

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE PSICOLOGIA

Ísis Gomes Vasconcelos

Comportamento moral e responder relacional

São Paulo

2020

ÍISIS GOMES VASCONCELOS

Comportamento moral e responder relacional

Versão original

Tese apresentada ao Instituto de Psicologia da Universidade de São Paulo para obtenção do título de Doutora em Psicologia.

Área de Concentração: Psicologia Experimental

Orientador: Prof. Dr. Marcelo Frota Lobato Benvenuti

São Paulo

2020

Trabalho parcialmente financiado pelo
CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, processo
#142217-2016-6

E realizado no contexto do
Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino
coordenado pela Profa. Deisy das Graças de Souza, processos FAPESP (#2008/57705-8) e
CNPq (#573972-2008-7).

Nome: Vasconcelos, Ísis Gomes
Título: Comportamento moral e responder relacional

Tese apresentada ao Instituto de Psicologia da
Universidade de São Paulo para obtenção do
título de Doutora em Psicologia

Aprovada em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dra. _____

Instituição: _____

Julgamento: _____

Prof. Dra. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Agradecimentos

O ofício de pesquisador é árduo. É um processo ao mesmo tempo criativo e rigoroso. Instigante e desgastante. Demanda de nós a energia para criar o que ainda não existe (a famosa contribuição original), apoiados nos ombros uns dos outros. Ao fim deste trabalho, um percurso de anos, tenho a felicidade de contar com várias pessoas maravilhosas e com oportunidades que me foram dadas e que fizeram este momento possível.

Agradeço primeiramente ao Cnpq, pelo incentivo por meio da bolsa de estudos. Esta bolsa foi condição necessária para a realização deste doutorado, do primeiro ao último dia. Sem ela, este trabalho que aqui apresento e o treinamento que recebi não seriam possíveis. A oportunidade de uma bolsa de estudos pode mudar vidas, abrir caminhos. Foi assim comigo, e por isso sou muito grata. Agradeço também ao departamento de psicologia experimental por oferecer as condições para a realização deste doutorado.

Entre as pessoas que fizeram parte dessa caminhada, preciso tecer agradecimentos especiais ao meu orientador neste trabalho, Marcelo e ao meu orientador do mestrado, João Claudio Todorov. Com vocês eu aprendi (não que este trajeto esteja concluído) o ofício do pesquisador e, em especial, o da condução de experimentos. Desde o trato com o texto, com a escrita, o planejamento de procedimentos, tudo o que faz parte do treino de um pesquisador. Aprendi também o respeito no debate de ideias. Aprendi que divergências teóricas e epistemológicas são terreno fértil para uma produção de conhecimento rigorosa e útil. Como Todorov sempre dizia: "Você pensa desse jeito, então escreva sobre isso!". Marcelo, você é um exemplo de professor e pesquisador. Obrigada por sua atenção e suas críticas, sempre pertinentes.

Agradeço também os professores Paula, Renato, Alexandre, Júlio, Thaís e todos os membros da banca por ter aceitado o convite para apreciar este trabalho e me ajudar a crescer com ele.

O ambiente do doutorado permitiu que eu conhecesse muitas pessoas, pesquisadores competentes que debateram este trabalho comigo e que também se tornaram amigos para a vida. Carla, César, Jéssica e Eduardo, vocês tornaram as aulas e a pesquisa muito mais instigantes e tornaram a vida em São Paulo muito mais feliz.

Família querida, minha mãe Lucineide, meu pai Edmilson, meus irmãos Filipe e Laís, este trabalho também é fruto do apoio incondicional que vocês sempre me deram para sair de Aracaju e procurar a formação em Análise do Comportamento que era meu objeto de desejo. É fruto do amor, carinho e companheirismo de vocês. Sem vocês eu não conseguiria.

Dante, meu querido, as palavras não chegam para expressar meu amor, admiração e gratidão a você. Você fez parte de todo esse percurso como um companheiro cuidadoso, e um leitor interessado e crítico. Pude me beneficiar grandemente de sua formação de pesquisador em filosofia, de seus questionamentos e suas revisões criteriosas sobre meus textos. Obrigada por aliviar a carga deste trabalho e do dia-a-dia. Me sinto muito feliz por compartilhar a vida com você.

“We do not choose survival as a value, it chooses us.”

B. F. Skinner

RESUMO

Vasconcelos, Í. G. (2020). *Comportamento moral e responder relacional* (Tese de Doutorado). Instituto de Psicologia, Universidade de São Paulo, São Paulo.

Este trabalho apresenta uma abordagem experimental da moralidade orientada pela noção de multideterminação do comportamento e o princípio de seleção por consequências. É composto por um trabalho de revisão da produção experimental sobre a moralidade e uma proposta empírica de investigação dos valores morais endereçada em dois experimentos. O estudo experimental da moralidade é dedicado à investigação de variáveis distais, presentes na história da espécie, e variáveis proximais, identificáveis na história de vida do organismo. A partir disso, oferece descrições como os seres humanos classificam as condutas em *certas* ou *erradas* e utilizam essas classificações para pautar o próprio comportamento e produzir consequências para o comportamento de seus pares a fim de coibir o egoísmo e favorecer interesses coletivos. No nível proximal de análise, o estudo dos valores que subordinam as condutas é majoritariamente apoiado numa análise dos usos dos termos valorativos, tendo o comportamento verbal como dado principal. Sustentamos neste trabalho que a diversidade de convenções culturais, dos temas que compõem normas de obrigações, permissões e proibições num grupo, pode ser investigada por meio de procedimentos de responder relacional, tomando como base a literatura sobre comportamento simbólico. A parte empírica deste trabalho propõe o uso do procedimento de avaliação de relações implícitas (*implicit relations assessment procedure* – IRAP) como uma alternativa ao uso de procedimentos de autorrelato para a identificação de valores de justiça e pureza. O dado básico do IRAP é a latência média da resposta a conjuntos de estímulos verbais consistentes ou não entre si. No primeiro estudo, 20 participantes primeiro responderam ao questionário de bases morais (*Moral Foundations Questionnaire* – MFQ) e, em seguida, a um IRAP apresentando os mesmos conteúdos sobre justiça e pureza. Os resultados diferiram a depender do procedimento. No questionário os participantes apresentaram alta frequência de concordância com valores de pureza e alta frequência de discordância com valores de pureza. No IRAP, a frequência de concordância foram aproximada entre os dois valores. O Estudo 2 introduziu uma comparação entre dois grupos pré-experimentalmente diferentes a fim de verificar se os procedimentos seriam capazes de corretamente diferenciar os participantes de um e outro grupo com base em suas respostas a valores de pureza e justiça. A variável pré-experimental verificada foi a presença ou ausência de vivência religiosa do participante. A literatura do MFQ demonstra que participantes religiosos tendem a demonstrar concordância com ambos os valores enquanto participantes não religiosos tendem a demonstrar concordância com valores de justiça e discordância de valores de pureza. Ambos os procedimentos apontaram diferenças estatisticamente significativas no padrão de resposta em função do grupo, mas o desempenho dos participantes foi, assim como no Estudo 1, diferente em função do procedimento. Os resultados do MFQ no Estudo 2 reproduziram o padrão observado na literatura enquanto os resultados do IRAP indicaram ausência de concordância com valores de pureza entre os participantes religiosos. Concluímos apresentando questões futuras para a pesquisa experimental sobre valores morais.

Palavras-chave: Moralidade. Responder relacional. Autorrelato. Valores

ABSTRACT

Vasconcelos, Í. G. (2020). *Moral behavior and relational responding* (Tese de Doutorado). Instituto de Psicologia, Universidade de São Paulo, São Paulo.

The current work presents an experimental approach of morality guided by the notion of multi-determination of behavior and the principle of selection by consequences. It is composed by a theoretical ensemble of the experimental production on morality and an empirical approach on moral values addressed in two experiments. The experimental study of morality is dedicated to the investigation of distal variables, presented in the history of the species and proximal variables, identifiable in the organism's life history. Parting from this, it offers descriptions of how human beings classify conducts in *right* or *wrong* and apply such classifications as a guide to its own behavior and to produce consequences to the behavior of its pairs in order to curb selfishness and favors collective interests. In the proximal level of analysis, the study of the values underlying conducts is mostly addressed in analysis of the uses of evaluative terms having verbal behavior as main data. We sustain that the diversity of cultural conventions, of the themes that composes norms of obligation, permission and prohibition in a group can be adequately described by procedures on relational responding, taking as a foundation the literature on symbolic behavior. The empirical part of this work proposes the use of the implicit relations assessment procedure – IRAP as an alternative to the use self-report procedures as means to the identification of values of fairness and purity. IRAP's basic data is the mean latency of response to sets of verbal stimuli consistent or not among them. In the first study, 20 participants first responded to the moral foundations questionnaire – MFQ and then to an IRAP presenting the same content on fairness and purity. Results differed depending on the procedure. In questionnaire participants indicated high frequency of agreement with values of fairness and high frequency of disagreement with values of purity. In IRAP, agreement biases were similar to both values. Study two introduced a comparison between two pre-experimentally distinct groups aiming to verify if both procedures would correctly differentiate participants from one or another group. The pre-experimental variable verified was the presence or absence of religious experience in participant's history. MFQ literature shows that religious participants tend to demonstrate agreement with both values whereas non religious participants tend to demonstrate agreement with values of fairness and disagreement with values of purity. Both procedures showed statistically significant differences in response patterns depending on the group, but participant's performance, as in study 1, differed between procedures. Results from MFQ reproduced the pattern observed in literature while results from IRAP indicated absence of agreement on purity values among religious participants. We conclude by presenting further research issues on the experimental research on moral values.

