinos. Nesta figura é possível observar que no geral, o Grupo Punição teve uma redução do número de blocos necessários para se atingir o critério ao longo de todos os blocos de treino, o que não ocorreu para os outros dois grupos.

**Figura 4**

*Proporção de blocos (Nº de blocos para atingir critério/Nº total de blocos que poderiam ser realizados) necessários para atingir o critério em cada um dos blocos de treino para todos os grupos. E a proporção total dos blocos de treino, considerando os participantes do Experimento 1 e 2.*

**Experimento 1**

Dois participantes do Grupo Reforço, três participantes do Grupo Reforço-Punição e quatro participantes do Grupo Punição conseguiram atingir o critério de 100% de acertos nas fases de treino e de teste das relações ensinadas (ver células realçadas em cinza na Tabela 3). No bloco de teste das relações de equivalência (BD/DB), três participantes do GR e do GR-P e dois do GP atingiram critério. No teste de manutenção, 35 dias depois, apenas uma pessoa do GR manteve as relações BD/DB (ver Tabela 4). A porcentagem de acertos dos participantes nos blocos do Teste BD/DB mostra que um participante do GR (107) chegou

próximo do critério para finalização do bloco, mas errou consistentemente nos blocos, não conseguindo atingir o critério.

**Tabela 4**

*Porcentagem de acertos dos participantes dos três grupos em todos os blocos da Fase 1 do Experimento 1.*

	AB	AC	CD	ABACCD	ABACCD (ext)	Teste BD/DB	Manutenção BD/DB
<b>Grupo Reforço</b>							
102	83 83 92 83 100	92 75 100	75 67 92 92 100	94 100	94 89 100	33 67 92	21 38 50 50 42
103	50 100	67 75 100	83 83 100	67 100	94 94 100	75 96	92
105	0 58 83 67 92	67 100	58 75 92 100	94 89 100	100	33 58 88 71 38	21 25 17 0 13
106	42 67 67 92 100	83 92 92 100	67 100	94 94 94 94 89	100	58 83 83 92	33 33 25 42 25
107	0 8 75 100	0 58 50 92 83	58 92 75 92 83	89 83 89 89 89	89 89 100	75 88 79 88 75	29 29 29 38 38
<b>Grupo Reforço-Punição</b>							
201	58 92 92 100	92 100	92 100	94 100	100	96	13 0 4 0 0
202	75 100	92 100	58 92 100	94 89 94 94 100	94 94 89 100	92	58 75 58 54 46
203	83 100	92 92 100	83 100	100	100	100	17 0 0 0 0
204	42 42 58 25 33	67 92 83 92 100	92 100	72 83 100	100	29 25 29 38 58	63 58 75 100
205	75 100	58 92 100	100	83 89 89 94 89	89 83 94 100	4 0 0 0 0	100
<b>Grupo Punição</b>							
301	75 100	92 100	100	100	100	88 100	8 29 4 0 71
302	83 83 92 100	83 100	83 100	89 100	100	67 58 46 83 33	33 33 29 29 25
303	75 92 100	75 100	67 100	94 100	100	33 96	50 50 50 50 50
304	33 8 25 0 92	67 92 100	100	83 100	100	13 0 0 0 0	0 0 4 4 0
305	92 100	92 92 100	83 92 100	89 100	100	0 0 0 0 0	4 0 0 0 0

*Nota.* Os valores marcados em cinza indicam o bloco no qual o participante atingiu o critério de número de acertos para finalizar o bloco.



## ***Experimento 2***

Com a exceção de dois participantes, todos os outros conseguiram atingir o critério nos blocos de treino (Treino AB, Treino AC, Treino CD e Treino Misto) e de teste das relações de linha de base (Teste Misto). As exceções foram no treino AB para o Grupo Reforço, no qual os participantes 111 e 112 não atingiram o critério (ver células realçadas em cinza na Tabela 5). Nos blocos de teste das relações emergentes (Teste BD/DB), três participantes do GR, um participante do GR-P e um do GP atingiram critério. Entretanto, no Sorting, um participante do Grupo Reforço, um do Grupo Reforço-Punição e três do Grupo Punição atingiram critério. Apenas um participante, do GR, manteve as relações no Teste de Manutenção (BD/DB).

**Tabela 5.**

*Porcentagem de acertos dos participantes dos três grupos em todos os blocos da Fase 1 do Experimento 2.*

	AB	AC	CD	ABACCD	ABACCD (ext)	Teste BD/DB	Sort.	Manutenção BD/DB	Sort.
<b>Grupo Reforço</b>									
109	33 92 100	58 92 92 100	58 100	100	100	21 50 79 92		13 21 17 17 33	
110	50 100	58 100	58 100	89 100	100	25 33 25 25 25		67 50 92	
111	0 0 0 8 0	92 83 83 92 100	67 100	78 83 94	100	79 88 92	X	75 58 96	
112	50 50 50 50 83	50 100	42 92 92 100	100	94	33 50 79 92		46 33 50 54 46	
<b>Grupo Reforço-Punição</b>									
206	58 100	67 100	92 100	83 94	100	8 8 0 0 0		0 0 0 0 0	
207	33 75 83 100	67 92 100	58 100	94	100	50 50 46 42 50		71 88 96	
208	50 100	67 92 92 92 100	83 100	100	100	88 100	X	88 92	
209	58 100	83 100	100	83 100	100	13 42 8 0 13		17 4 0 0 0	
<b>Grupo Punição</b>									
306	58 67 92 92 100	42 75 100	83 92 100	83 100	100	54 21 0 0 0	X	83 96	
307	42 83 83 100	83 100	92 100	83 94	100	0 0 0 0 0		25 92	
308	25 83 92 92 100	58 92 100	75 100	100	100	42 4 4 0 0	X	100	
309	67 92 100	83 100	92 100	100	100	96	X	25 17 0 0 4	

*Nota.* Os valores marcados em cinza indicam o bloco no qual o participante atingiu o critério de número de acertos para mudar de bloco. No caso do Sorting, as células marcadas em cinza indicam os participantes que agruparam os estímulos das classes de maneira consistente com as classes planejadas pelo experimentador.

## Discussão

Em todos os grupos ocorreu a formação das classes de equivalência para alguns dos participantes, apesar das diferenças no desempenho dos participantes no teste das relações emergentes (Teste BD/DB). Embora não tenha sido utilizado teste estatístico, um maior número de participantes (6) do Grupo Reforço atingiu critério neste teste em comparação aos outros dois grupos (4 e 3). Além disso, no Experimento 1, um dos participantes do GR que não atingiram critério ficaram com o desempenho próximo de 90% de acertos, o que indica que as relações provavelmente foram estabelecidas.

Um dado que se contrapõe aos resultados do Teste BD/DB é o do Sorting. No Experimento 2 foi utilizado esse procedimento para se testar outras relações emergentes que não estavam sendo testadas no Experimento 1, entretanto os resultados desse teste foram, de certa forma incongruentes com os resultados do Teste BD/DB. Apenas um dos participantes do GR e um do GR-P que atingiram critério no primeiro teste também atingiram critério no Sorting. E no GP, dois participantes que não haviam demonstrado as relações de equivalência, apresentaram as classes corretamente no Sorting.

Essa discrepância entre os dados também pode ser visualizada no Teste de Manutenção em ambos os experimentos. No Experimento 1, dois participantes do GR-P, que não haviam formado as classes, atingiram critério no teste de manutenção. E no Experimento 2 isso aconteceu com pelo menos um participante em todos os grupos. Uma hipótese para esses resultados dissonantes é o controle fraco das relações ensinadas, o que pode ter gerado desempenhos ambíguos, como apresentado anteriormente, além de aumentar a probabilidade de ocorrência de falsos-positivos.

Em um estudo sobre aprendizagem de classes por seleção condicional não reforçada (Saunders et al., 1988), os autores testaram se participantes com histórico de aprendizagem de classes de estímulos equivalentes responderiam de forma condicional diante de novas

relações que não fossem reforçadas. Os participantes não só apresentaram respostas de seleção condicional diante de estímulos que já pertenciam a uma classe pré-estabelecidas, como responderam de maneira condicional diante de estímulos novos, formando classes sem que a resposta de seleção tivesse sido reforçada. Variações deste estudo reafirmaram a possibilidade de os participantes apresentarem um comportamento generalizado de seleção condicional quando expostos a esse tipo de tarefa (Harrison & Green, 1990; Saunders et al., 1988; Williams et al., 1995).

Considerando estes estudos, no caso do Experimento 1 e 2, alguns participantes que atingiram o critério de no mínimo 92% de acertos podem não ter formado as classes de equivalência, ou seja, podem ter respondido de forma generalizada de acordo com o que aprenderam nas fases de treino. Um dado que complementa essa hipótese é a quantidade de participantes que ficaram com 0% de acertos no Teste de BD/DB em mais de um bloco de teste (205, 206, 304, 305, 306, 307, 308). Nestes casos, é provável que os participantes tenham aprendido que deveriam selecionar os estímulos de uma maneira consistente, o que provavelmente levou a um desempenho estável, apesar de estar incondizente com as relações aprendidas. Da mesma forma, esse processo pode ter ocorrido com os participantes que atingiram o critério no Teste de Equivalência: mesmo tendo atingido no mínimo 92% de acertos, eles podem ter generalizado o responder dos treinos, selecionando sistematicamente no teste apenas um dos estímulos que, neste caso, foi condizente com o ensinado. Uma forma de se diminuir a probabilidade desse tipo de falso-positivo é considerar como critério de formação das classes de equivalência dois testes consecutivos com no mínimo 92% de acertos (tanto para o Teste BD/DB quanto para o Teste de Manutenção). Desta forma, seria menos provável que os participantes, que atingiram o critério de acertos, tivessem apresentado um desempenho baseado unicamente na aprendizagem de responder de forma condicional generalizada.

Uma outra hipótese que se complementa à anterior é que o ensino das relações de apenas duas classes de equivalência prejudicou a aprendizagem. Mesmo tendo sido inserida uma terceira classe, que funcionava como estímulo distrator nas tentativas de treino e teste, as contingências implicavam na discriminação de apenas duas classes. Desta forma, é possível compreender melhor o desempenho de alguns participantes que tiveram uma emergência atrasada no Teste de Manutenção (Participante 204 – Experimento 1), ou então dos que passaram de desempenhos entre 0 e 30% de escolhas corretas no Teste BD/DB para 100% de acertos no Teste de Manutenção (Participantes 205, 110, 306, 307 e 308). Nestes casos é possível que o padrão de respostas dos participantes tenha invertido de um teste para o outro (ou então os participantes tenham passado a responder por exclusão), o que ocasionou a melhora do desempenho entre os testes, sem que as relações tenham sido retreinadas.

Além disso, algumas variáveis podem ter contribuído, pelo menos parcialmente, para esses resultados: a utilização exclusiva de estímulos abstratos, a estrutura de treino mista OTM/Linear, o uso do protocolo simultâneo, e o teste apenas das relações emergentes entre os estímulos B e D (relação de equivalência). Todas essas variáveis diminuem a probabilidade da formação das classes de equivalência (Arntzen, 2012) e, somadas às dificuldades no controle experimental sinalizadas anteriormente, provavelmente contribuíram para que poucos participantes atingissem critério nos testes realizados. Desta forma, utilizar procedimentos mais eficazes no ensino das relações pode não só contribuir com a melhora no desempenho geral dos participantes nas fases de treino, como também diminuir a probabilidade de resultados ambíguos nos testes das relações emergentes.

Outros testes que foram afetados pelo controle fraco das relações ensinadas foram o IRAP e o FAST. Os estudos considerados para a utilização desses procedimentos ensinavam as relações de equivalência utilizando imagens com significado prévio (Bortoloti & de Rose, 2012) ou então ensinando relações de classes com poucos estímulos abstratos (O'Reilly et al.,

2012).. Desta forma os procedimentos utilizados nessas pesquisas provavelmente ensinavam as relações de uma maneira mais eficaz do que o procedimento utilizando nos Experimentos 1 e 2. Além disso, em um estudo sobre a relação entre quantidade de treino e desempenho no FAST, Cummins et al (2018) verificaram que o valor do FAST era correlacionado positivamente à quantidade de treino: quanto mais treino os participantes de um grupo recebiam, melhor era o desempenho deste grupo no FAST. Ou seja, melhorar o procedimento de ensino das relações condicionais não só pode diminuir o controle por variáveis externas, como pode permitir um uso mais adequado do IRAP ou do FAST.

É importante destacar, entretanto, que o uso de um procedimento pouco eficaz no ensino das relações e a possibilidade de falso-positivos nos testes de equivalência não explicam totalmente o desempenho dos participantes do Grupo Punição, em que um menor número de participantes formou classes, em comparação ao Grupo Reforço (quando considerado o Teste BD/DB de ambos os experimentos). Essas diferenças sinalizam que o uso de consequências punitivas pode ter um efeito prejudicial na aprendizagem de relações condicionais, mesmo que o pequeno número de participantes e as diferenças entre os procedimentos não tenham permitido a realização de uma análise estatística adequada para se verificar a diferença entre os grupos com relação aos testes utilizados (incluindo o Diferencial Semântico, que não foi incluído na análise pelo tamanho da amostra).

Considerando esses resultados e as limitações dos Experimentos 1 e 2, foi planejado um novo experimento no qual as condições de ensino e de teste das relações fossem mais controladas, permitindo verificar de forma mais clara o efeito das diferentes consequências no desempenho dos participantes. Algumas mudanças tiveram o objetivo de controlar a probabilidade de ocorrência de resultados falsos-positivos, como: o ensino das relações entre estímulos de três classes (para diminuir a probabilidade de controle por rejeição) e o teste de todas as relações emergentes. E algumas das mudanças tiveram como objetivo melhorar o

desempenho dos participantes nos testes, aumentando a probabilidade de formação e manutenção das classes: redução do intervalo de tempo para realização do Teste de Manutenção, diminuição do número de estímulos por classe, e uso de um protocolo de ensino similar ao Simples para Complexo.

## Parte 2 – Gamificação do Procedimento e Experimento 3

Em março/2020 foi decretada a interrupção das atividades presenciais em diversas instituições, incluindo a UFSCar (Portaria UFSCar N°4380, de 20 de março de 2020), devido ao avanço da pandemia de Covid-19. Diante desse cenário, foram consideradas algumas opções para a continuação da coleta de dados de maneira online.

Partindo das variáveis estudadas nesta dissertação, foram identificados alguns problemas que poderiam ocorrer com a adaptação mais simples do procedimento para a coleta remota. Nos Experimento 1 e 2 uma das principais características da consequência punitiva era a utilização de um som aversivo. Isso foi controlado na pesquisa presencial utilizando o mesmo computador e o mesmo volume para todos os participantes. Além disso, o teclado era retirado da sala de coleta de dados para que o participante não pudesse aumentar ou diminuir o volume do computador. Desta forma, os participantes não tinham qualquer controle sobre a intensidade do som e, conseqüentemente, da punição.

Se a pesquisa fosse conduzida de maneira remota, apenas transmitindo a tela do computador do experimentador, os participantes teriam muito mais controle sobre a aversividade da punição em comparação aos experimentos anteriores. Eles poderiam diminuir o volume, tirar o som do computador, e mesmo que todos deixassem o mesmo nível de som, há variações na intensidade de volume entre os diferentes computadores.

Considerando que esta era uma das principais variáveis e que outros elementos da coleta de dados também seriam prejudicados pela falta de controle experimental na coleta remota, foi planejada uma adaptação do procedimento. Para esta finalidade foi planejada e construída<sup>5</sup> uma versão gamificada do procedimento de *Matching-to-Sample*, que está

---

<sup>5</sup> O desenvolvimento do jogo, realizado pelo autor, exigiu a aprendizagem de programação de jogos e programação em linguagem Python, como será detalhado mais tarde.



descrita na Parte 2.1. A Parte 2.2 abrange o Experimento 3 desta dissertação, na qual está descrita a continuação da pergunta de pesquisa adaptada ao novo procedimento utilizado.

### **Parte 2.1 – Criação do jogo “Miner Troubles”**

O *Matching-to-Sample* é comumente utilizado no estudo de equivalência de estímulos para o ensino de discriminações condicionais desde o primeiro trabalho realizado por Sidman, que ensinou uma criança atípica a ler utilizando este procedimento (Sidman, 1971). Neste estudo, foi utilizado uma máquina, que possuía uma matriz com nove janelas translúcidas que permitiam a apresentação dos estímulos e o registro da resposta dos participantes ao pressionar essas janelas. Os estímulos eram impressos em slides transparentes e o registro das respostas do participante era feito por uma caneta em um rolo de papel ligado à máquina.