Keywords: Morality. Relational Responding. Self-report. Values.

LISTA DE TABELAS

Tabela 1 - Trial types presented in practice phase.....	67
Tabela 2 - Trial types presented in test phase.....	69
Tabela 3 - Method to convert mean latencies from each participant in D – IRAP scores.....	71

Lista de Figuras

Figura 1 - Group means of agreement for each of the five foundations in MFQ.....	72
Figura 2 - Individual mean latencies for statements on fairness.....	74
Figura 3 - Individual mean latencies for statements on purity.....	76
Figura 4 - D – IRAP scores for each one of the trial types.....	77
Figura 5 - Frequency of agreement within the six statements in the questionnaire (explicit) and IRAP (implicit).....	80
Figura 6 - Média de concordância para cada uma das cinco bases morais por grupo.....	98
Figura 7 - Latências médias individuais nos blocos consistentes e inconsistentes para cada uma das três afirmações sobre justiça.....	101
Figura 8 - Latências médias individuais nos blocos consistentes e inconsistentes para cada uma das três afirmações sobre pureza.....	103

SUMÁRIO

INTRODUÇÃO GERAL.....	14
CAPÍTULO 1: Moral behavior: A theoretical and empirical bond for culturo-behavior sciences.....	28
Morality.....	31
Stimulus Relations and Relational Responding.....	34
Decision-Making in Moral Situations.....	37
Evolutionary Aspects of Morality.....	41
References.....	48
CAPÍTULO 2: An empirical investigation on moral values: explicit and implicit values of fairness and purity.....	57
METHOD.....	65
Participants.....	65
Aparatus, instruments e material.....	65
Procedure.....	66
Data analysis.....	70
Results.....	72
Discussion.....	80
References.....	83
CAPÍTULO 3: Religiosidade como preditora de valores morais de justiça e pureza em medidas explícitas e implícitas.....	89
MÉTODO.....	94
Participantes.....	94
Aparato, material e procedimentos.....	96
Análise dos dados.....	96
Resultados.....	97
Discussão.....	104
DISCUSSÃO GERAL.....	106
REFERÊNCIAS.....	112
APÊNDICES.....	124

Introdução Geral

Em *Beyond Freedom and Dignity*, Skinner (1971) destaca a questão dos usos da ciência do comportamento. A compreensão de que o comportamento é produto da interação do organismo com contingências ambientais, em oposição à explicação do agente autônomo como causa do comportamento, levantaria imediatamente questões da ordem dos valores morais: que tipos de contingências seriam ideais? E para quais finalidades? Se os efeitos de alguns comportamentos são melhores do que outros, de onde veem os critérios para tal classificação? O estabelecimento de permissões, obrigações e deveres estaria fora do escopo da ciência, que não deveria oferecer juízos de valor sobre o comportamento. Mas a construção de juízos de valor moral é também comportamento e, dessa forma, é um objeto para investigação em uma ciência do comportamento.

Moral, do latim *mores*, *moralis* ou *mos* significa costume (Japiassu & Marcondes, 1996). A moral enquanto disciplina nasce na filosofia e corresponde ao estudo ou ciência da conduta, dos costumes observados em um grupo, em especial, seus valores sobre *bem e mal*, *correto e incorreto*, virtudes, obrigações, responsabilidades e proibições socialmente estabelecidas (Abbagnano, 2007; Brugger, 1977; Chiesa, 2003; Ferrater-Mora 1982, 2001). Autores como Japiassu e Marcondes (1996) e Lalande (1999) propõem a distinção entre o estudo da moral e o da ética separando o estudo da conduta em duas vertentes: a descrição dos costumes e seus valores (bons ou maus) em uma dada época e lugar corresponde ao estudo moral; os juízos de apreciação dos atos em si corresponderia ao estudo da ética.

Outra distinção presente nos estudos sobre a apreciação de condutas seria o seu caráter normativo/prescritivo ou descritivo (Abib, 2002; Japiassu & Marcondes, 1996). Os estudos normativos/prescritivos se dedicam ao estabelecimento e justificação de hierarquias de valores cuja adoção seria desejável, oferecendo um guia para o comportamento (Chiesa, 2003; Dittrich, 2010). Os estudos descritivos, também chamados de metaéticos, estariam, por sua

vez, dedicados à análise da origem dos comportamentos morais e dos valores que os subjazem. Estudos descritivos são aqueles que se dedicam a identificar de que forma é feito o uso de termos valorativos como *bom* ou *certo* para classificar eventos, em outras palavras, que variáveis controlam os comportamentos de empregar esses termos (Chiesa, 2003; Dittrich, 2010; Pimentel-Pequeno, 2003). Dessa forma, o objeto de análise na perspectiva descritiva é fundamentalmente a linguagem.

Até aqui temos que a disciplina da moral se dedica ao estudo do comportamento de valoração de condutas. O presente trabalho propõe uma incursão ao tema da moralidade por meio do referencial conceitual e empírico da Análise do Comportamento. Uma proposta de leitura comportamental se apoia em achados experimentais e em modelos como o de seleção por consequências (Glenn & Malagodi, 1991). Utilizando distinção entre conteúdo e processo proposta por Glenn e Malagodi (1991), o comportamento é o conteúdo que se busca explicar. A análise do comportamento oferece explicações sobre conteúdos por meio da descrição dos processos que os subjazem. Dessa forma, uma ciência do comportamento pode investigar e ajudar a debater as ferramentas conceituais e empíricas para oferecer descrições sobre o comportamento moral.

Skinner (1971) oferece uma análise de contingências envolvidas nos usos de termos valorativos. Nesta obra, Skinner relaciona o comportamento verbal sobre o uso dos termos *bom* e *ruim* a uma análise do efeito das consequências do comportamento. Dessa forma, ele adiciona mais um elemento à análise da moralidade: para entender como as pessoas classificam o comportamento próprio e dos outros, é preciso conhecer os efeitos do comportamento a ser classificado.

Juízos de valor podem ser feitos a respeito de qualquer comportamento. Ao falar sobre moralidade, entretanto, nos deteremos nos comportamentos que recebem juízos de valor por se conformarem a convenções sociais, fazendo assim uma distinção entre preferências

peçoais e práticas culturais (Hocutt, 2013). Esta delimitação nos leva a aproximar o conceito de comportamento moral ao campo das ciências culturo-comportamentais (Cihon & Mattaini, 2019). Para tal, recorreremos às ferramentas conceituais disponíveis para a abordagem de questões culturais.

O termo prática cultural refere-se, na análise do comportamento, a comportamentos que são aprendidos no contato com membros de um grupo, repetidos por diferentes indivíduos e mantidos ao longo de gerações, sobrevivendo à troca dos membros desse grupo (Glenn et al, 2016; Sampaio & Andery, 2010).

Valores que subsidiam a propagação de algumas práticas e a proibição de outras seriam encontrados no *milieu* cultural. O termo proposto por Houmanfar & Rodrigues (2006) designa o conjunto de crenças, recursos materiais, políticas governamentais, regras, tradições, instituições e tecnologias e compreende a variável ambiental numa contingência.

Tradicionalmente, o comportamento verbal é apontado como o principal vetor por meio do qual o *milieu* cultural evoca comportamentos (Glenn, 1988, 1989; Houmanfar & Rodrigues, 2006). Entretanto, como dado de análise, o conteúdo verbal é bastante frágil: o controle do comportamento verbal está na comunidade verbal e não nos eventos aos quais ele se refere (Skinner, 1945, 1957). Adicionalmente, diante de temas polêmicos e controversos, a resposta verbal é altamente suscetível a distorções como a “desejabilidade social” (Paunonen & LeBel, 2012; Power, Barnes-Holmes, Barnes-Holmes & Stewart, 2009; Ziegler & Buehner, 2009). Desse modo, simplesmente perguntar às pessoas sobre seus valores morais não parece ser um modo válido e fidedigno de se conhecer sobre eles.

Aqui, vamos sugerir uma abordagem aos valores morais que envolve descrever a relação entre comportamento verbal e não verbal. No controle não verbal, destacamos a importância da noção de controle e classe de estímulos, que podem ser investigados empiricamente por estratégias experimentais típicas da análise experimental do

comportamento. Conforme aponta Sidman (1993) *“The treatment of language symbols as if they are the things they represent is so strongly and deeply ingrained that we often react to words with feelings of pain, and we use words to inflict pain”* (p.04). Essa passagem evidencia como a linguagem pode extrapolar o comportamento verbal e o estudo do comportamento simbólico pode ser um meio de investigar empiricamente fenômenos que, embora relacionados ao verbal, não se limitam a ele. Para Barnes e Holmes (1991) e de Rose (1993), pensamento e linguagem são repertórios complexos baseados na capacidade de agrupar estímulos em classes. A formação de classes de estímulos ocorre por meio do estabelecimento de relações entre os mesmos.

Considere o exemplo do racismo. Nas práticas verbais, é comum o discurso de que o racismo é errado e criminoso e deve ser rechaçado. Nas práticas não verbais, todavia, os dados mostram uma manutenção do perfil depreciativo atrelado ao negro, que se materializa em fatos como seletividade da truculência policial; desigualdade salarial; porcentagem da população carcerária que ainda aguarda julgamento (Flores, 2016; Gomes & Laborne, 2018; Santos, 2019), apenas para citar alguns exemplos. Como relacionar esses eventos à noção de racismo? Barnes-Holmes, Murphy, Barnes-Holmes & Stewart (2010) demonstraram que pessoas declaradamente antirracistas apresentavam um padrão de respostas que relacionava a imagem de homens negros armados à noção de perigo e a imagem de homens brancos armados à noção de segurança. Esse dado contraria o discurso pessoal dos participantes e corrobora o padrão hegemônico do racismo de relacionar o negro à violência.

Uma questão metodológica que se coloca é como acessar relações entre estímulos que expressem valores morais. Esse é um dos desafios principais do presente trabalho. Apresentamos aqui três procedimentos experimentais para o estudo de relações entre estímulos que podem contribuir com o estudo de práticas culturais e valores morais: o procedimento de formação de classes de equivalência de estímulos, o teste de velocidade da

aquisição de função (*Function Acquisition Speed Test* – FAST) e o procedimento de avaliação de relações implícitas (*Implicit Relational Assessment Procedure* – IRAP).

Classes de equivalência podem ser definidas como a emergência de relações condicionais não treinadas AC, CA, BA e CB, exibindo as propriedades de simetria, transitividade e reflexividade, a partir do ensino das relações condicionais AB e BC (Sidman & Tailby, 1982). O treino é comumente realizado por meio de um procedimento de emparelhamento com o modelo (*matching to sample* – MTS). As relações entre os estímulos são denominadas arbitrárias pois não envolvem semelhança física entre eles (Terrace, 1966) nem o estabelecimento da relação entre os estímulos por meio de uma resposta comum (de Rose, 1993; O'Reilly et al, 2012).

Os estudos que investigam relações de equivalência para a identificação de comportamentos relacionados a práticas culturais têm utilizado conteúdos socialmente controversos que são identificáveis no grupo social do qual o participante faz parte. No procedimento, esses estímulos conhecidos e que possuem relação entre si no *milieu* cultural são utilizados em procedimentos de formação de classes de equivalência conflitantes. Ou seja, as relações que se deseja estabelecer são incompatíveis com relações entre estímulos culturalmente estabelecidas (de Carvalho & de Rose, 2014; Dixon et al, 2006; Mizael et al, 2016; e Watt et al, 1991).

Em um estudo seminal sobre a emergência de novas relações entre estímulos culturalmente estabelecidos, Watt et al (1991) demonstraram as relações entre religião e nomes de pessoas em função do papel social dos participantes. No procedimento, participantes norte irlandeses católicos, norte irlandeses protestantes e ingleses protestantes foram expostos a um treino de MTS no qual foram ensinadas as relações AB (nomes com sobrenomes católicos/sílabas sem sentido) e BC (sílabas sem sentido/símbolos protestantes). Durante os testes para observar possíveis relações emergentes entre AC e CA, os ingleses

corretamente escolheram símbolos protestantes diante dos nomes católicos. Já entre os participantes norte irlandeses, 12 dos 19 escolhiam um novo estímulo (nome e sobrenome protestante) diante do símbolo protestante, ou seja, não demonstravam a relação emergente predita entre nomes católicos e símbolos protestantes. Os achados sugerem que a não emergência das relações esperadas decorreu da solicitação para que os participantes norte irlandeses respondessem de acordo com relações que contrariavam sua história em sociedade, ou seja, suas relações entre estímulos estabelecidas pré-experimentalmente.

De Carvalho e de Rose (2014) propuseram um treino de relações condicionais entre símbolos positivos, um símbolo abstrato e fotos de homens negros a crianças que apresentavam, na linha de base, um viés negativo para as fotos de homens negros. Três meninos e uma menina foram selecionados para o estudo. A tarefa experimental envolveu pré-teste, treino e pós-teste em emparelhamento com o modelo com atraso (*delayed matching to sample-DMTS*). Os estímulos utilizados foram fotos de homens negros ou brancos (conjunto C), fotos de símbolos abstratos (conjunto B) e figuras de símbolos positivos e negativos (conjunto A). O treino visou a estabelecer as relações AB e BC.

A partir disso, a predição é a de que a relação de equivalência entre fotos de homens negros e símbolos positivos poderia emergir. Os resultados mostraram que apenas um participante apresentou o padrão predito, exibindo reversão da relação homem negro/símbolo negativo. Dois participantes não apresentaram mudanças nessas relações e um deles apresentou uma frequência maior de relações homem negro/símbolo negativo no pós-teste do que no pré-teste.

Mizael et al (2016) realizaram uma replicação sistemática do estudo de de Carvalho e de Rose (2014). Treze crianças com linha de base com viés negativo para fotos de homens negros participaram do estudo. As etapas de treino e teste para possíveis reversões nas relações de equivalência foram idênticas às do procedimento de de Carvalho e de Rose

(2014). Além do treino, no pré-treino as crianças responderam a uma escala de prazer tendo como estímulos as fotos de homens negros e brancos utilizadas no treino e, no pós-teste, as crianças responderam a um IRAP que apresentava como estímulos rótulo as fotos de homens negros utilizadas no MTS e como estímulos-alvo fotos de cenas agradáveis ou de cenas desagradáveis. Além disso, metade dos participantes passou pela fase de treino com o emparelhamento com o modelo simultâneo (*simultaneous matching to sample* – SMTS) e a outra metade respondeu DMTS. Os resultados mostram uma mudança de atitude representada pela redução de frequência de escolhas do símbolo negativo diante da foto de homens negros comparativamente com os dados da linha de base. Esses resultados são reforçados pelo desempenho no IRAP que mostra a ausência de viés negativo diante das fotos de homens negros.

Em conjunto, os achados sugerem que a formação de relações entre estímulos durante a história pessoal culturalmente estabelecida da pessoa pode interferir na formação e emergência de novas relações entre estímulos. Além disso, testes de equivalência parecem ser procedimentos úteis para verificar a existência de relações entre estímulos previamente estabelecidas em contextos socialmente sensíveis (O'Reilly et al, 2012) bem como uma possibilidade instrumental para modificar essas relações (Mizael et al, 2016).

O'Reilly et al (2012) discutem que, a despeito da eficácia do modelo das relações de equivalência para o estudo de temas socialmente controversos, esta metodologia é pouco utilizada com essa finalidade. Para eles, isso acontece porque o uso do paradigma das relações emergentes demanda o conhecimento sobre o fenômeno da emergência de classes de equivalência e do treino de relações condicionais, conteúdos pouco discutidos entre não analistas do comportamento. Ainda de acordo com os autores, haveria limitações de ordem procedimental. Treinos e testes de equivalência normalmente são procedimentos demorados quando comparados com os testes implícitos mais populares e requerem muita atenção e

motivação por parte do participante para completar o procedimento. Além disso, o procedimento de formação de classes de equivalência também não seria o modelo mais indicado para o estudo de conteúdos socialmente controversos por carecer de *Procedural implicitness*. Na fase de teste para a verificação da emergência de relações, os dois estímulos de interesse são apresentados simultaneamente como modelo e comparação, o que torna as variáveis sob análise discrimináveis para o participante.

O segundo procedimento que possibilita o estudo de relações entre estímulos é o FAST. Ele foi idealizado por Roche e Ruiz (O'Reilly et al, 2012) e surge como uma proposta de procedimento experimental para o estudo da formação de relações implícitas entre estímulos que responde aos problemas usualmente apontados em procedimentos baseados na associação entre estímulos ao propor uma tarefa de discriminação simples.