Outros experimentos realizados alguns anos depois utilizaram o mesmo princípio do procedimento, mas com outros equipamentos. De Rose et al. (1996) também ensinaram leitura para crianças com dificuldade nesta tarefa e para isso, utilizaram pastas de fichário. Nelas, os estímulos de cada tentativa eram impressos e colados em folhas de papel e disponibilizados em sacos plásticos transparentes na sequência de apresentação dos treinos e dos testes. Os participantes ficavam virados de frente para a pasta e o experimentador deveria apresentar sucessivamente as tentativas, de forma manual. O registro das respostas dos participantes também era feito manualmente por pelo menos um outro experimentador, que ficava atrás da criança.

Alguns anos depois, com o avanço tecnológico e uma maior disponibilização de computadores, alguns programas foram criados para facilitar o controle dessas tarefas experimentais, como o MTS (Dube & Hiris, 1997) e o Mestre (Goyos & Almeida, 1994).

Utilizando esses *softwares*, as tarefas de programação do ensino e do registro das respostas dos participantes tornaram-se mais simples e fáceis, uma vez que não era mais necessária a presença de pelo menos dois pesquisadores durante toda a sessão experimental. Além disso esses programas computadorizados tornaram o registro das respostas dos participantes e as condições experimentais mais controladas, sem a possibilidade de erros humanos durante o procedimento.

Com o avanço das pesquisas na área e a necessidade de novos experimentos mais amplos e com controle de novas variáveis, outros *softwares* foram desenvolvidos. O programa LECH-GEIC (Capobianco et al., 2009) permitiu agrupar a programação das tentativas com a sincronização e salvamento de dados online dos participantes. Desta forma, os pesquisadores passaram a conseguir coletar dados com muitos participantes e a gerenciar as informações de uma maneira prática e segura. Outros *softwares*, como o Contingência Programada (Hanna et al., 2014) e o PsychoPy (Peirce, 2007) por sua vez, permitiram o planejamento das tentativas de treino e teste das relações condicionais, manipulando diversas variáveis e podendo ser feito sem a necessidade de se aprender uma linguagem de programação.

Entretanto, apesar das diferenças nos equipamentos utilizados, o processo básico e o *design* dos experimentos eram muito semelhantes. Em todos esses programas as tentativas iniciavam com a apresentação do estímulo modelo (visual ou auditivo). Caso o estímulo fosse visual, era geralmente exigida uma resposta de observação do participante a este estímulo, seja clicando com o mouse, pressionando com o dedo ou apontando. Após essa resposta, eram apresentados os estímulos de comparação, geralmente nos cantos ou na parte inferior da tela, da folha de papel ou da máquina. Então, caso fosse um procedimento de ensino, a resposta de seleção do participante era reforçada, se ele acertasse, ou punida/extinta, se ele escolhesse de maneira incorreta. Essas tentativas geralmente eram apresentadas em telas

monocromáticas e os estímulos utilizados eram, em sua maioria, em preto e branco. Essas tentativas eram repetidas até que se atingisse algum critério definido pelo experimentador.

Desta forma, o que foi modificado ao longo dos anos não foi o *design* das tentativas ou a forma como é implementado o procedimento. O que mudou foi o controle sobre as condições experimentais e os dados gerados por elas. Utilizando os *softwares* computadorizados, as medidas e os resultados ficaram muito mais precisos, em comparação aos primeiros estudos da área. Além disso, eles permitiram o registro de outras variáveis, como tempo de resposta, o tempo de visualização dos estímulos, o controle mais adequado do tempo entre apresentação dos estímulos modelo e estímulos comparação, uma maior diversificação dos tipos de estímulos utilizados e das consequências, entre outros. Uma das características que permitiu essas variações foi a simplicidade do procedimento. Como exemplo, os estudos de Rose et al. (1996) e de Reis et al. (2009), nos quais foi ensinada leitura e escrita para crianças com dificuldade nessa tarefa, mesmo utilizando equipamentos totalmente diferentes (pastas manejadas pelo experimentador, no primeiro; computador, no segundo), são muito semelhantes em termos do procedimento e dos resultados de aprendizagem que geraram.

Contudo, essa simplicidade tornou-se um problema durante a pandemia de Covid-19 e com a necessidade de realização de coletas de dados de forma remota. Nessa situação, os participantes muitas vezes estão em um ambiente no qual há pouco controle experimental. Além das variações de computador, conexão com a internet e habilidade com recursos tecnológicos, que já seriam um problema em uma pesquisa no laboratório, o ambiente físico está repleto de contingências competitivas. O participante pode ter que dividir o espaço com outras pessoas que podem interromper o procedimento, podem ocorrer interrupções de telefone, de interfone, de animais domésticos, barulhos externos, queda de energia, notificações do computador, entre outros.

E neste novo ambiente de coleta de dados, utilizar um procedimento computadorizado simples pode não só desmotivar os participantes, como pode distrair da tarefa de ensino, ao aumentar as operações motivacionais relacionadas a outras atividades, e assim comprometer a aprendizagem e os dados.

Uma forma de se contornar esse problema e que tem sido muito utilizada recentemente em procedimentos educacionais é a gamificação (de Sousa Borges et al., 2014). Segundo Kapp (2013), a gamificação caracteriza-se pela utilização de elementos ou *design* de jogos em outros contextos. Desta forma, criar um *Matching-to-Sample* gamificado poderia ajudar a tornar o procedimento mais imersivo, interativo e engajador, diminuindo o controle de variáveis externas ao procedimento durante as coletas remotas.

Com esse objetivo foi desenvolvido o programa Miner Troubles (MTs)<sup>6</sup>, uma versão gamificada do MTS. O software será descrito de acordo com as características experimentais e com uma taxonomia de elementos de gamificação (Toda et al., 2019).

### ***Software Miner Troubles***

O Miner Troubles foi desenvolvido utilizando o programa UPBGE (versão 0.3 alpha), um motor open-source de jogos 3D. As programações do experimento foram realizadas em linguagem Python (versão 3.9.6), que é compatível com o software da UPBGE. Modificações nas variáveis do procedimento podem ser realizadas por meio de um código fonte utilizando um leitor de *scripts* em linguagem Python. Alterações no estilo do jogo podem ser realizadas apenas pelo UPBGE. Os dados são salvos automaticamente em uma Planilha Google e podem ser salvos no computador no qual está instalado o programa.

---

<sup>6</sup> O jogo Miner Troubles foi desenvolvido pelo autor desta dissertação em parceria com o doutorando Filipe Cesar da Hora Carvalho. O jogo foi desenvolvido de forma totalmente autônoma, sem a contratação de empresas de desenvolvimento de jogos ou de programadores.

**História do jogo e Instruções.** O jogo inicia apresentando a história do personagem, os objetivos do jogador, as variáveis/métricas que o participante tem acesso ao longo do jogo e as instruções gerais, apresentadas por escrito:

“Miner é um mineiro que trabalha vendendo diamantes. Um dia, enquanto cavava, ele caiu em um buraco e ao olhar ao redor, percebeu que estava na tumba Namdis, muito antiga e famosa por ser um labirinto. Para sair dela, Miner precisa descobrir quais portas ele pode, ou não, abrir. Seu objetivo é ajudar Miner a sair da tumba e coletar o maior número possível de diamantes. O número total de diamantes está no canto superior direito. No canto superior esquerdo estão as portas corretas. Para passar de fase, você deve abrir 12 portas corretas consecutivamente. Use as setas do teclado para movimentar o Miner. Para abrir as portas, basta pressionar a barra de espaço próximo delas.”

**Ambientação do jogo.** A construção do cenário do jogo foi feita de modo que todos os estímulos ambientais, inclusive a música, passem a sensação de que o Miner está em um lugar fechado, escuro e abandonado. E a principal iluminação do jogo é a lanterna do mineiro, que possibilita o participante se focar nos estímulos individualmente (ver Figura 5). Foi utilizado esse contexto como uma forma de manter o foco na tarefa do jogo (que é aprender a relação entre os símbolos das portas) e complementar o objetivo inicial descrito para os participantes, de ajudar o Miner a sair desse local. Todas as paredes, o piso e o cenário são iguais durante as tentativas; os únicos elementos que mudam são os símbolos apresentados nas portas.

**Figura 5**

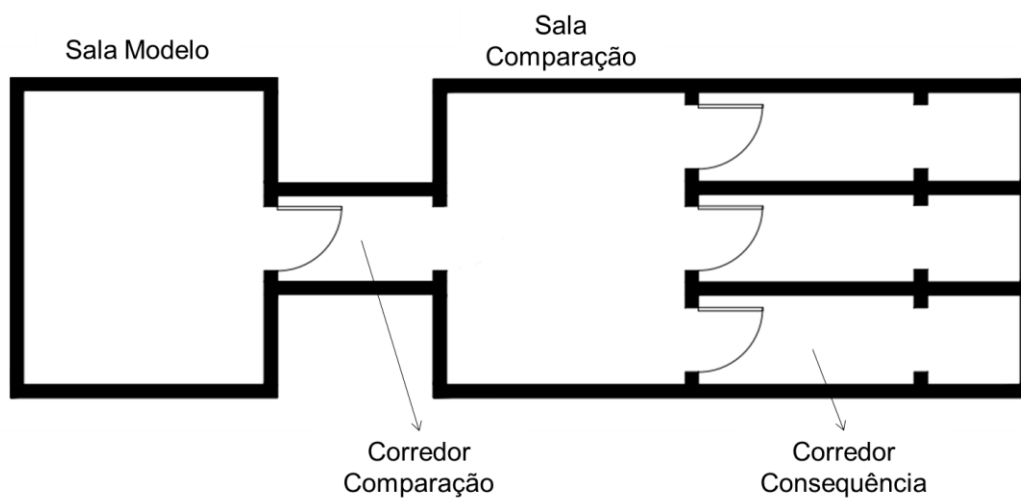
*Imagem do início do jogo Miner Troubles.*



**Tarefa experimental e estruturação do jogo.** Todas as tentativas de treino e teste do jogo eram apresentadas no mesmo espaço virtual dentro do jogo, composto por duas salas (Sala Modelo e Sala Comparação; ver Figura 6) e os corredores que interligavam essas salas (Corredor Comparação e Corredores Consequências).

**Figura 6**

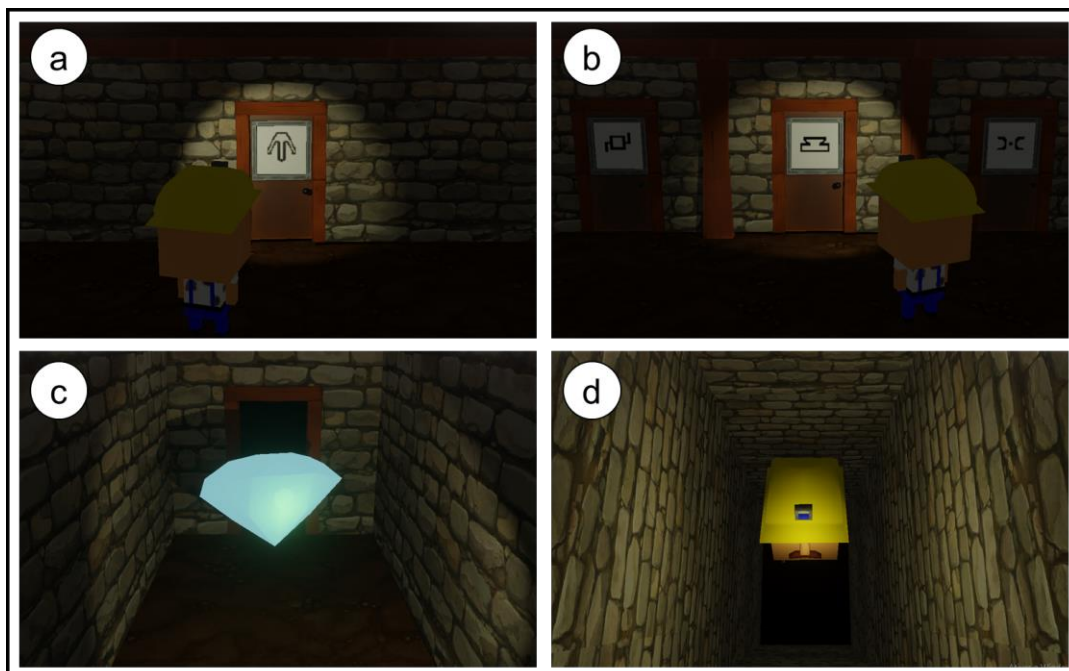
*Representação da planta baixa do jogo nas tentativas de treino e teste.*



Na Sala Modelo há uma porta que apresenta apenas um estímulo (no caso, o estímulo modelo; ver Figura 7a) e na Sala Comparação há três portas que apresentam um estímulo cada (ver Figura 7b). O Miner inicia todas as tentativas no início da sala modelo; ele deve abrir a Porta Modelo e escolher uma das três Portas Comparação. Durante as fases de treino, ao escolher a porta correta o participante encontra um diamante no corredor logo após a porta (Figura 7c) e continua seguindo o corredor até chegar novamente na Sala Modelo. Apesar de aparecer apenas um diamante no corredor, o número de diamantes que o participante pode ganhar é variável e pode ser definido pelo experimentador (ver “Métricas do Jogo” a seguir). Caso o participante erre, não é apresentado o diamante e, ao chegar no fim do corredor da consequência, o Miner cai em um buraco (Figura 7d), perdendo uma determinada quantidade de diamantes (definida pelo experimentador) e chegando novamente na Sala Modelo.

### Figura 7

*Imagens do Jogo Miner Troubles: (a) Apresentação da Porta Modelo, (b) Apresentação das Portas Comparação, (c) Consequência de acerto e (d) Consequência para erro.*



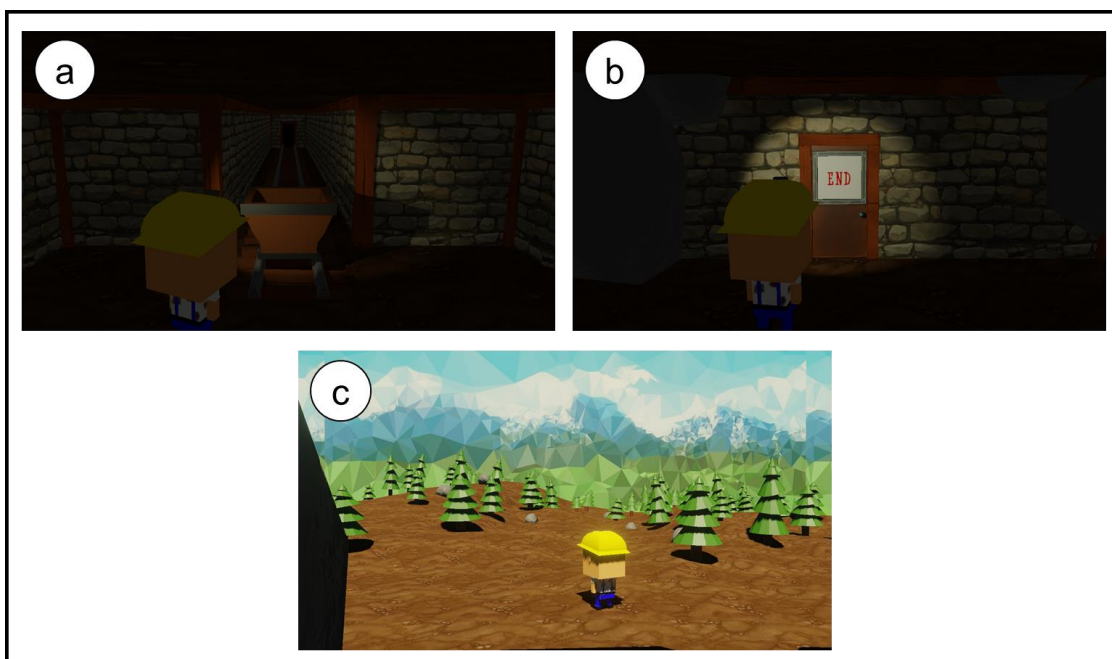
Ao finalizar qualquer bloco de ensino ou de teste, o Miner chega a uma sala de espera (ver Figura 8a), que possui um carrinho de mineração, caso o participante precise pausar o procedimento. A seguinte instrução é apresentada antes do participante chegar à sala:

“Esta é uma sala de descanso, caso queira dar uma pausa. Quando quiser continuar, basta clicar no vagão. Não feche o jogo agora pois ele não irá salvar.”