A proposta do FAST (O'Reilly et al, 2012) é realizar um teste que verifique a emergência de classes funcionais em vez de classes de equivalência de forma que o procedimento não requeira respostas em discriminação condicional tal como no emparelhamento com o modelo. Para isso, os estímulos devem ser apresentados em tentativas separadas e funções de resposta diferentes precisam ser estabelecidas para cada classe. A taxa em que o participante aprende a produzir a resposta comum a ambos os estímulos apresentados separadamente pode ser comparada à taxa de aprendizagem para produzir respostas diferentes para cada um dos estímulos em tentativas separadas.

Para verificar a viabilidade empírica da proposta, O'Reilly et al (2012) conduziram o procedimento FAST para avaliar a aquisição e a força de relações entre estímulos criadas em laboratório. O experimento contou com 18 participantes e consistiu em 4 fases. Na primeira foi realizado o MTS no qual foram reforçadas diferencialmente as relações A1-B1, B1-A1, A2-B2 e B2-A2. As fases 2 e 4 consistiram cada uma na exposição a um único bloco de treino a fim de estabelecer uma linha de base para o nível de aquisições de classes utilizando

estímulos novos e não relacionados. Essas linhas de base serviram de comparação para a velocidade de aquisição na fase 3. A fase 3 foi dividida em dois blocos, um deles apresentando relações entre estímulos tal como treinadas na fase 1 (teste de bloco consistente) e outro bloco apresentando relações que combinavam estímulos de classes diferentes (teste de bloco inconsistente).

Treze participantes mostraram aquisição de função mais rápida no bloco consistente. Com isso, os autores sugerem que o procedimento pode ser utilizado como uma metodologia comportamental para acessar a aquisição e a força de relações previamente estabelecidas entre estímulos. Além disso, o procedimento poderia servir para diferenciar repertórios de responder relacional que estejam bem estabelecidos ou em processo de emergência.

O'Reilly et al (2013) aplicaram o procedimento FAST como forma de testar a força da emergência de relações de equivalência treinadas experimentalmente. Após treinar as relações A1-B1, A1-C1, A2-B2 e A2-C2 e verificar a emergência de B1-C1 e B2-C2, os participantes passaram por um treino no qual uma resposta foi treinada diante da apresentação de B1-C1 (condição de resposta consistente) e outra resposta foi estabelecida diante da apresentação de B1-C2 (condição de resposta inconsistente). Os resultados mostram que a aquisição da resposta para a relação consistente com a relação emergente foi mais rápida.

O'Reilly et al (2012, 2013) enfatizam que o FAST representa um avanço procedimental no que diz respeito ao *procedural implicitness* e também poderia ser utilizado para a verificação de relações pré-experimentais. Entretanto, FAST permite acessar apenas a velocidade de aquisição das relações entre estímulos, mas não informa sobre o tipo de relação (substituibilidade, oposição e etc).

O terceiro procedimento para a identificação de relações entre estímulos é o IRAP. A tarefa do IRAP consiste em apresentar conjuntos de estímulos aos participantes em uma tela de computador. Os estímulos apresentados no topo da tela são os estímulos rótulo que vão

modificar a função do estímulo alvo. Usualmente as tentativas alternam dois estímulos rótulo que representam categorias distintas (certo/errado; agradável/desagradável; branco/negro; mulher/homem). Os estímulos alvo são apresentados no centro da tela, e podem ser palavras ou imagens. É a relação entre os estímulos rótulo e alvo que se deseja observar. Nos cantos inferiores da tela são apresentados os termos relacionais. Eles recebem este nome pois a escolha por um deles deve expressar a relação entre os estímulos rótulo e alvo. Eles são apresentados em valências opostas (verdadeiro/falso; igual/oposto).

A resposta solicitada do participante é escolher o termo relacional que melhor expresse a relação entre estímulo rótulo e estímulo alvo a partir de dois tipos de regra apresentadas blocos de tentativa distintos. Nos blocos consistentes o participante deve escolher o termo relacional que expresse a relação típica, socialmente estabelecida entre os estímulos rótulo e alvo. Nos blocos inconsistentes o participante é exposto aos mesmos pares de estímulos vistos antes, mas ele deve escolher o termo relacional contrário à relação típica.

Além disso, os participantes são solicitados a responder a esses blocos da forma mais acurada e o mais rapidamente possível. A pressão temporal tem a função de reduzir a “contaminação” da resposta por outros estímulos contextuais.

Estudos que utilizam a latência como medida da relação entre estímulos sugerem que a latência da resposta será mais curta em tentativas cuja relação entre estímulos seja do tipo consistente produzindo assim um padrão diferencial de latências entre as respostas consistentes e as inconsistentes, o chamado “efeito IRAP” (Power et al, 2009). Dessa forma, o IRAP compara a facilidade relativa, indicada por médias de tempos de reação, com que o participante responde aos pares de estímulos (Hussey, Thompson, McEnterggart, Barnes-Holmes & Barnes-Holmes, 2015).

O papel da latência como crítico para a identificação do responder implícito relacional tem sido fortalecido por alguns achados experimentais. Quando participantes são solicitados a

responder a medidas explícitas (autorrelatos) e implícitas (responder com pressão temporal) das mesmas relações, os resultados são diferentes (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010; Drake et al, 2010; Power et al, 2009). Além disso, em procedimentos nos quais os participantes foram orientados sobre a natureza do IRAP e a tentar fingir seu desempenho, ainda assim o efeito IRAP foi produzido de acordo com as relações típicas esperadas (Mc Kenna, Barnes-Holmes, Barnes-Holmes & Stewart, 2007).

Pela forma como o IRAP é organizado, considera-se que os dados obtidos por meio dele são avaliações automáticas. Já no caso de medidas de autorrelato, o que se tem são reações mais cuidadosamente elaboradas (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010). Assim, quando a medida implícita e a de autorrelato apontam resultados diferentes, argumenta-se que, no autorrelato, pela ausência de pressão temporal, o participante “rejeita” sua resposta relacional mais imediata e responde de acordo com relações mais elaboradas e influenciadas por variáveis contextuais (Barnes-Holmes et al, 2006; Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010; e Drake et al, 2010). Os achados experimentais têm sistematicamente demonstrado o efeito IRAP e as diferenças entre essa medida e procedimentos baseados em autorrelato (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart 2010; Dawson, Barnes-Holmes, Gresswell, Hart & Gore, 2009; Power et al, 2009, apenas para citar alguns).

A fim de testar a confiabilidade do IRAP diante de expectativas de “desejabilidade social” do participante, Barnes-Holmes, Murphy, Barnes-Holmes e Stewart (2010) verificaram se a variação do contexto no qual o IRAP é aplicado (saber que suas respostas serão públicas *versus* ser informado de que suas respostas não serão analisadas individualmente) influenciaria na obtenção do efeito IRAP sobre vieses raciais.

No experimento 2, 31 participantes (16 mulheres) adultos (média de idade de 29 anos) todos brancos e de cidadania irlandesa foram randomicamente designados para uma de duas

condições: contexto público e contexto privado. Os participantes da condição pública foram informados que após sua participação na pesquisa, o computador calcularia o seu nível de preconceito contra negros e os resultados seriam avaliados a partir de alguns *surveys*, além disso, o pesquisador permaneceria na sala durante toda a atividade. Na condição de contexto privado o participante era informado de que os dados produzidos por ele seriam calculados numa média com os dados de todos os outros participantes, que ninguém nunca saberia quais respostas ele deu e que o pesquisador ficaria fora da sala durante todo o procedimento.

Todos os participantes responderam a um IRAP e a três medidas de autorrelato. No IRAP os participantes foram expostos a 8 blocos de 24 tentativas sendo 4 do tipo resposta consistente e 4 do tipo resposta inconsistente. Os estímulos rótulo foram os termos “seguro” ou “perigoso”, os estímulo alvo foram fotos de homens negros e brancos segurando armas de fogo. Nos cantos inferiores da tela estavam os termos relacionais *verdadeiro* e *falso* acionados respectivamente pelas teclas D e K. Os participantes foram instruídos a responder em um intervalo de latência de até 2 segundos. Nas condições de resposta consistente se esperava que os participantes apresentassem menores latências diante dos pares: seguro-branco-verdadeiro; seguro-negro-falso; perigoso-branco-falso; e perigoso-negro-verdadeiro. Nas condições de resposta inconsistente se esperava que as latências fossem mais longas diante dos mesmos pares de estímulos mas com termos relacionais trocados.

Após o IRAP, os participantes responderam às escalas de autorrelato: uma escala sobre discriminação racial e diversidade, outra escala sobre racismo moderno e uma escala likert elaborada com os estímulos do IRAP. Os resultados mostraram que, nas duas condições, as três escalas de autorrelato mostravam um viés positivo por negros. O efeito IRAP, por sua vez indicou um viés pró-branco e antinegro.