Ao final do procedimento, o Miner aparece em uma sala final, e ao abrir a porta desta sala, ele sai da tumba e chega em um espaço aberto (ver Figura 8b e 8c, respectivamente). Esse final ocorre independente do desempenho do participante nos testes, para que o final do procedimento não seja negativo, com o personagem preso dentro da tumba.

### Figura 8

*Imagens do Jogo Miner Troubles: (a) Tela de pausa do jogo entre os blocos de treino/teste, (b) Sala final, (c) Última cena do jogo.*



**Métricas do jogo.** As variáveis a que o participante tem acesso ao longo do jogo são duas: o número de portas corretas consecutivas e o número de diamantes (ver Figura 9, no canto superior esquerdo e direito, respectivamente). A primeira tem relação com o critério para passar de fase dentro do jogo. Caso o participante acerte a resposta em uma tentativa, é somado um ponto; caso o participante erre, o valor é zerado, independente de quantas



respostas corretas tenha feito anteriormente. Ao atingir o critério (no caso da Figura 9 são 12 tentativas corretas consecutivas), o participante passava para a sala de pausa e, depois, seguia o procedimento. O número de diamantes também é modificado de acordo com os acertos e os erros do participante e o jogo continua mesmo que este valor chegue a zero. Ambas as variáveis podem ser modificadas de acordo com o experimento, podendo-se aumentar ou diminuir o critério, e alterar o número de diamantes que são recebidos ou perdidos de acordo as escolhas dos participantes. Além disso, estas variáveis são apresentadas apenas nos blocos de treino do jogo, nos blocos de teste o participante não tem acesso a qualquer informação sobre o seu desempenho.

### Figura 9

*Imagem do jogo Miner Troubles com apresentação das variáveis a que o participante tem acesso ao longo dos blocos de treino (parte superior da imagem).*



Outras variáveis que também são registradas no fim de cada tentativa, são: nome do bloco, número da tentativa, se o participante acertou ou errou, o estímulo modelo, os estímulos de comparação e a porta selecionada. Também são registrados os intervalos de tempo: total da tentativa; entre o início da tentativa e o participante abrir a Porta Modelo; entre abrir a Porta

Modelo e chegar à Sala Comparação; e entre chegar à Sala Comparação e escolher uma das portas.

**Elementos de gamificação.** Utilizando uma taxonomia de elementos de gamificação (Toda et al., 2019), foram selecionadas as características do jogo relacionadas a outros procedimentos gamificados na área educacional: objetivos, progresso, pontos, puzzles, sensação, *status* e *storytelling*. Os objetivos do jogo são descritos logo no começo pelas instruções e continuam os mesmos até o fim do jogo: o participante é informado que ele deve coletar o maior número possível de diamantes e ajudar o Miner a sair daquele local. O progresso é feito pela passagem por diferentes fases dentro do jogo por meio dos blocos de treino e de teste, e pode ser acompanhado pelo número de portas corretas consecutivas. E os pontos no jogo são os diamantes.

A classificação de *puzzles* também se encaixa com o jogo pois ele exige que o participante aprenda que existe uma relação entre as imagens das portas, e responda de acordo com essas relações. Desta forma o que é exigido é uma aprendizagem simbólica e não uma aprendizagem motora, como ocorre em outros tipos de jogos (como jogos de plataforma). As sensações são relacionadas à ambientação do jogo (tanto visuais quanto sonoras). O *status* são as métricas a que o participante tem acesso, pois elas disponibilizam informações sobre o desempenho na tarefa. E por fim, o *storytelling* é o que agrupa todas essas variáveis descritas anteriormente em uma história de forma coesa e com significado para o participante.

## Parte 2.2 – Experimento 3: O efeito da magnitude das consequências no ensino de relações de equivalência

A adaptação dos Experimento 1 e 2 utilizando o jogo Miner Troubles requereria que algumas das contingências do jogo não estivessem presentes dependendo do grupo do qual o participante faria parte, como o ganhar diamantes ao abrir a porta correta ou perder diamantes ao abrir a porta errada. Neste último caso seria necessário se estabelecer uma quantidade total de diamantes que seria perdida ao longo da partida. Entretanto além da dificuldade de se estabelecer esse número, considerando que o procedimento não foi testado, alguns participantes poderiam chegar a zero diamantes caso o número inicial de diamantes não fosse o suficiente, visto que este desempenho é algo variável. Além disso, algumas das instruções e parte da história teriam que ser modificadas de acordo com o grupo. Todas essas alterações, somadas ao procedimento novo, poderiam prejudicar a análise da pergunta de pesquisa.

Uma maneira de se contornar esse problema foi apresentar os *feedbacks* de acerto e erro para todos os participantes. Desta forma, modificando a magnitude das consequências, não seria necessário realizar qualquer alteração no jogo ou na estrutura do procedimento. Esta variável é geralmente descrita como a “quantidade” da consequência, seja pela quantidade física, o valor reforçador/punitivo ou o tempo de apresentação da consequência (Bonem & Crossman, 1988; Cooper et al., 2020). No caso do jogo Miner Troubles, o tipo de variação da magnitude mais adequado seria pelo número de diamantes (pontos) ganhos ou perdidos de acordo com as escolhas. Deste modo, adaptando o procedimento, o Grupo Reforço ganharia mais pontos ao acertar, o Grupo Punição perderia mais pontos ao errar e o Grupo Reforço-Punição ganharia e perderia a mesma quantidade de pontos ao acertar ou errar. Entretanto, considerando que neste procedimento todos os grupos tiveram feedback para acerto e erro os nomes dos grupos foram alterados para: Grupo Mais-Ganho (GmG), Grupo Mais-Perda (GmP) e Grupo Balanceado (GB), respectivamente.

O objetivo do Experimento 3, portanto, foi igual ao dos Experimento 1 e 2: avaliar o efeito do uso de consequências reforçadoras e/ou punitivas na formação e manutenção de relações de equivalência. Entretanto, no Experimento 3 foram manipuladas diferentes magnitudes das consequências utilizadas (reforço e punição). Neste experimento não foram utilizados o Diferencial Semântico, o Fast ou o IRAP por conta dos problemas encontrados anteriormente e pelo contexto da coleta online. Entretanto, os participantes preencheram um questionário no fim do procedimento avaliando o jogo.

### ***Método***

**Participantes.** Participaram da pesquisa 56 pessoas (29 do gênero masculino e 27 do gênero feminino), e a média de idade dos participantes era de 27,8 anos (variando de 18 a 57 anos). Dois participantes haviam terminado o Ensino Médio (3,6%), 24 participantes estavam cursando uma graduação (42,8%), 7 participantes estavam cursando uma pós-graduação (12,5%), e 23 participantes estavam trabalhando (41,1%). Os critérios de inclusão foram: acesso à internet e a um computador com sistema Windows instalado. Os critérios de exclusão foram: já ter participado de pesquisas em equivalência de estímulos, ter cursado psicologia ou ser estudante de psicologia a partir do segundo ano da graduação.

**Equipamentos e Instrumentos.** Foi utilizando um computador específico para a coleta de dados. O computador foi montado pelo autor desta dissertação com o auxílio de um técnico em informática e conta com: Processador Intel I5 9400F, adequado para programação de aplicativos e com alto desempenho; Memória Ram Corsair Vegeance LPX 8gb 2400mhz; Disco Rígido SSD Western Digital 240gb; e Placa de vídeo GTX 1050 MSI, capaz de processar gráficos 3D e realizar *renderização* de maneira segura e rápida. Foi utilizada uma internet com conexão de fibra ótica e o computador foi conectado ao roteador utilizando um cabo de par trançado. Também foi utilizado o software Parsec para a transmissão da tela do computador do experimentador aos participantes. Este software é especializado na

transferência de tela em jogos online e consegue manter uma alta taxa de atualização da tela do computador além de permitir modificações nas configurações de apresentação e de rede. Todas essas características foram planejadas para que as condições do experimento fossem mais semelhantes possível entre os participantes, diminuindo o número de variáveis que poderiam prejudicar o procedimento.

Os estímulos utilizados foram os mesmos do Experimento 1 e 2, entretanto, neste experimento, o participante foi exposto ao ensino das relações de três classes de estímulos, com três membros cada (A1B1C1, A2B2C2 e A3B3C3). Foi utilizado o jogo Miner Troubles para a coleta de dados e os participantes foram distribuídos aleatoriamente em três grupos: Grupo Mais-Ganho (GmG), Grupo Mais-Perda (GmP) e Grupo Balanceado (GB). Estes Grupos diferiram em relação à quantidade de pontos (diamantes) que eles ganhavam/perdiam ao escolher as portas corretamente/incorrectamente de acordo com o estabelecido pelo experimentador. O Grupo Mais-Ganho recebia quatro diamantes ao acertar e perdia um diamante quando errava. O Grupo Mais-Perda recebia um diamante quando acertava e perdia quatro diamantes quando errava. E o Grupo Balanceado ganhava ou perdia um diamante ao acertar ou errar, respectivamente. Todos os participantes iniciavam o jogo com 20 diamantes.

Para todos os grupos, o critério para finalização dos blocos de treino no jogo era o número de respostas corretas consecutivas (ver descrição do jogo). Era acrescentado 1 ponto a essa variável a cada acerto e os participantes deveriam chegar a 12 respostas corretas consecutivamente para finalizar os blocos de treino. Todos os participantes iniciavam com zero respostas corretas e sempre que o participante errava, esse valor era zerado. Nos blocos de teste os participantes não tinham acesso ao número de diamantes ou ao número de respostas corretas consecutivas.

**Procedimento.** Todas as etapas do procedimento foram realizadas na casa do próprio participante. O pesquisador entrava em contato com o participante e enviava um formulário do *Google Forms* com todas as informações relativas à pesquisa. Após ler o Termo de Consentimento Livre e Esclarecido (Processo N°.: 40154520.1.0000.5504) e caso a pessoa aceitasse participar, ela era encaminhada para uma página com as instruções sobre como realizar o *download do software Parsec*, utilizado para transmitir a tela do computador do experimentador para o participante.

Todas as sessões foram realizadas individualmente e o formulário de instruções incluía sugestões para o participante procurar um local o mais tranquilo e silencioso possível, se sentar em frente ao seu computador com uma postura adequada e em uma mesa apenas com o computador. Também era sugerido ao participante avisar as pessoas com quem dividia o espaço que ele estaria fazendo uma pesquisa, para que não o interrompessem, e que deixasse o celular no modo silencioso e fora do seu alcance visual. Os participantes também eram instruídos a não anotar nada durante o procedimento. Após o envio do formulário, o experimentador concedia o acesso do computador da pesquisa ao participante e era iniciado o procedimento. O experimentador observava todas as ações dos participantes ao longo do jogo pela própria tela e ficava disponível durante todo o procedimento, caso o participante precisasse ou caso ocorresse algum problema técnico.

A sequência de blocos de treino e teste foi composta por um bloco de treino das relações AB, um bloco de treino das relações BC, um bloco de teste das relações simétricas (BA e CB), o bloco de treino misto das relações de linha de base e um bloco de teste das relações de transitividade e equivalência (ver Tabela 6). Desta forma, o bloco de teste das relações emergentes foi dividido em dois para que não ficasse muito extenso. Os blocos de treino eram finalizados quando o participante atingisse 12 tentativas corretas consecutivas. Os

blocos de teste tiveram um número fixo de 18 tentativas. O critério para considerar a formação da classe de equivalência era um mínimo de 88,9% de acertos no Teste de Equivalência (16 acertos em 18 tentativas). Os participantes foram convidados a realizar novamente o Teste de Equivalência sete dias depois. Todos os participantes que finalizaram o procedimento retornaram e a média entre o primeiro e o segundo teste foi de 7 dias ( $sd = 0.36$ ).

**Avaliação da pesquisa.** Após o Teste de Manutenção os participantes preencheram um questionário de *feedback* com cinco perguntas em uma escala *Likert* de cinco pontos sobre: costume em jogar jogos eletrônicos; estabilidade da internet/do jogo durante o procedimento; quanto o participante gostou do jogo; quão fácil ou difícil foi o jogo; e quanto o participante se sentiu focado no jogo durante a pesquisa. Também havia um espaço para os participantes escreverem um comentário, caso quisessem. Essas perguntas foram feitas para verificar se alguma dessas variáveis poderia ter interferido no desempenho dos participantes.

### Tabela 6

*Sequência de treino, número de tentativas em cada bloco, critério para finalização do bloco e relações treinadas ou testadas do Experimento 3.*

Bloco	Nº de tentativas	Critério para finalização do bloco	Relações
Treino AB	12	12 tentativas corretas consecutivas	A1B1, A2B2, A3B3
Treino BC	12	12 tentativas corretas consecutivas	B1C1, B2C2, B3C3
Teste de Simetria	18	X	B1A1, B2A2, B3A3, C1B1, C2B2, C3B3
Treino Misto	12	12 tentativas corretas consecutivas	A1B1, A2B2, A3B3, B1C1, B2C2, B3C3
Teste de Equivalência	18	X	A1C1, A2C2, A3C3, C1A1, C2A2, C3A3

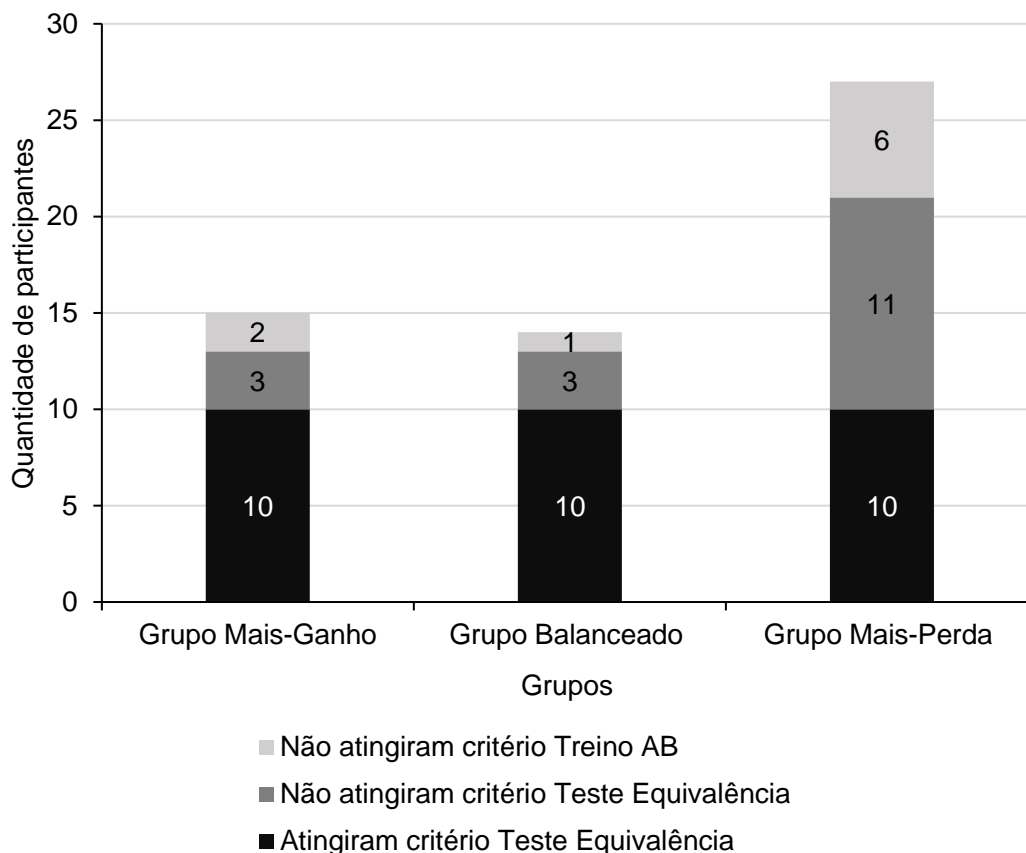
## ***Resultados***

Novos participantes foram sendo recrutados até que se chegasse a 10 participantes em cada grupo que atingissem o critério no Teste de Equivalência. Desta forma, o número total de participantes em cada grupo foi variável. A Figura 10 mostra o número total de cada grupo dividido em participantes que: atingiram critério no Teste de Equivalência; chegaram até o fim do procedimento, mas não atingiram critério no último teste; não atingiram o critério no primeiro bloco de treino (Treino AB). Neste último caso os participantes não compreenderam a tarefa ou não conseguiram aprender as relações do Treino AB. No total, o GmG precisou de 15 participantes, o GB precisou de 14 e o GmP precisou de 27 para que se chegasse a dez participantes que atingissem critério no Teste de Equivalência (respectivamente, 66,7%, 71,4% e 37,0% dos participantes). O Grupo Mais-Perda teve entre três e quatro vezes mais participantes que não atingiram critério no Teste de Equivalência ou no Treino AB, quando comparado aos outros dois grupos. Para todas as análises seguintes serão desconsiderados os participantes que não atingiram critério no Treino AB e as análises serão focadas nos dados dos 10 participantes de cada grupo que atingiram o critério no Teste de Equivalência.