O efeito IRAP também parece ser sensível a histórias de reforçamento produzidas em laboratório como foi descrito anteriormente no estudo de Mizael et al (2016). Além desse

estudo, Bortoloti e De Rose (2012) demonstraram que até relações nunca treinadas, como as relações emergentes estabelecidas experimentalmente por classes de equivalência também produzem o efeito IRAP.

Neste estudo, 19 participantes foram designados para duas condições de treino de relações arbitrárias: SMTS ou DMTS. Todos os participantes passaram por treino das relações condicionais AB, AC e CD. Os estímulos do conjunto A eram figuras de faces expressando felicidade, raiva ou expressão neutra e os estímulos B, C e D eram sequências de quatro letras sem sentido. Os participantes que atingiram o critério de acerto no teste de emergência de relações foram expostos a um teste adicional com um IRAP. No teste com o IRAP, os estímulos da classe D foram apresentados como rótulos (no topo da tela) e as figuras da classe de estímulo A foram apresentadas como estímulo-alvo.

Os resultados mostraram que os participantes treinados com o procedimento de DMTS respondiam mais rápido em tentativas consistentes do que nas inconsistentes caracterizando assim o efeito IRAP com diferença estatisticamente significativa. Não houve diferença estatisticamente significativa entre as latências de resposta dos participantes treinados no SMTS. Os achados foram interpretados como demonstrativos de que treinos em DMTS produzem relações entre estímulos mais fortes do que aquelas produzidas via treino por SMTS.

O IRAP tem se mostrado um procedimento bem sucedido em lidar com temas socialmente controversos ou sobre questões clínicas diante das quais o relato verbal é pouco acurado como medida; e um preditor de condutas não verbais mais eficiente do que instrumentos de autorrelato (Carpenter, Martinez, Vadhan, Barnes-Holmes & Nunes, 2012; Hooper, Villatte, Neofotistou & Mc Hugh, 2010; Leech, Barnes-Holmes & Madden, 2016; Perez et al, 2017; Rönspies, Schmidt, Melnikova, Krumova, Zolfagari & Banse, 2015; Vahey, Barnes-Holmes, Barnes-Holmes & Stewart, 2009).

Além disso, parece pouco suscetível à influência da “desejabilidade social” e tentativas de falseamento de resposta (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010; McKenna et al, 2007). Dessa forma, mesmo também carecendo do *procedural implicitness*, essa característica do procedimento não parece afetar os dados. Neste trabalho, o IRAP foi o procedimento escolhido como ferramenta experimental para o estudo dos valores morais.

A estrutura do presente trabalho foi organizada em forma de capítulos que constituem investigações independentes, embora relacionadas. O Capítulo 1 é dedicado ao exame do referencial experimental para o estudo da moralidade. Ele apresenta de forma resumida três linhas de pesquisa experimental para o estudo dos valores morais: pesquisas sobre controle de estímulos, pesquisas sobre tomada de decisão e processamento da informação e pesquisas sobre a evolução da moralidade. O capítulo destaca as principais contribuições em cada uma dessas linhas para a compreensão empírica da moralidade e propõe alguns pontos de discussão para uma intersecção entre esses achados a fim de explicar fenômenos culturais. O Capítulo 2 apresenta um estudo experimental de valores morais a partir do IRAP tendo como conteúdo os elementos da Teoria das Bases Morais. O Capítulo 3 aprofunda a investigação experimental dos valores morais por meio do IRAP introduzindo a religiosidade como uma variável preditora de padrões morais.

Capítulo 1

Moral behavior: A theoretical and empirical bond for culturo-behavior sciences¹

¹ Capítulo no formato original em que foi submetido como artigo.

Abstract

Morality can be functionally defined as patterns of classifying behavior as *right* or *wrong*, behaving accordingly, and to provide consequences to the behavior of others based on those classifications. Considering the widespread presence of such patterns in social contingencies, they may be of interest in the exam on how individual behavior is related to social behavior and cultural practices. This paper briefly presents findings from three fields of experimental research that have investigated the moral phenomenon and proposes the culturo-behavior sciences as the ideal context to integrate different causal conceptions about morality. The first group of findings encompasses evidence produced from stimulus relation procedures (equivalence relations and the implicit relational assessment procedure) showing how relational responses expresses social practices such as intolerance, racism and prejudice and is an alternative to verbal data. In the second group, we describe investigations that focus on decision-making processes in moral dilemmas. This set of findings highlights the context-dependency in moral decisions. The third group provides comprehensive discussions on the evolutionary roots of morality. Evidence from these diverse findings support the idea that morality is influenced by multiple causes including proximal and distal causes. The notions of social behavior as interlocking interactions and cultural practices as evolved systems may address the confluence of these causes.

Keywords: Morality. Stimulus relations. Moral decision making. Distal causes. Proximal causes. Experimental research. Culturo-behavioral science.

Behavior is highly influenced by social contingencies. Although recognition about the importance of social and cultural context was never explicitly neglect, traditional research on learning was developed by studying the behavior of single organisms in a strict controlled environment (e.g, Skinner, 1938). Social variables were only gradually incorporated as definitive aspects of procedures to understand basic elements of behavior interactions. How and why people use reinforcement, punishment or create contextual cues to deal with other people's behavior is an essential aspect of social and cultural life.

Recent theoretical effort to understand the relationship between behavior processes that describe individual behavior and cultural practices have as an effect, the need for a refinement of known terms and production of new ones. This stream of concepts encompasses terms as group control, controlling agencies, cultural practices, interlocking behavior contingencies, metacontingencies, macrocontingencies to mention a few, and also a specific behavioral branch of research, the culturo-behavioral sciences (Cihon & Mattaini, 2019).

As a complex net of interlocking contingencies, social behavior demands that people reinforce, punish and provide contextual control to other's people behavior. This is made usually in a consistent way, following either explicit (spoken) or implicit (symbolic but not always described) logic. Based on these assumptions about social behavior and cultural practices and in the body of experimental research on the topic of morality, we claim in this paper that (a) the wide notion of social behavior and the stream of concepts that comprehends the study of social and cultural phenomena are bind by the notion of morality. (b) morality as an interdisciplinary field address distal and proximal causes of behavior and moral behavior can only be fully explained by the comprehension of such overlapping. (c) a real understanding of social behavior and interlocking behavior contingencies would be achieved by an investigative effort of the explicit and implicit logic underlying social behavior and cultural practices.

In order to sustain such claims, we will first provide some initial definitions of morality and then explore three fields of investigative efforts about morality based on different epistemologies. The first group of findings encompasses evidence produced from stimulus relation procedures (equivalence relations and the implicit relational assessment procedure) used to capture symbolic nets on practices such as intolerance, racism and prejudice. In the second group, we describe investigations that focus on decision-making processes in moral dilemmas. Such processes are described in terms of neuropsychological dispositions affected by immediate features of the context of decision. The third group provides comprehensive discussions on the evolutionary roots of morality. Evidence from these diverse findings support the idea that morality is influenced by multiple causes as proximal and distal causes and may be a helpful tool to provide some insights to current debates on interlocking behavior contingencies and cultural practices.

Morality

A typical element of human coexistence is the tension between individual interests and collective well-being. How humans mitigate this tension is usually examined from the perspective of morality (Skinner, 1971).

In Behavior Analysis, the subject of morality or ethics has been examined from a theoretical view on the basis of the works of Skinner (1953, 1961, 1971, 1981). Skinner's analysts on moral behavior have discussed either the meta-ethical issues embedded in a behavioristic approach of morality or the so called Skinnerian ethical system present in the discussion on cultural evolution. Here we focused on the first approach, the behavioral meta-ethics, since our goal is to propose an empirical route to a behavioral analysis of morality.

This first branch has explored Skinner's model of identifying moral values from the identification of the effect of consequences (Skinner, 1971). Skinner (1971) stated that *good* is

everything that produces positive reinforcement and remove aversive stimulation. Following this reasoning, *bad* is everything that produces or maintains aversive stimulation. From this, authors dedicated to a behavioral analysis of moral values have presented analyses of the uses of the qualifying terms *good*, *bad*, *right* and *wrong* (Dittrich, 2004; Dittrich, 2010; Garret, 1979; Graham, 1977; Hocutt, 1977; Hocutt, 2013; Leigland, 2005; Rottschaefer, 1980; Ruiz, 2013; Waller, 1982). Some of them propose the discussion about learning contingencies for moral behavior (Dittrich, 2004; Hayes, Gifford & Hayes, 1998).

These authors use the terms ethics or moral sometimes interchangeably and with no differentiation between individual and social contingencies. That is, any behavior can be considered good or bad, so any behavior is ethical behavior. Here, however, we will assume Hocutt's (1977, 2013) view in isolating moral behavior as behaviors that are socially installed and maintained. The author proposes a distinction among the uses of the terms *good/bad* and *right/wrong* stating that the first pair is usually applied to qualify individual preferences, while the second pair is applied to designate behavior socially accepted. For Hocutt, only the second type may constitute moral behavior. Contingencies for moral behavior may be informal are enforced by group members constituting social practices. Considering these brief presentation, we will refer to values deem personal preferences as ethical behavior and values deem to informal social enforced contingencies as moral behavior.