**Figura 10**

*Número total de participantes em cada grupo do Experimento 3, de acordo com o desempenho na Fase 1.*



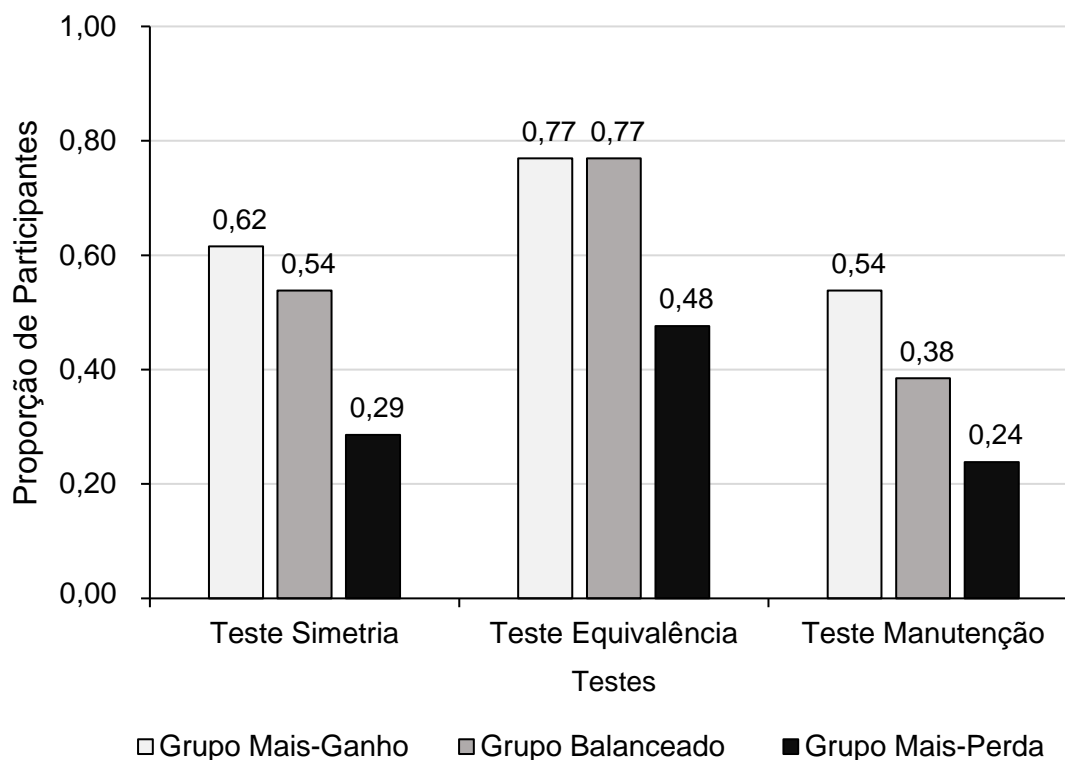
**Testes das relações emergentes.** Foram realizados três testes de relações emergentes:

Teste de Simetria, Teste de Equivalência e Teste de Manutenção (idêntico ao Teste de Equivalência, realizado 7 dias depois). Considerando o critério de no mínimo 89% de acertos (máximo de 2 erros no bloco de 18 tentativas) para ser considerado que o participante estabeleceu as relações emergentes e incluindo todos os participantes que chegaram ao fim do procedimento, é possível calcular a proporção de participantes que, em cada grupo, estabeleceram essas relações (ver Figura 11). O Grupo Mais-Ganho (GmG) foi o que demonstrou maior proporção de participantes que atingiram o critério em todos os testes. O Grupo Balanceado (GB), apesar de apresentar proporção semelhante no Teste de Equivalência ao GmG, teve um desempenho pior que este grupo no Teste de Simetria e no Teste de Manutenção. O Grupo Mais-Perda (GmP) foi o que teve o pior resultado em todos

os testes: menos de 50% dos participantes atingiram o critério no Teste de Equivalência e menos de 30% atingiram o critério nos outros dois testes.

### Figura 11

*Proporção de participantes de cada grupo do Experimento 3 que atingiram o critério em cada teste das relações emergentes.*

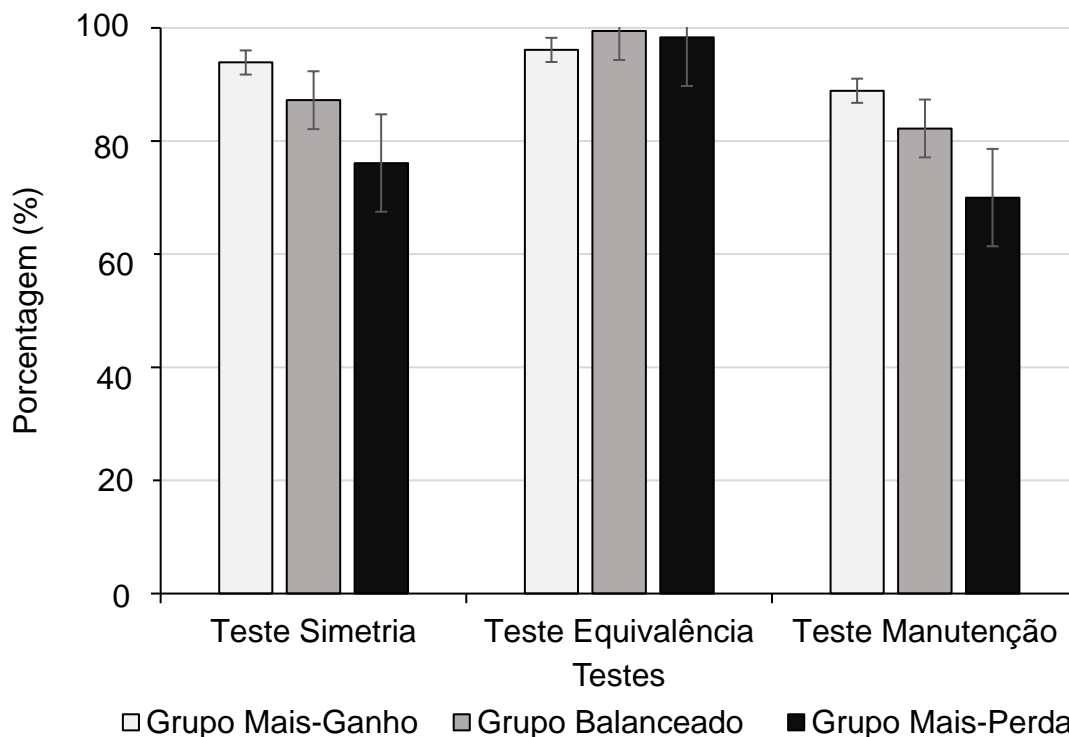


Entretanto, considerar apenas a proporção total de participantes que atingiram o critério de no mínimo 89% de acertos, inclui todos os participantes, inclusive os que não estabeleceram as classes de equivalência. Outra forma de se observar o resultado de cada grupo é verificando a média do desempenho nos testes dos dez participantes de cada grupo que atingiram o critério no Teste de Equivalência (ver Figura 12). Considerando apenas estes participantes, a média de acerto dos três grupos no Teste de Equivalência fica entre 90 e 100% de acertos, como esperado por conta do critério de seleção. Entretanto, ao se observar o Teste de Simetria e o Teste de Manutenção, os dados apresentam um padrão semelhante ao encontrado na proporção: o Grupo Mais-Ganho tem uma média de acertos superior e uma variabilidade menor dos dados (ver barras de erro padrão – Figura 12); o Grupo Balanceado

apresenta um desempenho intermediário entre os dois outros grupos; e o Grupo Mais-Perda tem uma variabilidade maior e o pior desempenho em comparação aos outros dois grupos.

**Figura 12**

*Porcentagem média de acertos nos testes dos participantes de cada grupo do Experimento 3 que atingiram o critério no Teste de Equivalência.*



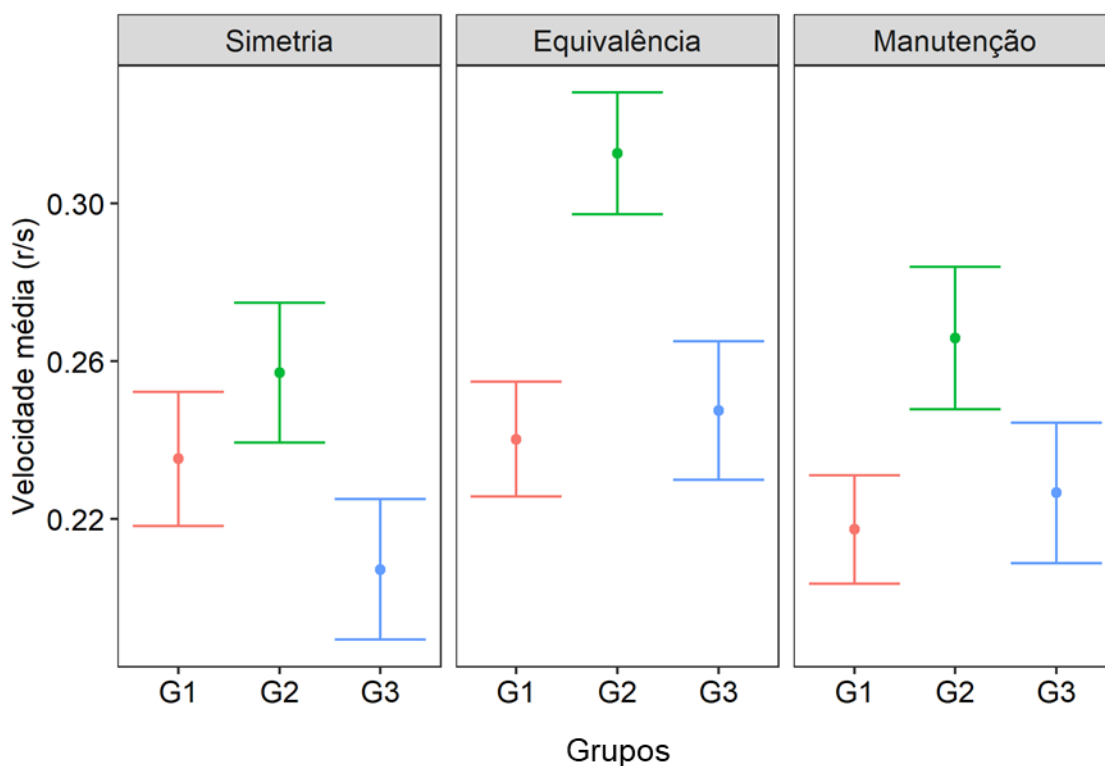
*Nota.* As barras de erro são relativas ao erro padrão.

**Velocidade de resposta.** Outra forma de se verificar o desempenho nos testes das relações emergentes é analisando a velocidade de resposta dos participantes diante dos estímulos de comparação (Spencer & Chase, 1996). O responder com maior velocidade pode ser tomado como indicativo de que as relações estejam melhor estabelecidas, ou seja, de que o grau de relacionamento entre os estímulos seja maior. A Figura 13 apresenta a velocidade de respostas dos três grupos, de acordo com o teste realizado (neste caso foram utilizados apenas os dados dos participantes que atingiram o critério no Teste de Equivalência). A velocidade é calculada dividindo-se a resposta (no caso, cada resposta individual do participante) pelo intervalo de tempo entre ele visualizar os estímulos comparação e escolher um deles (velocidade de resposta =  $1 / \text{latência da resposta}$ ). Neste caso, se a velocidade é maior, mais

rapidamente o participante escolheu o estímulo comparação e provavelmente mais forte é a relação. Nos gráficos é possível observar que o Grupo Balanceado foi o que apresentou a maior velocidade média de resposta, em comparação aos outros dois grupos. O Grupo Mais-Reforço apresentou uma velocidade média pouco maior do que o Grupo Mais-Punição no Teste de Simetria, entretanto nos outros testes a velocidade média de resposta desses dois grupos foi muito semelhante (ver Análise Estatística para mais detalhes das diferenças entre os grupos).

### Figura 13

*Velocidade média das respostas nos testes dos participantes de cada grupo do Experimento 3 que atingiram o critério no Teste de Equivalência.*

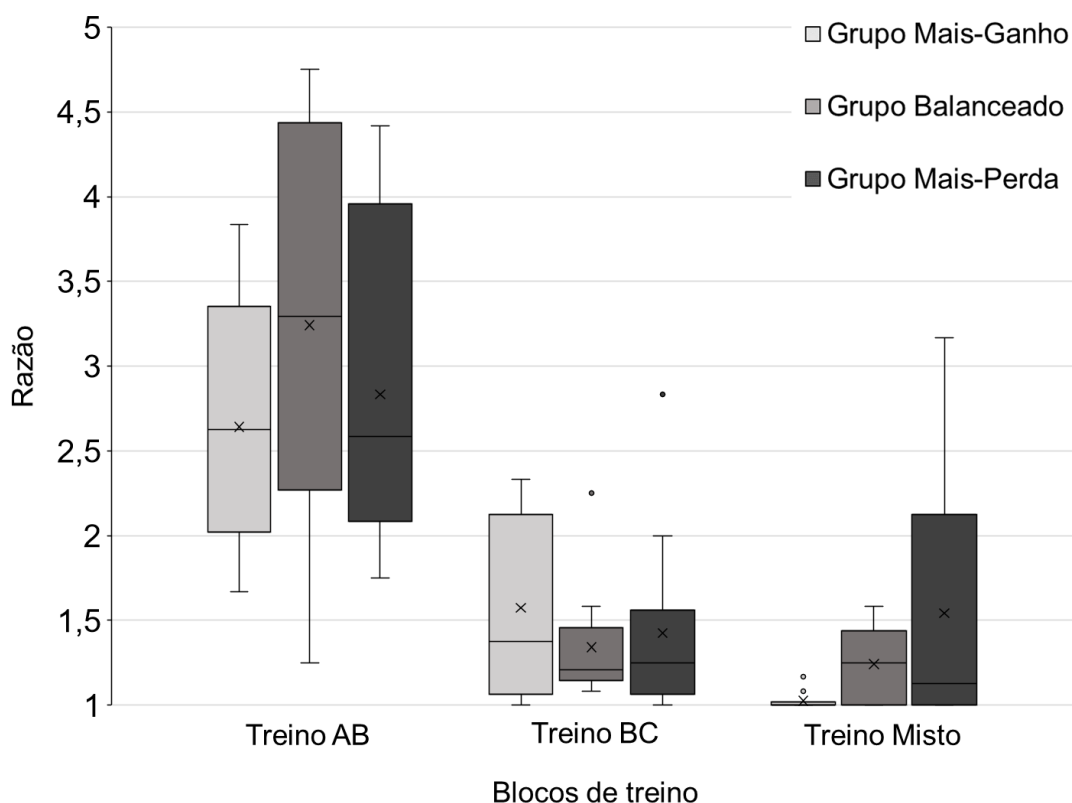


**Número de tentativas nos blocos de treino.** Neste experimento o critério para finalização dos blocos de treino foi a realização de 12 tentativas corretas consecutivas. Esse critério permitiu uma análise mais adequada da quantidade de tentativas que cada participante/grupo precisou para atingir o critério e passar de fase de treino, já que não dependia da porcentagem de acertos em cada bloco individual de treino (como ocorria nos Experimento 1 e 2).

Considerando que o número mínimo de tentativas realizadas pelos participantes em cada bloco de treino era 12, uma forma de se analisar o dado é calculando a razão do número de tentativas mínimas que foi necessário para o participante atingir o critério em cada bloco de treino (e.g., caso o participante tenha precisado de 12 tentativas, a razão das tentativas realizadas será igual a 1; caso ele precisasse de 36 tentativas, a razão será igual a 3). A Figura 14 mostra a distribuição em *boxplot* da razão de tentativas necessárias para que os participantes de cada grupo atingissem o critério. Nesta figura foram utilizados os dados dos participantes que apresentaram formação de classes no Teste de Equivalência.