There is also a third class of qualifying terms, that terms stated by controlling agencies, where intentional hierarchical sets of rules are established via sanctioning contingencies. This class encompasses terms as *legal/illegal*, *virtuous/sinful* and so on, depending on the agency. This distinction in three classes of evaluative terms finds foundation in Skinner's distinctions among individual contingency, group control and controlling agencies respectively (Skinner, 1953).

The analysis of the use of evaluative terms is useful to the description of social contingencies concerning moral issues. Yet, as an empirical data, verbal descriptions are highly sensitive to contextual influence that weakens its predictive value on non-verbal behavior. In this work, we propose an alternative basic data to verbal reports, the relational responding, that have been proving itself useful specially in handling socially controversial topics.

Based on these initial definitions, we will restrain moral behavior to a functional set of repertoires of classifying behaviors as *right* or *wrong*, as means to adjust its own behavior and to present differing consequences to the behavior of others (Hocutt, 2013). Interest in morality issues has mobilized research across many disciplines, from the earliest stages of knowledge until now. As a result, knowledge about morality is related to a variety of epistemologies, even inside the perspective of an experimental psychology.

An attempt to integrate different epistemological traditions on the study of morality can be to understand the role of different "causes" regarding the psychological phenomenon. Researchers as Tinbergen (1963) and Killeen (2001), for example, suggests the classical four Aristotelian causes as a model for understanding the myriad of theoretical epistemologies in psychology. Aristotle proposes that the causes of any phenomenon can be separated into four types: material, final, efficient, and formal. Applied to the debate in psychology, Killeen (2001) considers that material causes are organic substrates necessary for the behavior; formal causes are models and theories that describe relationships among events; final causes are functional explanations; and efficient causes are triggers, or stimuli that initiate actions. Tinbergen (1963), aiming a systematization of the study of behavior in the field of ethology, proposed that behavior can be explained by the study of its mechanisms, ontogeny, adaptive function and phylogeny. Wilson (1975) refined Tinbergen's classification by putting the first two causes as proximal variables, encompassing learning processes, development and

physiological functioning across an individual's lifespan, and the next two as distal variables, related to the adaptive value of behavior established by the biological evolution of human species.

Far from an attempt to separate research fields by types of causes, it is yet possible to consider that the different models and procedures that have been applied to the study of morality are focusing more on one or some of these causes. As such, the integration of proximal and distal perspectives may allow a deeper understanding of the moral phenomenon.

As stated by Rachlin (1992), evolutionary thinking is not incompatible with learning related to social behavior. Evolution is responsible for *mechanisms* (e.g., kin selection, reciprocity etc, see Nowak, 2006 for a review) that explain social behavior at a populational level. Behavior analysis is more concerned about final causes for behavior across learning process. Final causes may be restrict to, when we are able to identify, specific consequences for specific classes of behavior. Final causes may also be broader, when we are able to reach molar consequences that are generally intrinsic to behavior extend in time.

Stimulus Relations and Relational Responding

Procedures such as equivalence classes and the implicit relational assessment procedures (IRAP) have been used with the purpose of capturing group phenomena in laboratory-based basic investigations (e.g., Watt, Keenan, Barnes, & Cairns, 1991; Dixon, Rehfeldt, Zlomke, & Robinson, 2006; Mizael et al, 2016; Rabelo, Bortoloti, & Souza, 2014, to mention some). These procedures have shown to be sensitive to social learning that occurs outside the laboratory and reveal explicit and implicit aspects of moral behavior. They address moral behavior by showing relations between stimulus class and behavior (i.e. stimulus control).

In a seminal study, Watt et al (1991) sought to analyze the phenomenon of religious intolerance in Northern Ireland through an equivalence class formation procedure. In this study, participants from three ethnic groups, Northern Irish Catholics, Northern Irish Protestants, and English Protestants were trained to match Catholic names to nonsense syllables and the same nonsense syllables to Protestant symbols. The paradigm of equivalence class formation predicted that relations between Catholic names and Protestant symbols should emerge. However, such relations would conflict with the history of socially-mediated reinforcement of the Northern Irish groups, but not the English group. Indeed, English participants correctly matched Protestant symbols to Catholic names, but twelve out of nineteen Northern-Irish did not show the predicted emergent relation. Given these results, it can be argued that the emergent relation contradicted the social interaction history of the Irish participants and conflicted with their pre-experimental learned stimulus relations. The symbolic relations captured in the procedure illustrate the cultural practice in Northern Ireland of separating Catholics from Protestants. Other studies exploring the same premises found similar results while exploring issues such as racism (de Carvalho e de Rose, 2014; Mizael et al, 2016), attitudes towards soccer teams (Haydu, Camargo & Bayer, 2015), and terrorism (Dixon et al, 2006).

The IRAP consists of a computer-based procedure in which the participant is presented with two stimuli (label and target) and two relational terms used to indicate the relation between stimuli (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010; Power et al, 2009). These pairs of related stimuli are presented under two different blocks; consistent and inconsistent. Under the consistent condition, participants must follow a rule to select the relational term indicating the typical relation between label and target. For example, when given “black,” and “white” the participant must select “different” and not “same.” Conversely, in the inconsistent blocks, participants are instructed to select the opposite, non-

typical relation. When given “black” and “white” participants would select “same”. The two blocks generally produce different patterns of response in respect to mean latencies with shorter means indicating a relation between the two stimuli. Thus, by the IRAP it is possible to observe symbolic relations, and then drawing conclusions about personal classifications of a behavior as *right* or *wrong* and its descriptive function of determined cultural practices.

In a study conducted by Rabelo, Bortoloti, and Souza (2014), children (average age 8 years) responded to an IRAP about types of toys and gender. Stimuli presented were photographs of dolls and cars paired with photos of boys and girls. Results showed a bias toward a relation stating that dolls are for girls but are forbidden for boys, yet no gender bias was observed regarding cars. Other several IRAP experiments have investigated social issues underlying moral considerations towards money and capitalism (O’Shea, 2015), racism (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010; Mizael et al, 2016), and the dehumanization of women (Hussey et al, 2016) to mention a few.

Empirical evidence produced via IRAP indicates it can be an alternative procedure to self-report measures. While the IRAP can capture automatic evaluations, self-report measures risks capturing carefully elaborated reactions. Such distinction is important when results from IRAP and self-report measures are compared with performance in non-verbal tasks. Some data suggests that only the IRAP can predict behaviors related to the content of the stimulus relations under analysis (Bast & Barnes-Holmes, 2015; Carpenter et al., 2012; Hooper, Villatte, Neofotistou, & Mchugh, 2010; Leech, Barnes-holmes, & Madden, 2016; Perez et al., 2017).

Equivalence class formation and IRAP are procedures that deal with symbolic relations. In cultural practices, symbols are the very pillars of the social institutions, fundamental to the understanding of the control of behavior (Geertz, 1973). Advances in experimental analysis of behavior enhanced our comprehension on how symbols are formed and how they are related

to behavior and learning (de Rose, 2016). So, experimentation on stimulus relations and relational responding concerning social issues produces direct contributions to the comprehension of cultural matters as a complex net of symbolic relations related to daily life behaviors. One advantage of the stimulus relations and relational responding procedures is its potential to predict other behaviors related to the content of the symbolic net under analysis. By mapping stimulus class and stimulus relations it may be possible to draw predictions about several related non-verbal behavior critical for the comprehension of cultural practices.

Decision-Making in Moral Situations

Experimental procedures into moral decision-making processes focus on the understanding of immediate contextual factors that affect the judgment of a behavior as *right* or *wrong*. Typically, a procedure consists of the presentation of written moral dilemmas to participants who must choose one out of two alternatives of action that involves harming one or more persons (May, 2014).

Studies on moral decision-making are mainly based on assumptions of information-processing theories (e.g., Simon, 1955, 1978) with emphasis on the dual-model classification system that places moral decision-making in either intuitive or deliberative categories. Intuitive processes are characterized by quick, automatic processing with little or no effort and no perception of voluntary control. Deliberative processes are rational and entail the voluntary allocation of attention and awareness to the decision-making process (see also Kahneman, 2011; Simon, 1955, 1978).