**Figura 14**

*Razão das tentativas de treino (nº de tentativas/12) para cada bloco de treino dos participantes que atingiram o critério no Teste de Equivalência para todos os grupos do Experimento 3.*

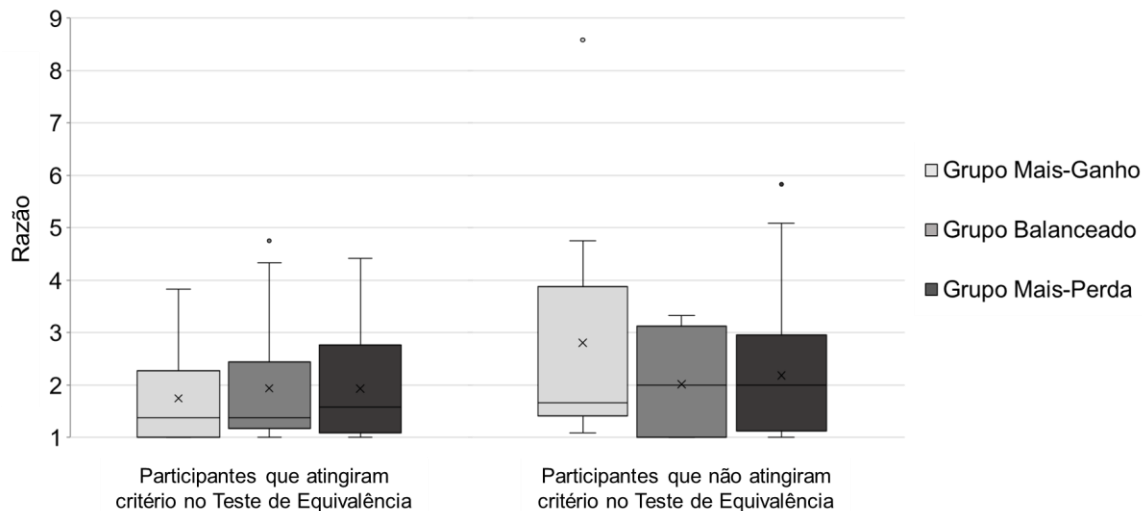


Na Figura 14, é possível observar primeiramente a redução, para todos os grupos, no número de tentativas necessárias para ser atingido o critério no Treino BC, quando comparado ao Treino AB. Provavelmente essa diminuição ocorreu devido à adaptação dos

participantes ao procedimento no primeiro bloco de ensino. Ou seja, no bloco de ensino AB, os participantes tinham que aprender não só como funcionava o jogo (incluindo se atentar aos estímulos nas portas), como também aprender a regra de condicionalidade (e.g., se A1 então B1, se A2 então B2, se A3 então B3), o que provavelmente fez com que os participantes levassem um maior número de tentativas para atingir o critério. Além disso, cada grupo precisou de mais tentativas em determinados blocos de treino: O Grupo Mais-Ganho teve um número médio de tentativas e uma distribuição dos dados maior em comparação aos outros grupos no Treino BC; o Grupo Balanceado apresentou o mesmo padrão no Treino AB; e o Grupo Mais-Perda também apresentou esse padrão de média acima dos outros grupos e maior distribuição dos dados no Treino Misto. Desta forma, apesar das diferenças entre os grupos no bloco no qual os participantes precisaram de mais ou menos tentativas de treino, todos os grupos tiveram uma quantidade semelhante de blocos de treino. Esse resultado fica mais claro ao se observar a razão total de tentativas para os todos os blocos de treino somados (ver Figura 15). Para todos os grupos, os participantes que atingiram o critério no teste de equivalência fizeram uma quantidade total de tentativas de treino semelhante. Além disso os dados dos participantes que não atingiram critério no Teste de Equivalência (ver Figura 15 – lado direito) são muito semelhantes aos dados dos participantes que atingiram critério (com exceção do Grupo Mais-Ganho). Este resultado sugere que as diferenças entre os grupos e entre os participantes que formaram as classes e os que não formaram provavelmente não foram consequência da quantidade de treino ao qual os participantes foram expostos (ver item Análise Estatística para mais detalhes).

**Figura 15**

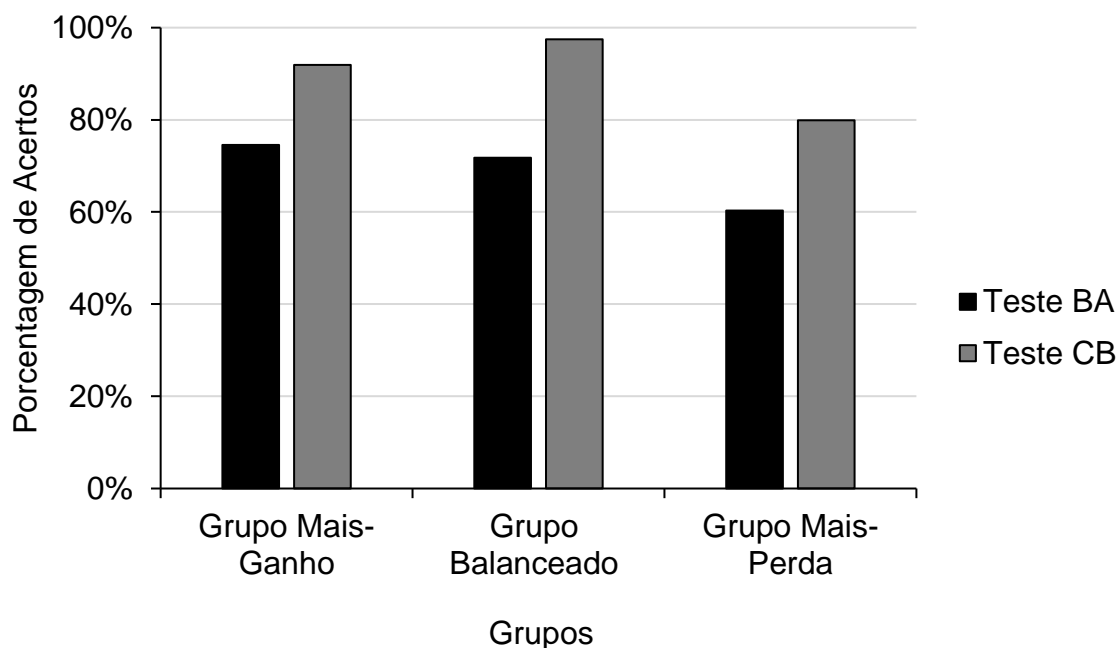
Razão do total das tentativas de treino ( $n^\circ$  de tentativas/12) para todos os participantes de cada grupo do Experimento 3.



**Teste de Simetria.** A Figura 16 apresenta o resultado de todos os participantes dos três grupos no Teste de Simetria. Como esse teste foi realizado depois do treino das relações AB e BC, foram testadas as relações BA e CB, que estão apresentadas separadamente na figura. Para todos os participantes, independente do grupo, o desempenho nos testes para a relação BA foi pior do que o desempenho para a relação CB.

**Figura 16**

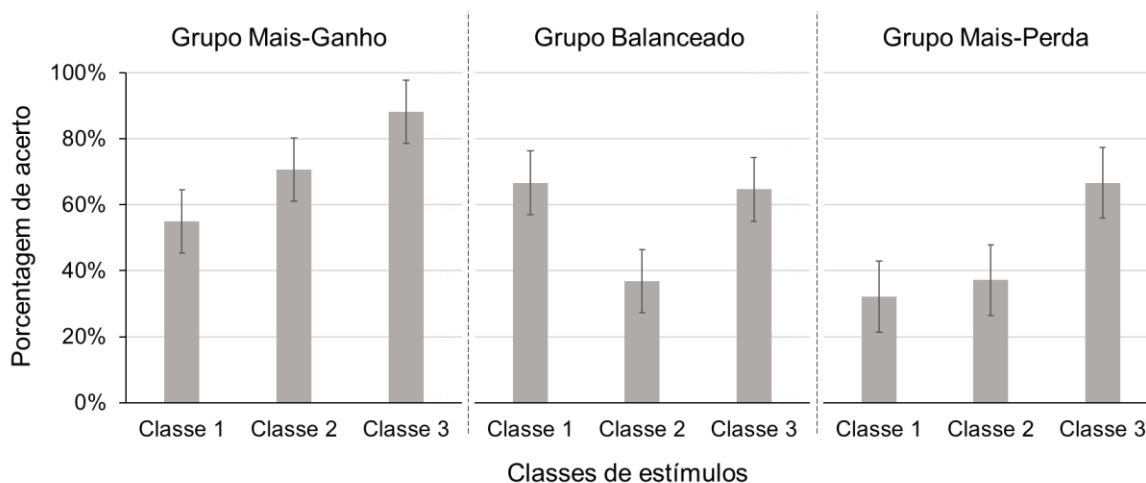
Porcentagem de acerto no Teste de Simetria das relações BA e CB para todos os participantes de cada grupo do Experimento 3.



**Análise de erros.** Para os participantes que formaram as classes, mas que não atingiram o critério no Teste de Manutenção, foi realizada uma análise de erros para se verificar se havia algum controle específico pelos estímulos utilizados. Para isso, foram calculadas as porcentagens de acertos para cada classe de estímulos. A Figura 17 apresenta a média da porcentagem de respostas corretas para as relações de cada classe (e.g., caso todos os participantes do GB tivessem acertado as relações A1C1 e C1A1, a média da Classe 1 para esse grupo seria igual a 100%). Os dados mostraram que para o GmG e o GmP os participantes tiveram um melhor desempenho (maior número de acertos) com as relações da Classe 3, enquanto o GB teve um desempenho pior (menor número de acertos) com os estímulos da Classe 2 em comparação às relações das outras classes. De forma geral, é possível perceber que há diferenças no desempenho dos participantes com relação às classes, principalmente da Classe 3 para o GmG e o GmP. Este resultado sugere que outras fontes de controle de estímulos podem ter ocorrido neste teste e que provavelmente algumas das características dos estímulos podem ter controlado as respostas dos participantes.

### Figura 17

*Porcentagem de acerto total do Teste de Manutenção das Classes 1, 2 e 3 para todos os participantes de cada grupo do Experimento 3 que não tenham atingido critério neste teste.*



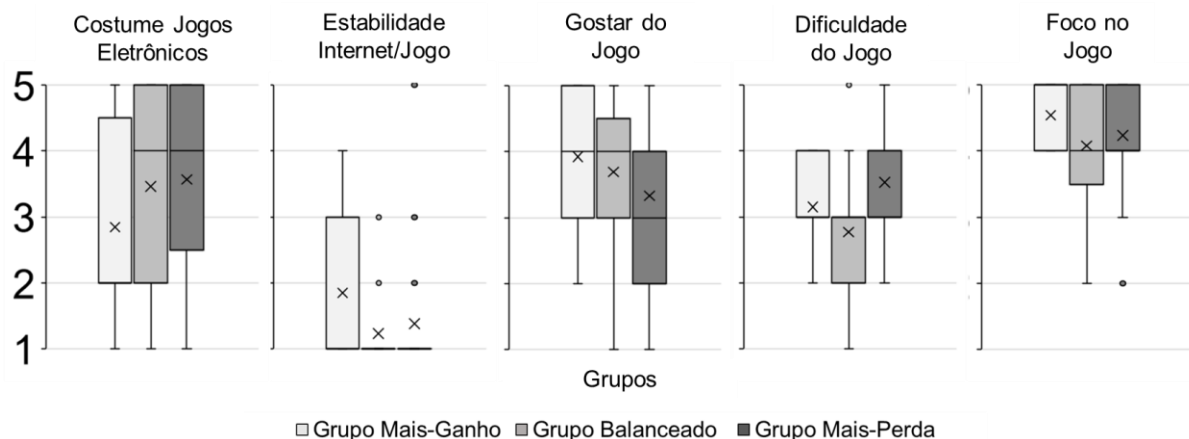
*Nota.* As barras de erro são relativas ao erro padrão.



**Avaliação do jogo pelos participantes.** Após o Teste de Manutenção os participantes respondiam um teste com algumas questões relativas ao procedimento (Figura 18). O Grupo Mais-Ganho foi o grupo que teve, em média, pessoas que jogavam menos jogos eletrônicos menos frequentemente e o grupo no qual mais ocorreram instabilidades na internet/jogo. Entretanto, de acordo com os dados dos testes e dos treinos, isso não pareceu ter influenciado o desempenho deste grupo. Com relação aos participantes gostarem do jogo, o Grupo Mais-Perda foi o que teve a menor média e uma distribuição maior de participantes que não gostaram do jogo, e o Grupo Mais-Ganho foi o grupo no qual os participantes, no geral gostaram mais do jogo, com uma maior proporção dos participantes sinalizando que gostaram do jogo. A média da avaliação da dificuldade do jogo ficou muito semelhante para todos os grupos (próxima de 3), entretanto, no Grupo Balanceado mais participantes avaliaram o jogo como mais fácil, em comparação aos outros dois grupos, que tiveram avaliações semelhantes. Por fim, de forma geral, a maioria dos participantes avaliou como tendo se mantido focados no jogo durante o procedimento. O Grupo Balanceado teve mais participantes avaliando como mantendo menos foco, entretanto, de acordo com os resultados dos treinos e testes, isso não pareceu ter influenciado no desempenho desse grupo. É importante destacar, entretanto, que apesar das diferenças apresentadas as avaliações dos grupos ficaram, no geral, muito semelhantes.

**Figura 18**

*Distribuição das respostas de todos os participantes de cada grupo do Experimento 3 para cada pergunta do questionário sobre a pesquisa.*



**Análise estatística.** As análises estatísticas foram realizadas de acordo com as hipóteses levantadas pelos dados. Para isso, foram realizados dois tipos de análises: análise de frequência e análise de regressão logística. Todas as análises foram realizadas utilizando o *software* RStudio.

**Análise de frequência.** Foi utilizado o Teste de Shapiro-Wilk para se avaliar a normalidade dos dados dos treinos e testes de todos os grupos e verificou-se que os dados não apresentavam uma distribuição normal ( $p < .05$ ); desta forma foram realizadas análises não paramétricas dos resultados. Para calcular o tamanho do efeito foi utilizada a estimativa Eta-quadrado, baseada na estatística H gerada pelo teste Kruskal-Wallis, o qual indicou que a magnitude do tamanho do efeito foi considerada forte para o teste de simetria e moderada para os testes de equivalência e manutenção.

O teste Kruskal-Wallis (variação não paramétrica do ANOVA One-way) indicou que o desempenho no Teste de Simetria foi estatisticamente diferente entre os grupos  $H(2, 30) = 6.51, p = .04, \eta_H^2 = .17$ , justificando análises post-hocs. O desempenho dos grupos no Teste de Equivalência  $H(3, 30) = 5.79, p = .12, \eta_H^2 = .11$  e no Teste de Manutenção  $H(8, 30) = 10.8, p = .22, \eta_H^2 = .13$ , foram estatisticamente equivalentes.

O tamanho do efeito da amostra foi estimado a partir do teste Cohen's (d) com correção de Hedges (g) para amostras menores que 20 (Espírito Santo & Daniel, 2017). O teste de Wilcoxon para amostras não pareadas demonstrou que não houve diferenças estatísticas na comparação entre os grupos no Teste de Simetria. Entretanto, a diferença de desempenho entre o GmG e o GmP ficou próximo do critério de diferença estatística ( $d = 1.02$ ,  $p = 0.079$ ).