This kind of discussion relating data from cognitive sciences with the ones more related to a behavioristic tradition in research may generate several interesting epistemological and philosophical debates (Rachlin, 1989). Our interest, however, is to highlight empirical evidence that strongly demonstrate that moral decisions are not entirely dependent on verbal

behavior. They may be related to verbal behavior in the sense that people need to make decisions via verbal behavior and evaluate the moral aspects of verbal statements. But, moral decisions may also be intuitive, implicit and occur under the strong influence of incidental variables as time restrictions (thereby decreasing the influence of verbal behavior in decision making). Verbal behavior may help to organize and amplify morality, but morality probably “came first” in the same sense that Hayes and Sanford (2014) argue that cooperation, as evolved forms of social behavior, are prerequisites of cultural organization and social cognition.

To further the distinction between intuition and deliberation, moral decisions can be classified as utilitarian or deontological (Bartels et al, 2015; Cushman, 2013). Utilitarian judgments are those that maximize gains (Bartels, et al., 2015; Gleichgerrcht, Torralva, Rattazzi & Marengo, 2013; Gleichgerrcht & Young, 2013). They would be aligned with a consequentialist ethical system, according to which an action is considered correct if, and only if, the greatest possible benefit with smallest loss is produced (to put it colloquially, “the ends justify the means;” Bartels et al., 2015). In deontological moral decisions, the value of an action is attributed to the qualities of that action (if it complies with or breaks a moral rule) independent of its consequences. In an example exerted from Bartels et al (2015), if a person follows the moral rule, “To pollute the environment is wrong” with a deontological perspective, she may agree that, “Companies should not be allowed to buy the right to pollute the earth even if pollution credits reduce pollution”. Thus, the deontological judgment is consequence-insensitive. In the two alternatives of action offered in the dilemmas, one always addresses the utilitarian values and the other the deontological values. Thus, a moral dilemma would be characterized as a situation in which there is conflict between the two processing systems. Some studies, such as Gleichgerrcht Torralva, Rattazzi & Marengo (2013), Gleichgerrcht and Young (2013), Cushman (2013), and Suter and Hertwig (2011),

consider that automatic/emotional processes lead to deontological judgments and rational/controlled processes produce utilitarian judgments.

Morality has always been treated as a different matter compared to personal choices or social conventions, because it is taken as the most compulsory rules (Bartels et al, 2015). There is more tolerance for variety in personal tastes and cultural traits than for morally deviant behavior. However, evidence about moral judgment reinforces the view of a flexible morality, showing how moral judgment is strongly influenced by manipulations of contextual variables that affect the presentation of the dilemma and the conditions under which decision-making was made. A widely explored aspect in these studies is the effect of time pressure on moral decision-making. Suter and Hertwig (2011) evaluated the effect of time pressure on moral decision-making across two experiments. In Experiment 1, participants were divided between two conditions: with and without temporal pressure. All participants read a dilemma about injury caused to others and then were asked to morally judge the agent's action. In the time pressure condition, the participants had to present their evaluation in less than 8 seconds. In the condition without time pressure, participants had 3 minutes to present their evaluation. Experiment 2 was similar, but participants were not presented with predetermined time limits to decision making. Participants were again divided into two groups and were directed to either make a decision as quickly as they could (according to their intuition) or to deliberate over the dilemma as long as they needed to. Results showed that shorter decision times were correlated with higher occurrences of deontological judgments.

Rand, Greene, and Nowak (2012) also found similar results in experiments that manipulate cooperation using behavioral games. In the first experiment, 212 participants received 40 cents and were randomly divided into groups of four to play a *Public Goods Game*. Each participant was asked to choose how many cents they would contribute to the common fund, if any. They were informed that the total amount of the common fund would

be doubled and then divided among the four members of the group. The participants were then scored according to the amount donated and the amount of time each one took to make the donation. A correlational analysis found a negative correlation between the contribution amount and decision-making time, meaning as decision-making time increased, the contribution amount decreased. In Experiment 2, temporal pressure conditions were introduced. In the condition with temporal pressure, participants were informed that they had a maximum of 10 seconds to make a contribution or not. In the condition without temporal pressure, participants were asked to think carefully about their decision and to make the donation after at least 10 seconds. The results were a strong replication of the findings in the previous experiment with a negative correlation between decision-making time and the amount of the donation. This is counterintuitive if we consider that verbal behavior and contextual social (deliberative elements) control must inhibit selfish behavior and immediate control by reinforcement (Skinner, 1969). One solution for this apparent paradox is to consider that morality is social, but not necessarily verbal. These studies analyzing the relation between time pressure and moral decisions suggest a parallel between utilitarian judgments and lower frequencies of cooperation, and deontological judgments and higher frequencies of cooperation. Gleichgerrcht and Young (2013) also showed a relation between lower levels of empathy and higher frequencies of utilitarian judgments.

Adding to this line of research, studies concerning moral judgment investigate elements of the dilemma's description, such as attributions of causality and intentionality of the agent, and the presence or absence of physical contact between agent and target of damage. Cushman and Young (2011) investigated three descriptive frames to moral judgments. Under the frame of the *doctrine of doing and allowing*, actions are morally worse than omissions. The *doctrine of double-effect* states that damage caused as a means to an end is morally worse than as a collateral damage. The last frame states that directly causing physical harm to

another human being is worse than harm that is caused without physical contact. Forty undergraduate college students were randomly assigned across two conditions. In the moral condition, the dilemmas presented life and death issues. In the non-moral condition, the dilemmas presented interactions between the agent and inanimate objects. Participants read 30 scenarios and were requested to indicate the degree of responsibility of the agent concerning the production of the action; the degree of intention of the agent in producing the result; and if the action should be forbidden, permissible, or obligatory. Results showed that in the moral condition, participants judged agents in actions (in opposition to omissions) as more responsible, more intentional, and as playing a bigger causal role in the result. In the non-moral condition, differences showed no statistical significance in any of the three dimensions concerning the difference between actions and omissions. Thus, the moral element of the scenarios influenced the attribution of causality and intentionality.

Such findings demonstrate how moral decision-making is context-flexible. Besides, this influence does not need to be exerted explicitly or verbally. This line of empirical investigation shows how efficient and final causes play a large role in how a person decides a course of action or judges other people's behavior.

Evolutionary Aspects of Morality

Biological perspectives of morality are based on evidence indicating an evolutionary root for moral behavior. Good candidates for an evolutionary predisposition should be behaviors that are observed: (1) in different species; (2) in highly diverse cultural groups; (3) in early childhood or in a clear point in development. An old dichotomy between learning and innate structures in psychology may erroneously suggest that an evolutionary root for specific behavior implies that such behavior needs little or no learning to emerge. On the contrary, the inquiry into biological roots of morality highlights the need to study the ontogenesis

(Tinbergen, 1963; Tomasello, 2016) and cultural heritage (Boyd & Richerson, 1985) of morality as a form of social behavior.

It is argued in this research field that social coexistence and, more specifically, social informal norms that dictate the obligatory and prohibited nature of certain acts of conduct (e.g., murder or adultery), have an evolutionary root, (i.e., distal causes; Haidt, 2001, 2008; Haidt & Graham, 2007; Tomasello, 2016). Thus, such arguments support a naturalistic and historical analysis of the development of the “sense of morality” in humans and other species (Tomasello, 2016). We briefly present here three models of evolutionary roots of morality, to highlight marked differences in the way the evolutionary propositions about moral behavior have been taken. They are all data-based discussions, but for the sake of economy, we don’t mention the experiments that demonstrate those models.

Tomasello (2016) proposes the *hypothesis of interdependence* suggesting that cooperation and morality in human behavior are the result of two great ecological, or selective, pressures. The first was that which induced the first human beings to collectively forage for food, which led to an increased chance of survival and procreation across members of the group. Thus, the context allowed for the emergence of cooperation repertoires, because survival depended on coexistence in groups and collaboration on specific survival-related activities.

The second occurred with an increase in population growth in human groups. With large groups arises the need for division of labor across members and some social skills seemed to have been essential for maintenance of the social organization of an expanding population.

Thus, a natural history of morality suggests that when a modern human being is born into a social group, it must: a) exhibit the capacity to self-regulate his or her behaviors according to social norms, b) identify with norms and institutions, and c) feel obliged to

follow and reinforce such norms. Group members who act in misalignment with such standards, but wish to maintain their identity as a group member, must justify their deviant action by pointing conflicts between different values of the group. Otherwise, they will be considered moral violators.

The second evolutionary model for the origin of morality is addressed in the *moral foundations theory* - MFT (Graham et al., 2012; Haidt & Graham, 2007). This theory posits that all social organizations maintain, through differing levels of force, five foundations that establish the rules for coexistence inside the group. These foundations can be characterized according to the following pairs of values: harm/care, fairness/reciprocity, loyalty/ingroup, authority/respect and purity/sanctity. According to MFT, society may emerge when a large number of people interact with one another under conditions controlled by a set of institutions and technologies that were created based on these five foundations (Graham et al., 2011). By its evolutionary function, the moral foundations would be universal to some extent. Cultural and personal variations would occur depending on the strength in which each foundation is enforced within the group in its proximal history. Variability is also influenced by historical events within the culture.