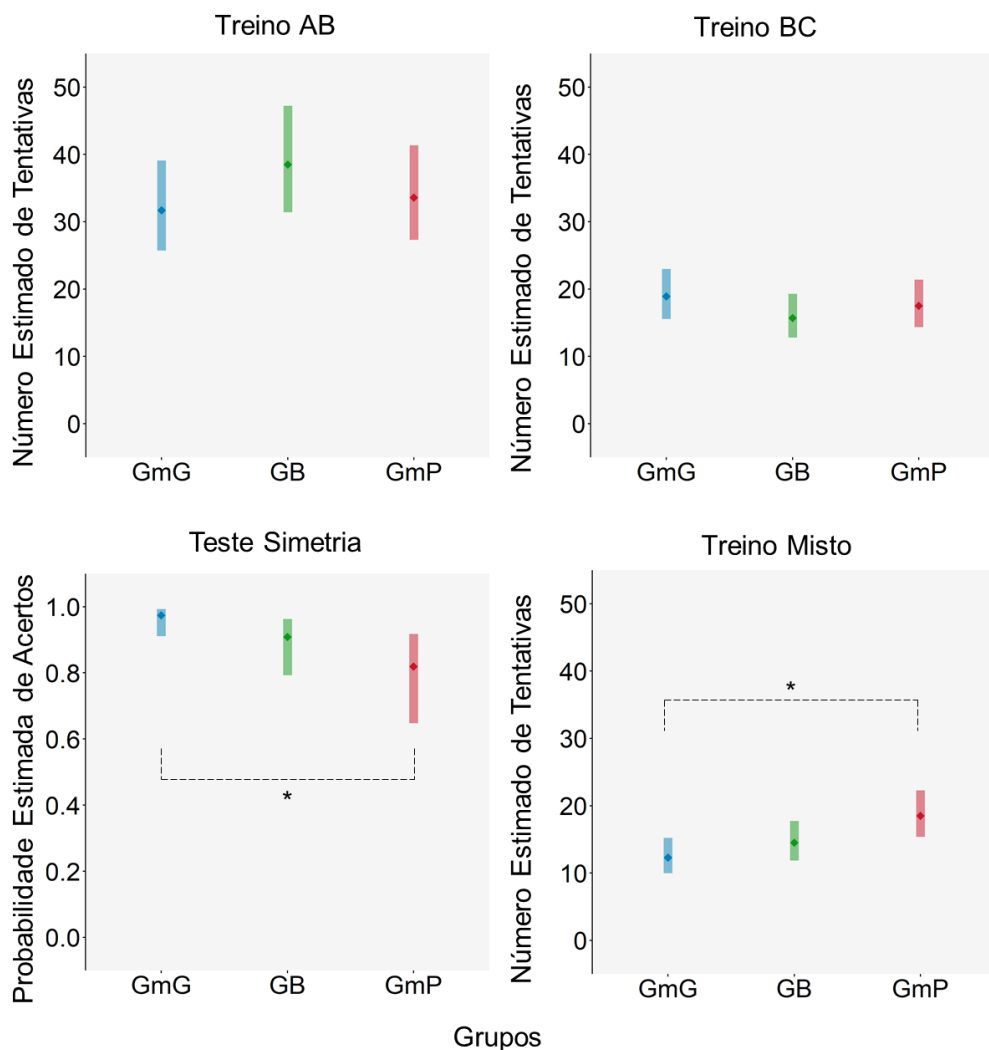
Para os testes de equivalência e manutenção, foi conduzido o teste não-paramétrico *Anova Type Statistic* (ATS), um modelo equivalente à *Two-way ANOVA*. As análises demonstraram que não há diferenças significativas quando os testes são comparados individualmente entre os grupos ( $p > 0.92$ ). Contudo, há uma diferença significativa quando observados apenas os testes de Equivalência e de Manutenção para todos os grupos ( $p < 0.03$ ). Além disso, há uma diferença marginalmente significativa quando considerada a interação entre os testes e os grupos ( $p < 0.065$ ). O teste *Wilcoxon Ranked* confirmou a diferença significativa entre os testes de equivalência e manutenção ( $p < 0.002$ ). E o *Simple effect test* demonstrou que não houve diferença entre os grupos quando foram analisados os testes de equivalência ( $p < 0.12$ ) e de manutenção ( $p < 0.52$ ) individualmente.

Não foram encontradas diferenças significativas entre os grupos para a quantidade de tentativas nos blocos de treino. Desta forma, foram conduzidas análises de correlação entre o número de tentativas de treino e o desempenho em cada um dos testes utilizando o teste de correlação de Spearman para amostras não paramétricas. Para o GmG, a quantidade de tentativas foi negativamente correlacionada de forma significativa com o desempenho em todos os testes, de simetria ( $\rho = -0,57$ ,  $p < 0,03$ ), de equivalência ( $\rho = -0,52$ ,  $p < 0,065$ ) e de manutenção ( $\rho = -0,56$ ,  $p < 0,05$ ). Para os outros grupos, a quantidade de tentativas também foi negativamente correlacionada com os testes, entretanto esse resultado foi significativo apenas para um teste para cada grupo, o Teste de Manutenção para o GB ( $\rho = -0,72$ ,  $p < 0,002$ ), e o Teste de Simetria para o GmP ( $\rho = -0,56$ ,  $p < 0,002$ ).

**Análise de Regressão.** Foram utilizados modelos lineares generalizados para a análise dos dados de número de tentativas de treino e o desempenho nos testes de relações emergentes. Este modelo pode ser utilizado na análise de dados que não sigam uma distribuição normal (como é o caso dos resultados do Exp. 3) por considerar diversos tipos de distribuições de acordo com as características da distribuição dos dados, além de permitir estimar a probabilidade de uma variável dependente de acordo com as variáveis independentes. Foram realizadas análises de desvios (Qui-quadrado de Wald Tipo II) e de variância (Anova Tipo II) e comparações pareadas *post hoc* de Tukey. Não houve diferença estatística entre os grupos no número de tentativas do Treino AB e do Treino BC. No Teste de Simetria e no Treino Misto, entretanto, houve efeito do grupo no desempenho dos participantes: O Grupo Mais-Ganho teve um desempenho melhor e estatisticamente diferente do Grupo Mais-Perda. É provável que as relações não foram bem estabelecidas para os participantes do GmP, o que ocasionou um desempenho pior no Teste de Simetria e, conseqüentemente, uma necessidade de mais tentativas no Treino Misto para que os participantes atingissem o critério para finalização do bloco, quando comparado aos outros grupos.

**Figura 19**

*Probabilidade do número de tentativas necessárias para cada um dos blocos de treino (Treino AB, BC e Misto) e probabilidade estimada de acertos no Teste de Simetria para cada um dos grupos.*



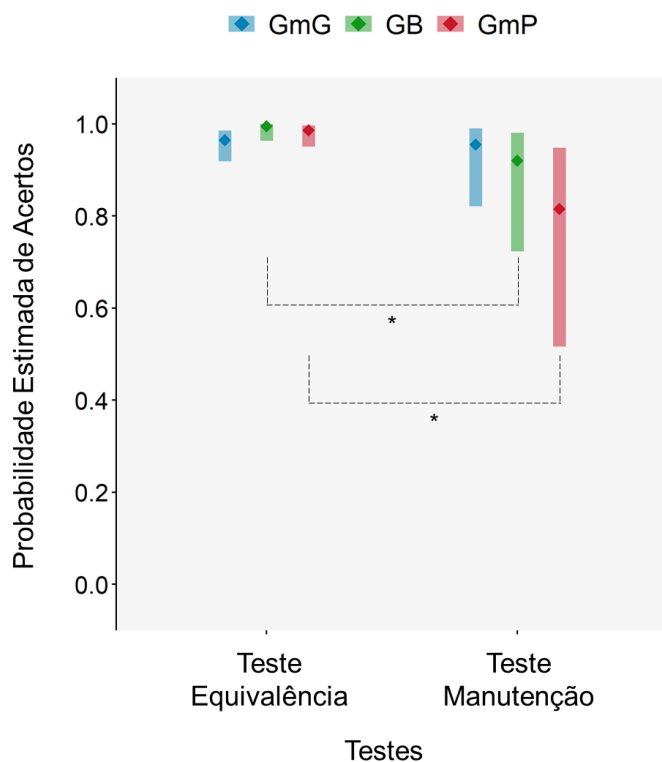
*Nota.* Foram utilizados os dados apenas dos participantes que atingiram o critério no Teste de Equivalência. As barras de erro denotam o intervalo de 95% de confiança.

O desempenho no Teste de Equivalência e no Teste de Manutenção não apresentou diferença entre os grupos, entretanto foi encontrado um efeito para os testes. Desta forma, apesar de não terem sido encontradas diferenças significativas entre os grupos de acordo com os testes individualmente, foram encontradas diferenças significativas quando foi comparado o desempenho de cada grupo de acordo com o teste. O Grupo Balanceado e o Grupo Mais-Perda apresentaram um desempenho menor e estatisticamente diferente quando comparado o

Teste de Equivalência com o Teste de Manutenção. Ou seja, o GB e o GmP tiveram uma piora significativa do desempenho no teste das relações AC e CA após uma semana.

### Figura 20

Probabilidade estimada de acertos no Teste de Equivalência e no Teste de Manutenção para cada um dos grupos.



*Nota.* Foram utilizados os dados apenas dos participantes que atingiram o critério no Teste de Equivalência. As barras de erro denotam o intervalo de 95% de confiança.

A análise da velocidade das respostas dos participantes nos testes das relações condicionais (ver Figura 13) também apresentou uma diferença significativa entre o desempenho dos participantes do Grupo Balanceado em comparação aos outros grupos para todos os testes. A única exceção foi o Teste de Simetria, no qual o desempenho do Grupo Mais-Ganho e do Grupo Balanceado não foi estatisticamente diferente, e houve uma diferença significativa entre o desempenho do Grupo Mais-Ganho e do Grupo Mais-Perda. Desta forma, esse dado corrobora os resultados do Teste de Simetria, demonstrando que neste teste o GmP não só teve um desempenho pior, quando comparado ao GmG, como teve uma menor velocidade de resposta dos participantes em comparação aos outros dois grupos.

## *Discussão*

O Experimento 3 teve como objetivo avaliar o uso de consequências reforçadoras e/ou punitivas na formação e manutenção de relações de equivalência, verificando o efeito de diferentes magnitudes de reforçamento e punição na aprendizagem de relações condicionais e, conseqüentemente, na emergência das relações não treinadas. Além disso, este experimento foi realizado com um procedimento de *Matching-to-Sample gamificado*, que foi adaptado para ser utilizado em situações com menor controle ambiental das condições experimentais. Desta forma, um dos aspectos mais importantes dos resultados, apesar de não ter sido mencionado anteriormente, foi que o procedimento gamificado foi capaz de ensinar e testar as relações condicionais de forma efetiva. Mesmo desconsiderando as diferentes contingências estabelecidas para cada grupo, mais da metade dos participantes (53,6%) aprendeu as relações arbitrárias e apresentou a emergência das relações não treinadas no Teste de Equivalência.

Esse número, entretanto, foi diferente para os grupos. O Grupo Mais-Perda foi o que apresentou o pior resultado entre os três grupos. Ele não só precisou de mais participantes para conseguir chegar até o número mínimo de 10 participantes que tivessem apresentado as relações de equivalência corretamente, como estes participantes, que formaram as classes, tiveram desempenhos menos acurados, em relação aos outros grupos, nos testes das relações emergentes de simetria e de manutenção. Os outros dois grupos (Mais-Ganho e Balanceado) apesar das diferenças nos procedimentos, apresentaram desempenhos superiores aos demais em testes diferentes. O Grupo Mais-Ganho teve uma perda menor do desempenho entre o Teste de Equivalência e o de Manutenção em comparação aos outros grupos, e o Grupo Balanceado apresentou uma velocidade de escolha dos estímulos comparação superior aos outros grupos (o que sugere uma maior força dessas relações para esse grupo). E como destacado nos resultados e na análise estatística, essa diferença entre os grupos

provavelmente ocorreu por conta das diferenças nas contingências de ensino, uma vez que não houve diferenças entre os grupos no número total de tentativas de treino.

Desta forma, os dados deste experimento vão, de certa forma, de encontro aos resultados encontrados anteriormente sobre o efeito do uso de estímulos punitivos na aprendizagem discriminativa (Brackbill & O'Hara, 1958; Penney, 1967; Penney & Lupton, 1961). Nestes estudos o comportamento discriminativo era apresentado mais rapidamente quando eram utilizadas consequências punitivas para erros em comparação a utilizar consequências reforçadoras para acertos. Entretanto, além de isso não ocorrer para os participantes do Grupo Mais-Perda, que tiveram um número de tentativas de treino semelhantes aos demais grupos, o uso de punição mais intensa afetou a emergência das relações não treinadas. Em adição, o Grupo Mais-Ganho que, considerando a hipótese levantada por esses estudos, deveria levar mais tempo para que os participantes aprendessem as relações, desempenhou de forma semelhante aos outros grupos nos treinos e teve um desempenho melhor nos testes em comparação ao Grupo Mais-Perda.

É importante destacar, entretanto, que o Experimento 3 não comparou, como nos estudos citados, o uso de reforçamento e/ou de punição isoladamente (como também foi feito nos Experimentos 1 e 2). No caso do Experimento 3 foram utilizados *feedbacks* para acertos e para erros para todos os grupos e foi manipulada a magnitude dessas consequências de acordo com o grupo. Em uma revisão de literatura da área de magnitude de reforçamento, Bonen e Crossman (1988) discutem que, mesmo após a realização de diversos estudos, a efetividade da manipulação exclusiva da magnitude de reforçamento ainda é uma questão em debate. Segundo os autores, enquanto alguns estudos demonstram efeitos muito pequenos da manipulação da magnitude de reforçamento, outros estudos demonstram efeitos robustos e duradouros. Essa análise foi reforçada por experimentos posteriores (Lerman et al., 2002; Reed, 1991), sugerindo que provavelmente o efeito da magnitude das consequências não é



um fator determinante do comportamento, quando não é utilizado em conjunto com outras variáveis. Como exemplo, alguns estudos realizados com pombos sobre esquemas concorrentes de reforçamento demonstraram que o comportamento é mais sensível à frequência do reforçamento do que à magnitude (Todorov, 1973; Todorov et al., 1984), quando estas duas variáveis são manipuladas em conjunto. Desta forma, seria necessário identificar quais as variáveis que modificam o comportamento quando utilizadas em conjunto com a manipulação da magnitude de reforçamento.

Entretanto, esses resultados, em sua maioria, são relativos à magnitude do reforçamento, e não da punição. E os estudos focados na diferença do efeito da magnitude de reforçamento e de punição sugerem que essas duas variáveis não são simétricas (Kubaneck et al., 2015; Rasmussen & Newland, 2008). Como exemplo, Kubaneck et al (2015) testaram o comportamento de escolha em uma tarefa de discriminação simples. Os participantes ouviam sons de clicks no ouvido direito e esquerdo (transmitidos por um fone), e deveriam escolher entre duas teclas do computador (do lado esquerdo ou direito) baseado no lado no qual ouviram mais sons. A dificuldade da tarefa era manipulada ao longo do estudo, de forma que a taxa de acertos/erros ficasse em 60/40 (ou seja, os participantes acertavam 60% das tentativas). Além disso, foi manipulada a magnitude de acertos e de erros para cada tentativa e foi registrada a porcentagem de vezes que os participantes repetiam a mesma tecla de acordo com a consequência da tentativa anterior. Os resultados mostraram que a magnitude de reforçamento tinha um efeito positivo e linear na escolha da mesma tecla, ou seja, quanto maior a magnitude do reforço, maior era a probabilidade de os participantes repetirem a tecla da tentativa anterior. Entretanto, a consequência de erro apresentou um efeito semelhante sobre a mudança de tecla (ou esquiva da tecla pressionada anteriormente), independente da magnitude. Segundo os autores, essa diferença demonstra uma assimetria: enquanto há uma função linear positiva entre o reforçamento e a repetição da resposta (quanto maior o valor do

estímulo reforçador, maior a probabilidade de o participante repetir aquela escolha), a punição apresenta uma função linear constante de supressão (independentemente do valor do estímulo punitivo, o participante tem uma probabilidade menor de repetir a escolha).

Contudo, não foi isso que ocorreu no Experimento 3. Todos os grupos foram expostos à contingência de punição, entretanto, foi apenas o grupo no qual a magnitude do estímulo punitivo era maior que o desempenho nas tarefas de aprendizagem discriminativa condicional foi prejudicado. Considerando a hipótese levantada pelos estudos citados anteriormente, a diferença na magnitude do estímulo punitivo não deveria alterar o padrão de respostas dos participantes de forma que isso interferisse no desempenho. Entretanto, outros estudos demonstram que essa a função linear de supressão da resposta ocorre apenas a partir de um ponto crítico (de Souza et al., 1984; Perone, 2003), no qual a punição afeta o comportamento do sujeito. De Souza et al (1984) verificaram o efeito da intensidade de choque elétrico no comportamento de esquiva sinalizada em ratos. A intensidade dos choques variou de valores muito baixos (0,1 mA) a valores muito altos (8 mA) e, de forma geral, o comportamento de esquiva dos ratos só começou a ser aprendido e mantido a partir de um valor específico (entre 0,8 e 1,0 mA). Aumentos subsequentes na intensidade de choque dos ratos mantinham taxas de respostas similares, assim como encontrado por Kubanek et al (2015). E choques com intensidade muito baixa (menores de 0,8 mA) não mantinham o comportamento dos ratos, ou seja, eles não se esquivavam do estímulo (que, provavelmente, ainda não apresentava caráter punitivo/aversivo). No caso do estudo de Kubanek et al (2015) esse padrão provavelmente não foi encontrado porque não foram manipuladas diferenças tão grandes do valor dos estímulos punitivos e reforçadores. Além disso, a consequência punitiva era relativamente simples, quando comparada a outras consequências como choque ou sons aversivos, que podem causar efeitos emocionais desagradáveis (Bradley & Lang, 2007).

Considerando os resultados destes estudos de que, a punição em intensidades muito baixas pode não gerar respostas de esquiva, e que a partir de um determinado valor estas respostas passam a ocorrer independente do aumento na intensidade da punição (de Souza et al., 1984; Kubanek et al., 2015; Perone, 2003), uma hipótese para os resultados do Experimento 3 é que a magnitude do estímulo punitivo utilizado para o Grupo Mais-Perda foi mais aversiva, prejudicando o desempenho dos participantes deste grupo, enquanto para o Grupo Mais-Ganho e para o Grupo Balanceado o estímulo punitivo pode não ter apresentado funções aversivas (ou seja, a intensidade da punição pode ter ficado abaixo do limiar necessário para se gerar respostas de esquiva). Desta forma, é provável que os participantes do GmP tenham emitido respostas de esquiva das consequências para erros (selecionando, conseqüentemente, o estímulo comparação correto), mas que estas respostas não contribuíssem para a aprendizagem dos participantes e para a emergência das relações não treinadas. Enquanto para o GmG e o GB, o uso de uma punição com menor intensidade pode ter favorecido a aprendizagem dos participantes ao sinalizar as respostas incorretas sem gerar comportamento de esquiva da consequência para erro. Resultado que corrobora estudos anteriores que demonstraram o efeito benéfico da punição na aprendizagem discriminativa, aumentando a acurácia dos participantes (Brackbill & O'Hara, 1958; Meyer & Offenbach, 1962; Penney, 1967; Penney & Lupton, 1961). Um dado que corrobora essa análise é a avaliação do jogo pelos participantes: o GmP foi o grupo que teve um maior número de participantes avaliando o jogo de uma forma negativa, quando comparado aos outros grupos. Além disso, alguns participantes do GmP relataram, ao fim da pesquisa, sentimentos como “irritação”, “receio de errar”, “desconforto” e “raiva” durante as fases de treino. Nos outros grupos, esse relato apareceu apenas para um participante do GB, que estava se referindo ao bloco de teste, no qual não havia consequências para acerto ou erro.

Pode-se argumentar que esses efeitos não são exclusivos de procedimentos que utilizam punição. Alguns estudos já demonstraram que mesmo em condições de ensino envolvendo apenas reforçamento positivo, os sujeitos podem se esquivar da tarefa (Azrin, 1961; Hunziker, 2017; Perone, 2003). Entretanto, os estudos que abordam essa questão utilizam, em sua maioria, ratos ou pombos como sujeitos experimentais. Se, como destacado por Todorov (2001), o estudo do efeito do uso de punição tem sido “relegado”, essa lacuna é ainda maior para estudos sobre comportamentos complexos, como no caso da discriminação condicional e formação de classes de equivalência com humanos.

Desta forma, a hipótese de que a perda de pontos funcionou como um estímulo punitivo apenas para o Grupo Mais-Perda, ou outras explicações sobre o que fez com que o desempenho do GmP fosse menos acurado do que o desempenho dos outros grupos ainda são questões a serem verificadas; é preciso considerar, ainda, que, se a perda de pontos/diamantes funcionou como punição, deveria ter favorecido um desempenho mais (e não menos) acurado (no caso do GmP). Estudos futuros podem manipular a magnitude da punição ampliando os valores utilizados e verificando se há um valor crítico no qual os estímulos punitivos adquirem propriedades aversivas (gerando comportamentos de esquiva). E apesar de alguns estudos sugerirem que o uso de punição melhora o desempenho em tarefas discriminativas simples (Brackbill & O’Hara, 1958; Penney, 1967; Penney & Lupton, 1961), novas pesquisas devem ser realizadas para se compreender melhor o efeito do uso de punição em tarefas de discriminação, principalmente utilizando outras populações e diferentes estímulos punitivos. Além disso, ainda não é claro o efeito que o uso de punição tem na aprendizagem discriminativa condicional, que é mais complexa, e tem outras consequências, como: a emergência de novas relações (Matos, 1999; Sidman, 2000) e a transferência de funções (Bortoloti et al., 2019; Silveira et al., 2016, 2021), que podem alterar a força das relações e a estabilidade delas à passagem de tempo. Desta forma, a generalização dos resultados de

pesquisas em discriminação simples para discriminação condicional deve ser tomada com cautela.

**Magnitude de reforçamento.** Em contraposição aos efeitos da maior magnitude da punição, a maior magnitude de reforçamento não pareceu influenciar o desempenho dos participantes do Grupo Mais-Ganho. Apesar de este grupo ter apresentado desempenho mais acurado no teste de manutenção em comparação aos outros grupos, seu desempenho foi muito semelhante ao do Grupo Balanceado. Comparando-se a avaliação da velocidade de resposta de escolha dos estímulos comparação, o GmG foi mais lento do que o GB, o que pode indicar uma maior força das relações para este último grupo. Apesar de esses resultados corroborarem alguns estudos e revisões de literatura da área, que demonstram que a magnitude de reforçamento tem efeitos pequenos em esquemas de reforçamento mais simples (Bonem & Crossman, 1988), como foi o caso do Experimento 3, uma maior parte dos estudos foi realizada com animais não humanos, e mesmo os estudos feitos com seres humanos geralmente são realizados com crianças em pesquisas aplicadas (Lerman et al., 2002; Trosclair-Lasserre et al., 2008). Além disso, no Experimento 3 foram utilizados pontos como estímulos reforçadores, e não foram manipulados diferentes valores da magnitude de reforçamento, como realizado em alguns estudos (Kubanek et al., 2015). Desta forma, os resultados do Experimento 3 com relação à magnitude de reforçamento não são conclusivos, nem esgotam as possibilidades de pesquisa nessa área. Novos estudos devem ser produzidos para se verificar o papel dessa variável em populações adultas e verbais, e em tarefas de discriminação condicional.

**Teste de simetria.** O Teste de Simetria verificava a emergência de duas relações (BA e CB), e para todos os grupos o desempenho dos participantes foi menos acurado para a relação BA em comparação à relação CB. Em um estudo sobre generalização de relações de

equivalência, Ayres-Pereira, Canovas, Varella e de Souza (2018) encontraram o mesmo resultado ao treinar primeiramente as relações AB e posteriormente as relações BC ou AC. Neste caso, foi utilizado um procedimento de sondas múltiplas, nas quais eram testadas todas as relações potencialmente emergentes, antes e depois de cada fase de ensino das relações. Dos seis participantes, apenas um apresentou 100% de acertos no teste de simetria da primeira relação aprendida (AB) após o ensino da segunda relação (BC ou AC). Nesse estudo também foi utilizado um bloco de treino misto para que o desempenho dos participantes se estabilizasse sem diferenças entre as relações ensinadas antes da continuação do procedimento. Desta forma, além de demonstrar que esse processo também pode ocorrer com adultos, os resultados do Experimento 3 confirmam a importância de se realizar um bloco de treino misto de todas as relações treinadas, em preparação para os testes. Novas pesquisas devem ser realizadas para se compreender melhor as variáveis que afetam essa diferença no teste de simetria (e.g., passagem de tempo, ensino de uma nova relação com pelo menos um membro do ensino anterior, exposição a uma tarefa de discriminação, entre outros).

**Características dos estímulos.** Em pesquisas básicas sobre classes de equivalência, os estímulos utilizados são, geralmente, abstratos<sup>7</sup>. Desta forma, é provável que as imagens utilizadas não possuam funções discriminativas específicas. Entretanto, isso não significa que os estímulos deixem de apresentar características que possam adquirir funções durante o treino das relações. Um dado que indica esse processo é o maior número de acertos dos estímulos da Classe 3 no Teste de Manutenção das relações AC e CA, para os participantes que tinham atingido critério no Teste de Equivalência do GmG e do GmP. Todos os estímulos utilizados eram figuras abstratas e foram escolhidas de forma semi-aleatória, para que nenhuma imagem da mesma classe se parecesse muito uma com a outra. Entretanto, esse

---

<sup>7</sup> A denominação “familiar” ou “não familiar” é mais adequada do que o uso do termo “abstrato”, afinal trata-se de uma pressuposição considerar que o estímulo é abstrato para o participante. Entretanto, muitos estudos utilizam essa definição, e não é o objetivo desse texto tratar conceitualmente dessa diferença de nomenclatura.

resultado demonstrou que é necessário se randomizar os estímulos utilizados no ensino de cada classe entre os participantes. Alguns estudos já fizeram esse tipo de controle dos estímulos utilizados (Dougher et al., 1994), escolhendo os estímulos de cada classe de forma randômica para cada participante. Desta forma, todos os participantes eram expostos a uma condição de ensino com estímulos e relações diferentes. Entretanto, essa é uma prática ainda pouco realizada e que pode influenciar os resultados das pesquisas. No caso do Experimento 3, o padrão de respostas, apesar de ter sido diferente entre as classes, foi muito semelhante entre os grupos, o que provavelmente não prejudicou os resultados. Contudo não é possível afirmar se essa variável pode interferir ou pode já ter interferido em outros estudos.

## Conclusão Geral

Este estudo tinha como objetivo avaliar o efeito do uso de consequências reforçadoras e/ou punitivas na formação e manutenção de relações de equivalência. No Experimento 1 e 2 foi utilizado um procedimento de uso exclusivo de punição, uso exclusivo de reforçamento ou do uso de reforçamento e punição de forma conjunta. E no Experimento 3 essa condição foi replicada utilizando um procedimento com diferentes magnitudes de reforçamento e de punição combinadas. Com base em estudos anteriores, com discriminação simples, a hipótese era a de que a punição facilitaria a aprendizagem das relações de linha de base e, também a formação de classes e que aumentaria o grau de relacionamento entre os estímulos nas classes formadas experimentalmente. No entanto, de forma geral, considerando os dados dos três experimentos, o uso de punição ou de uma maior magnitude da punição prejudicou a aprendizagem. Um menor número de participantes conseguiu chegar até o fim do procedimento e um menor número de participantes atingiu o critério nos testes de emergência das relações não treinadas, nos grupos nos quais foram utilizadas consequências punitivas ou consequências punitivas mais intensas. Além disso, não houve diferenças na quantidade de tentativas de treino de acordo com as contingências utilizadas (reforçamento, punição ou manipulação da magnitude das consequências). Entretanto, os resultados ainda não podem ser estendidos para o uso exclusivo de punição e/ou reforçamento; desta forma, novos estudos devem avaliar essa questão, além de verificar o efeito de uma manipulação mais ampla de diferentes magnitudes de reforçamento na aprendizagem e na manutenção de relações condicionais.

Para viabilizar este estudo, o Experimento 3 também contribuiu com a criação de um procedimento gamificado de *Matching-to-Sample*. Esse procedimento demonstrou ser eficaz no ensino e no teste de relações de equivalência. Mesmo utilizando uma estrutura de treino linear e apenas estímulos abstratos, ambas variáveis que dificultam a aprendizagem das



relações (Arntzen, 2004, 2012; Arntzen & Holth, 1997), considerando apenas o Grupos Maior-Ganho e o Grupo Balanceado, 77% dos participantes que chegaram ao fim do procedimento formaram as classes. Mesmo considerando que a pesquisa foi conduzida de forma remota, o que pode ter prejudicado a concentração dos participantes, o desempenho nos testes das relações emergentes foi mais acurado para os participantes que utilizaram o software Miner Troubles do que para os participantes que realizaram a coleta de maneira presencial e utilizando um software padrão de *Matching-to-Sample* (considerando os participantes do Grupo Reforço e Grupo Reforço-Punição dos Experimentos 1 e 2, 53% dos participantes atingiram critério no teste de equivalência). Entretanto, os experimentos tiveram muitas diferenças de procedimento (número de classes ensinadas, número de estímulos por classe, etapas do procedimento, testes realizados) que não permitem comparações diretas, apesar de apontar algumas tendências que devem ser avaliadas em novas investigações. Novas pesquisas são necessárias para se comparar o procedimento tradicional de MTS com o jogo Miner Troubles, e para se avaliar a viabilidade de realização de pesquisas online utilizando estes procedimentos.

É importante destacar que a criação de procedimentos gamificados é uma tendência na área educacional (de Sousa Borges et al., 2014), e considerando todas as aplicações para o ensino de leitura e escrita que já foram realizadas (de Rose et al., 1996; de Souza et al., 2009; Reis et al., 2009), é imprescindível que estudos sobre gamificação sejam realizados na área de equivalência de estímulos. Considerando os resultados do Experimento 3, novos estudos podem verificar o ensino de leitura e escrita para crianças com dificuldade nessa tarefa utilizando o jogo Miner Troubles ou realizando novas adaptações gamificadas do procedimento de *Matching-to-Sample*.

## Referências

- Albright, L. K., Fields, L., Reeve, K. F., Reeve, S. A., & Kisamore, A. N. (2019). Relatedness of Equivalence Class Members: Combined Effects of Nodality and Relational Type. *The Psychological Record*, *69*(2), 277–289. <https://doi.org/10.1007/s40732-019-00329-6>
- Almeida, J. H. de, Bortoloti, R., Ferreira, P. R. dos S., Schelini, P. W., & de Rose, J. C. C. (2014). Análise da Validade e Precisão de Instrumento de Diferencial Semântico. *Psicologia: Reflexão e Crítica*, *27*(2), 272–281. <https://doi.org/10.1590/1678-7153.201427207>
- 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. <https://doi.org/10.1002/jeab.19>
- Arntzen, E. (2004). Probability of Equivalence Formation: Familiar Stimuli and Training Sequence. *The Psychological Record*, *54*(2), 275–291. <https://doi.org/10.1007/BF03395474>
- Arntzen, E. (2012). Training and Testing Parameters in Formation of Stimulus Equivalence: Methodological Issues. *European Journal of Behavior Analysis*, *13*(1), 123–135. <https://doi.org/10.1080/15021149.2012.11434412>
- Arntzen, E., Granmo, S., & Fields, L. (2017). The Relation Between Sorting Tests and Matching-to-Sample Tests in the Formation of Equivalence Classes. *The Psychological Record*, *67*, 81–96. <https://doi.org/10.1007/s40732-016-0209-9>
- Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, *47*(2), 309–320. <https://doi.org/10.1007/BF03395227>
- Arntzen, E., Nartey, R. K., & Fields, L. (2015). Enhanced equivalence class formation by the

- delay and relational functions of meaningful stimuli. *Journal of the Experimental Analysis of Behavior*, *103*(3), 524–541. <https://doi.org/10.1002/jeab.152>
- Arntzen, E., Norbom, A., & Fields, L. (2015). Sorting: An Alternative Measure of Class Formation? *The Psychological Record*, *65*, 615–625. <https://doi.org/10.1007/s40732-015-0132-5>
- Ayres-Pereira, V., Canovas, D. S., Varella, A. A. B., & de Souza, D. G. (2018). Generalization of equivalence relations from photos to objects by preschool children. *Journal of the Experimental Analysis of Behavior*, *109*(2), 394–407. <https://doi.org/10.1002/jeab.313>
- Azrin, N. H. (1961). Time-out from Positive Reinforcement. *Science*, *133*(3450), 382–383. <https://doi.org/10.1126/science.133.3450.382>
- Barnes-Holmes, D., Finn, M., McEnteggart, C., & Barnes-Holmes, Y. (2018). Derived Stimulus Relations and Their Role in a Behavior-Analytic Account of Human Language and Cognition. *Perspectives on Behavior Science*, *41*(1), 155–173. <https://doi.org/10.1007/s40614-017-0124-7>
- Barnes-Holmes, D., Keane, J., Barnes-Holmes, Y., & Smeets, P. M. (2000). A derived transfer of emotive functions as a means of establishing differential preferences for soft drinks. *Psychological Record*, *50*(3), 493–511. <https://doi.org/10.1007/BF03395367>
- Bonem, M., & Crossman, E. K. (1988). Elucidating the Effects of Reinforcement Magnitude. *Psychological Bulletin*, *104*(3), 348–362. <https://doi.org/10.1037/0033-2909.104.3.348>
- Bortoloti, R., Rodrigues, N., Cortez, M. D., Pimentel, N., & de Rose, J. C. D. (2013). Overtraining increases the strength of equivalence relations. *Psychology & Neuroscience*, *6*(3), 357–364. [https://www.scielo.br/scielo.php?pid=S1983-32882013000300013&script=sci\\_arttext](https://www.scielo.br/scielo.php?pid=S1983-32882013000300013&script=sci_arttext)
- 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., de Almeida, R. V., de Almeida, J. H., & de Rose, J. C. (2021). A Commentary on the Dynamics of Arbitrarily Applicable Relational Responding Involving Positive Valenced Stimuli and its Implications for the IRAP Research. *The Psychological Record*, *71*(3), 481–486. <https://doi.org/10.1007/s40732-020-00413-2>
- Bortoloti, R., & de Rose, J. C. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential. *The Psychological Record*, *59*(4), 563–590. <https://doi.org/10.1007/BF03395682>
- 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. <https://doi.org/10.1080/15021149.2011.11434359>
- 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). *Psychological Record*, *62*(1), 41–54. <https://doi.org/10.1007/BF03395785>
- Bortoloti, R., & de Rose, J. C. 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. <https://doi.org/10.1590/s0102-79722011000200020>
- Brackbill, Y., & O’Hara, J. (1958). The relative effectiveness of reward and punishment for discrimination learning in children. *Journal of Comparative and Physiological Psychology*, *51*(6), 747–751. <https://doi.org/10.1037/h0038282>
- Bradley, M. M., & Lang, P. J. (2007). The International Affective Digitized Sounds Affective Ratings of Sounds and Instruction Manual. *Technical Report B-3*. University of Florida, Gainesville, Fl., 29–46.

- Capobianco, D., Teixeira, C., Bela, R. E., Orlando, A. F., de Souza, D. G., & de Rose, J. C. (2009). *LECH-GEIC. Sistema web gerenciador de ensino individualizado por computador.*
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). Positive Reinforcement. In *Applied behavior analysis* (pp. 252–286). Pearson UK.
- Cortez, M. D., da Silva Ferreira, M., de Lima, P. M. P., da Silva, M. G. A., & dos Reis, M. D. J. D. (2020). Efeitos de reforçadores de magnitude aumentada e do controle aversivo na manutenção do seguimento de regras em estudantes universitários. *Revista Brasileira de Terapia Comportamental e Cognitiva*, 22. <https://doi.org/10.31505/rbtcc.v22i1.1409>
- Costantini, A. F., & Hoving, K. L. (1973). The effectiveness of reward and punishment contingencies on response inhibition. *Journal of Experimental Child Psychology*, 16(3), 484–494. [https://doi.org/10.1016/0022-0965\(73\)90009-X](https://doi.org/10.1016/0022-0965(73)90009-X)
- Cummins, J., Roche, B., Tyndall, I., & Cartwright, A. (2018). The relationship between differential stimulus relatedness and implicit measure effect sizes. *Journal of the Experimental Analysis of Behavior*, 110(1), 24–38. <https://doi.org/10.1002/jeab.437>
- Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals.* Oxford University Press, USA.
- de Rose, J. C., de Souza, D. G., & Hanna, E. S. (1996). Teaching Reading and Spelling: Exclusion and Stimulus Equivalence. *Journal of Applied Behavior Analysis*, 29(4), 451–469. <https://doi.org/10.1901/jaba.1996.29-451>
- de Rose, J. C., McIlvane, W. J., Dube, W. V., Galpin, V. C., & Stoddard, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior*, 50(1), 1–20. <https://doi.org/10.1901/jeab.1988.50-1>
- de Sousa Borges, S., Durelli, V. H., Reis, H. M., & Isotani, S. (2014). A systematic mapping

- on gamification applied to education. *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, 216–222. <https://doi.org/10.1145/2554850.2554956>
- de Souza, D. G., de Moraes, A. B. A., & Todorov, J. C. (1984). Shock intensity and signaled avoidance responding. *Journal of the Experimental Analysis of Behavior*, 42(1), 67–74. <https://doi.org/10.1901/jeab.1984.42-67>
- de Souza, D. G., de Rose, J. C., Faleiros, T. C., Bortoloti, R., Hanna, E. S., & McIlvane, W. J. (2009). Teaching Generative Reading Via Recombination of Minimal Textual Units: A Legacy of Verbal Behavior to Children in Brazil. *International Journal of Psychology and Psychological Therapy*, 9(1), 19–44.
- de Souza Reis, T., Souza, D. G. de, & de Rose, J. C. (2009). Avaliação de um programa para o ensino de leitura e escrita. *Estudos Em Avaliação Educacional*, 20(44), 425. <https://doi.org/10.18222/ea204420092038>
- 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. <https://doi.org/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. <https://doi.org/10.1901/jeab.1994.62-331>
- Dube, W. V., McIlvane, W. J., Mackay, H. A., & Stoddard, L. T. (1987). Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior*, 47(2), 159–175. <https://doi.org/10.1901/jeab.1987.47-159>
- Dube, W. V., McIlvane, W. J., Maguire, R. W., Mackay, H. A., & Stoddard, L. T. (1989). Stimulus class formation and stimulus-reinforcer relations. *Journal of the Experimental*

- Analysis of Behavior*, 51(1), 65–76. <https://doi.org/10.1901/jeab.1989.51-65>
- Dube, W. V., & Hiris, J. (1997). *Matching to Sample Program (version 11.08)*.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal Of Personality and Social Psychology*, 17(2), 124–129. <https://psycnet.apa.org/doiLanding?doi=10.1037%2Fh0030377>
- Engelmann, A. (1978). Os estados subjetivos: uma tentativa de classificação de seus relatos verbais. In *pesquisa.bvsalud.org*. Ática. <https://pesquisa.bvsalud.org/portal/resource/pt/psa-119095>
- Espirito Santo, H., & Daniel, F. B. (2017). Calcular e apresentar tamanhos do efeito em trabalhos científicos (1): As limitações do  $p < 0,05$  na análise de diferenças de médias de dois grupos. *Revista Portuguesa de Investigação Comportamental e Social*, 1(1), 3–16. <https://doi.org/10.7342/ismt.rpics.2015.1.1.14>
- Fields, L., & Arntzen, E. (2018). Meaningful Stimuli and the Enhancement of Equivalence Class Formation. *Perspectives on Behavior Science*, 41(1), 69–93. <https://doi.org/10.1007/s40614-017-0134-5>
- Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior*, 64(2), 129–145. <https://doi.org/10.1901/jeab.1995.64-129>
- Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior*, 42(1), 143–157. <https://doi.org/10.1901/jeab.1984.42-143>
- Gershoff, E. T. (2017). School corporal punishment in global perspective: prevalence, outcomes, and efforts at intervention. *Psychology, Health & Medicine*, 22(sup1), 224–239. <https://doi.org/10.1080/13548506.2016.1271955>
- Goyos, C., & Almeida, J. C. B. (1994). *Mestre (Versão 1.0)*.

- Goyos, C. (2000). Equivalence class formation via common reinforcers among preschool children. *Psychological Record*, *50*(4), 629–654. <https://doi.org/10.1007/BF03395375>
- Hanna, E. S., Batitucci, L. A. V., J., & Batitucci, J. D. S. L. (2014). Software Contingência Programada: utilidade e funcionalidades. *Revista Brasileira de Análise Do Comportamento*, *10*(1), 97–104.  
<http://periodicos.ufpa.br/index.php/rebac/article/view/3949>
- Harrison, R. J., & Green, G. (1990). Development of conditional and equivalence relations without differential consequences. *Journal of the Experimental Analysis of Behavior*, *54*(3), 225–237. <https://doi.org/10.1901/jeab.1990.54-225>
- Hayes, S. C., Kohlenberg, B., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental Analysis of Behavior*, *56*(1), 119–137.  
<https://doi.org/10.1901/jeab.1991.56-119>
- Hunziker, M. H. L. (2017). Dogmas sobre o controle aversivo. *Acta Comportamentalia*, *25*(1), 85–100. <https://www.redalyc.org/pdf/2745/274550025006.pdf>
- Kapp, K. M. (2013). *The gamification of learning and instruction fieldbook: Ideas into practice*. John Wiley & Sons.
- Klevens, J., Mercer Kollar, L., Rizzo, G., O’Shea, G., Nguyen, J., & Roby, S. (2019). Commonalities and Differences in Social Norms Related to Corporal Punishment Among Black, Latino and White Parents. *Child and Adolescent Social Work Journal*, *36*(1), 19–28. <https://doi.org/10.1007/s10560-018-0591-z>
- Kubaneck, J., Snyder, L. H., & Abrams, R. A. (2015). Reward and punishment act as distinct factors in guiding behavior. *Cognition*, *139*, 154–167.  
<https://doi.org/10.1016/j.cognition.2015.03.005>
- Lerman, D. C., Kelley, M. E., Vorndran, C. M., Kuhn, S. A. C., & LaRue, R. H. (2002).



- Reinforcing magnitude and responding during treatment with differential reinforcement. *Journal of Applied Behavior Analysis*, 35(1), 29–48.  
<https://doi.org/10.1901/jaba.2002.35-29>
- Matos, M. A. (1999). Controle de estímulo condicional, formação de classes conceituais e comportamentos cognitivos. *Revista Brasileira de Terapia Comportamental e Cognitiva*, 1(2), 159–178.
- Meyer, W. J., & Offenbach, S. I. (1962). Effectiveness of reward and punishment as a function of task complexity. *Journal of Comparative and Physiological Psychology*, 55(4), 532–534. <https://doi.org/10.1037/h0049119>
- O'Reilly, A., Roche, B., Gavin, A., Ruiz, M. R., Ryan, A., & Campion, G. (2013). A function acquisition speed test for equivalence relations (FASTER). *Psychological Record*, 63(4), 707–724. <https://doi.org/10.11133/j.tpr.2013.63.4.001>
- O'Reilly, A., Roche, B., Ruiz, M. R., Tyndall, I., & Gavin, A. (2012). The function acquisition speed test (FAST): A behavior analytic implicit test for assessing stimulus relations. *Psychological Record*, 62(3), 507–528. <https://doi.org/10.1007/BF03395817>
- Orwell, G. (2021). *A Revolução dos Bichos*.
- Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*.
- Peirce, J. W. (2007). *PsychoPy — Psychophysics software in Python*. 162(1), 8–13.  
<https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Penney, R. K. (1967). Effect of Reward and Punishment on Children's Orientation and Discrimination Learning. *Journal of Experimental Psychology*, 75(1), 140–142.  
<https://doi.org/10.1037/h0024921>
- Penney, R. K., & Lupton, A. A. (1961). Children's discrimination learning as a function of reward and punishment. *Journal of Comparative and Physiological Psychology*, 54(4), 449–451. <https://doi.org/10.1037/h0045445>

- Perez, W. F., Fidalgo, A. P., Kovac, R., & Nico, Y. C. (2015). The transfer of Cfunc contextual control through equivalence relations. *Journal of the Experimental Analysis of Behavior*, *103*(3), 511–523. <https://doi.org/10.1002/jeab.150>
- Perez, W. F., dos Reis, M. J. D., & de Souza, D. G. (2010). Efeitos de história experimental com diferentes instruções e do controle por contingências sobre o seguimento de instruções. *Acta Comportamentalia*, *18*(1), 55–85.  
<http://revistas.unam.mx/index.php/acom/article/view/18162>
- Perone, M. (2003). Negative effects of positive reinforcement. *The Behavior Analyst*, *26*(1), 1–14. <https://doi.org/10.1007/BF03392064>
- Rabelo, L. Z., Bortoloti, R., & Souza, D. H. (2014). Dolls are for girls and not for boys: Evaluating the appropriateness of the Implicit Relational Assessment Procedure for school-age children. *The Psychological Record*, *64*(1), 71–77.  
<https://link.springer.com/content/pdf/10.1007/s40732-014-0006-2.pdf>
- Rasmussen, E. B., & Newland, M. C. (2008). Asymmetry of Reinforcement and Punishment in Human Choice. *Journal of the Experimental Analysis of Behavior*, *89*(2), 157–167.  
<https://doi.org/10.1901/jeab.2008.89-157>
- Reed, P. (1991). Multiple determinants of the effect of reinforcement magnitude on free-operant response rates. *Journal of the Experimental Analysis of Behavior*, *55*(1), 109–123. <https://doi.org/10.1901/JEAB.1991.55-109>
- Santos, S. L. D., & de Rose, J. C. C. (2018). Investigating the impact of stimulus equivalence on children's food choice and preference. *Trends in Psychology*, *26*(1), 1–14.  
[https://www.scielo.br/scielo.php?pid=S2358-18832018000100001&script=sci\\_arttext](https://www.scielo.br/scielo.php?pid=S2358-18832018000100001&script=sci_arttext)
- Saunders, R. R., & Green, G. (1992). The nonequivalence of behavioral and mathematical equivalence. *Journal of the Experimental Analysis of Behavior*, *57*(2), 227–241.  
<https://doi.org/10.1901/jeab.1992.57-227>

- Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior*, *50*(2), 145–162.  
<https://doi.org/10.1901/jeab.1988.50-145>
- Sidman, M. (1971). Reading and auditory-visual equivalence. *Journal of Speech and Hearing Research*, *14*(1), 5–13.
- Sidman, M. (1987). Two choices are not enough. *Behavior Analysis*, *22*(I), 11–18.
- Sidman, M. (1989). *Coercion and its fallout*. Authors Cooperative.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*.
- 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*(I), 5–22.
- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*(1), 127–146.
- 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., Mackay, H. A., & de Rose, J. C. (2018). Measuring the “transfer of meaning” through members of equivalence classes merged via a class-specific reinforcement procedure. *Learning and Behavior*, *46*(2), 157–170. <https://doi.org/10.3758/s13420-017-0298-6>
- Silveira, M. V., Mackay, H. A., & de Rose, J. C. (2017). Estabelecimento e reorganização de relações arbitrárias derivadas após o treino em matching-to-sample de identidade com consequências específicas. *Revista Brasileira de Análise Do Comportamento*, *13*(2), 1–18. [www.paulekman.com](http://www.paulekman.com)
- Skinner, B. F. (1965). The technology of teaching. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain)*, *162*(989), 427–443. <https://doi.org/10.1098/rspb.1965.0048>
- Spencer, T. J., & Chase, P. N. (1996). SPEED ANALYSES OF STIMULUS EQUIVALENCE. *Journal of the Experimental Analysis of Behavior*, *65*(3), 643–659. <https://doi.org/10.1901/jeab.1996.65-643>
- Toda, A. M., Oliveira, W., Klock, A. C., Palomino, P. T., Pimenta, M., Gasparini, I., Shi, L., Bittencourt, I., Isotani, S., & Cristea, A. I. (2019). A taxonomy of game elements for gamification in educational contexts: Proposal and evaluation. *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019*, 2161-377X(September), 84–88. <https://doi.org/10.1109/ICALT.2019.00028>
- Todorov, J. C. (2001). Quem tem medo de punição? *Revista Brasileira de Terapia Comportamental e Cognitiva*, *3*(1), 37–40.
- Todorov, J. C. (1973). Interaction of frequency and magnitude of reinforcement on concurrent performances. *Journal of the Experimental Analysis of Behavior*, *19*(3), 451–458. <https://doi.org/10.1901/jeab.1973.19-451>
- Todorov, J. C. (2011). Quem tem medo de controle aversivo? *Revista Latina de Análisis de*

*Comportamiento, 19, 5–7.*

Todorov, J. C., Hanna, E. S., & Sá, M. C. N. B. de. (1984). Frequency versus magnitude of reinforcement: new data with a different procedure. *Journal of the Experimental Analysis of Behavior, 41(2)*, 157–167. <https://doi.org/10.1901/jeab.1984.41-157>

*Analysis of Behavior, 41(2)*, 157–167. <https://doi.org/10.1901/jeab.1984.41-157>

Trosclair-Lasserre, N. M., Lerman, D. C., Call, N. A., Addison, L. R., & Kodak, T. (2008).

Reinforcement magnitude: an evaluation of preference and reinforcer efficacy. *Journal*

*of Applied Behavior Analysis, 41(2)*, 203–220. <https://doi.org/10.1901/jaba.2008.41-203>

UFSCar, R. (n.d.). *Portaria GR N°4380 de 20 de março de 2020*. Retrieved February 16,

2021, from <https://www.diariodareitoria.ufscar.br/?p=9004>

Williams, D. C., Saunders, K. J., Saunders, R. R., & Spradlin, J. E. (1995). Unreinforced

Conditional Selection Within Three-Choice Conditional Discriminations. *The*

*Psychological Record, 45(4)*, 613–627.