The third theoretical model is presented by Curry (2004) and may be called *the morality/cooperation hypothesis*. According to this hypothesis, human beings might be endowed with adaptations for cooperation that make them social, cooperative, and altruistic beings under certain circumstances. Moral values may be products of these adaptations. The author identifies morality with cooperation, indicating that acting morally results in increased cooperation. The hypothesis points to four types of cooperation: altruism by kinship, coordination for mutual advantage, reciprocity, and conflict resolution. Curry advocates that from these four types, it would be possible to adequately describe and understand the diversity of philosophical theories on morality as well as the problems of meta-ethics.

The evolutionary approach highlights the distal, final causes for moral behavior. The theories we choose to describe in this section are all in some way related to the evolution of cooperation (Hayes & Sanford, 2014). Such causes are seen to produce dispositions for behavior, but moral behavior remains strongly influenced by proximal variables.

The Challenge to Integrate Different Causes of Behavior

Our review suggested that the recent experimental literature about morality supports the idea that morality is influenced by multiple causes including proximal and distal causes. There is some convergent evidence that may be abstracted as, a) Considerations about *right* or *wrong* are influenced by incidental, immediate environmental features thereby causing moral behavior to be context-dependent; b) Such considerations are also under the control of symbolic nets subsumed to social and cultural influence; c) Moral values do not need to be verbal or explicit, they may be implicit or intuitive and specific behavioral procedures proved to be a useful strategy to demonstrate this; and, d) Distal causes, that is, the evolutionary roots of morality, are not in opposition of cultural evolution and social influences. On the contrary, they provide a framework to integrate human nature to reach a broader understanding of morality through considering the possibility of a disposition to learn moral behaviors.

Explanatory models for moral behavior can integrate distal and proximal causes, or distinct causes in the original Aristotelian sense (final, formal, efficient, and material), making the opposition among these different epistemologies unnecessary. Different causes may be integrated by behavior sciences, more than stated in necessary opposition as explanations of moral behavior. We argue here that capturing the interplay among epistemologies benefits psychological research.

Stimulus control and relational responding add key information to understand proximal causes to moral behavior. A deep theoretical questions arise when we ask if these areas of

research may also deal with final or efficient causes. Killeen (2001), for example, stated that final causes comprise functional explanations and addresses evolutionary explanations of behavior. Rachlin (1992) put final causes as explanations that use, “external objects and goals to explain overt behavior” (p.1371). We do not intend to solve this complex question. But this debate may extend the general idea of final causes from evolution to learning considering that both part from the principle of selection by consequences. As stated by Baum (2012), while evolution deals with the population of individuals in the species, behavior sciences usually deal with populations of behavior.

The Culturo-Behavioral Science

Foundational discussions about culturo-behavioral science suggest that interlocking behaviors are the main unit of analysis of social behavior and cultural practices (e.g., Cihon & Mattaini, 2019; Glenn et al., 2016; Houmanfar & Rodrigues, 2006; Mattaini, 2006). This position does not preclude behavior and learning principles such as reinforcement, punishment and different modalities of stimulus control and relational responding. On the contrary, they suggest the need for a synthesis between principles that regulate individual interactions with principles that describe and explore interlocking behavior and how they are related to self regulated systems (e.g., Mattaini, 2004, 2006), cultural practices maintained by the external environment (Malott & Glenn, 2006; Glenn et al, 2016) and the cultural *milieu* (Houmanfar, & Rodrigues, 2006). There are, of course, discussions to be made about what are the foundational elements of cultural practices and systems (see, for example, recent discussions about the notion of metacontingency and some critical views about the notion in Glenn et al, 2016; Mattaini, 2006; Krispin, 2016, 2017; Couto & Sandaker, 2016; Zilio, 2019). But it is safe to assume that learning principles will have a place among such elements.

The findings about symbolic and implicit relations may also suggest the need to refine the role of verbal behavior in maintaining interlocking behavior in cultural practice. Rule following is a key aspect in social transmission in Houmanfar and Rodrigues (2006) proposition about cultural *milieu*. For Glenn's (1991) account on metacontingencies, verbal behavior is the "glue" that maintains complex interlocking behavior relations in cultural practices (p. 60), distinguishing human cultures from protocultures of humans and other primates. The present analysis about morality may suggest that moral behavior, rather than verbal behavior *per se*, is the real glue that maintains social behavior. Reasons to punish or reinforce another person's behavior based within a particular cultural context do not need to be formulated as rules. Further, these rules do not need to be rational or even explicit in order to affect moral behavior. Verbal behavior may influence and control moral behavior, but moral behavior persists in the absence of rational, explicit verbal behavior.

Symbolic relations may enhance the comprehension of the features of the cultural *milieu* where interlocking behaviors take place. According to Houmanfar and Rodrigues (2006) social interactions and cultural practices occurs in a "cultural *milieu* of beliefs, material resources, organizational as well as overarching governmental policies, rules, traditions, morals, institutions, technological progress, and environmental competition" (p. 23). Procedures to investigate symbolic behavior appears to capture the "state of art" of moral values within a group, so, they may be an economic way to access the cultural *milieu* when interlocking behavior contingencies are occurring.

Another important contribution of the present paper is the consideration of how contextual-dependent moral decisions can be. If the frame of presentation of a moral issue can influence how we respond to it, such knowledge may be of great help in the planning of social contingencies for cultural change.

Our analysis about morality in this paper is based on recent experimental literature that challenges behavioral scientists to connect and discuss different causes of behavior. Psychological literature emphasizes final and efficient causes of behavior, but the four causes are interconnected in the study of social behavior and culture.

References

- Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A Sketch of the implicit relational assessment procedure (IRAP) and the relational elaboration and coherence (REC) model. *The Psychological Record*, *60*(3), 527-542. doi: 10.1007/BF03395726
- Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y. & Stewart, I. (2010). The implicit relational assessment procedure: exploring the impact of private versus public pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record*, *60*(1), 57–80. doi:10.1007/BF03395694B
- Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, P. (2015). Moral judgment and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of judgment and decision making*. Chichester, UK: Wiley. doi: 10.1002/9781118468333.ch17
- Bast, D. F., & Barnes-Holmes, D. (2015). Developing the implicit relational assessment procedure (IRAP) as a measure of self-forgiveness related to failing and succeeding behaviors. *The Psychological Record*, *65*(1), 189–201. doi: 10.1007/s40732-014-0100-5
- Baum, W. M. (2012). Rethinking reinforcement: allocation, induction, and contingency. *Journal of The Experimental Analysis of Behavior*, *97*(1), 101-124. doi: 10.1901/jeab.2012.97-101
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
- Carpenter, K. M., Martinez, D., Vadhan, N. P., Barnes-Holmes, D., & Nunes, E. V. (2012). Measures of attentional bias and relational responding are associated with behavioral treatment outcome for cocaine dependence. *The American Journal of Drug and Alcohol Abuse*, *38*(12), 146–154. doi:10.3109/00952990.2011.643986

- Cihon, T. C., & Mattaini, M. A. (2019). Emerging cultural and behavioral systems science [Editorial]. *Perspectives on behavior science*, 42, 699-711. doi: 10.1007/s40614-019-00237-8
- Couto, K. C., & Sandaker, I. (2017). Natural, behavioral and cultural selection-analysis: an integrative approach. *Behavior and Social Issues*, 25, 54-60. doi: 10.5210/bsi.v25i0.6891
- Curry, O. S. (2004). *Morality as natural history: an adaptationist account of ethics* (doctoral dissertation). Retrieved from core.ac.uk/download/pdf/40091549.pdf
- Cushman, F. A. (2013). Action, outcome, and value: a dual-system framework for morality. *Personality and Social Psychology Review*, 17(3), 273–292. doi: 10.1177/1088868313495594
- Cushman, F. A., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, 35(6), 1052–1075. doi: 10.1111/j.1551-6709.2010.01167.x
- de Carvalho, M. P., & de Rose, J. C. (2014). Understanding racial attitudes through the stimulus equivalence paradigm. *The Psychological Record*, 64(3), 527-536. doi: 10.1007/s40732-014-0049-4.
- de Rose, J. C. (2016). A importância dos respondentes e das relações simbólicas para uma análise comportamental da cultura. [The importance of respondents and of symbolic relations to a behavioral analysis of culture]. *Acta Comportamentalia*, 24(2), 201-220. Retrieved from www.revistas.unam.mx/index.php/acom/article/view/55783
- Dittrich, A. (2004). A ética como elemento explicativo do comportamento no behaviorismo radical. [Ethics as a explanatory element of behavior in radical behaviorism]. In Maria Zilah Brandão et al (Eds.), *Sobre comportamento e cognição* vol. 13 (pp. 21-26). Santo André:ESEtec.

- Dittrich, A. (2010). Ética e comportamento. [Ethics and behavior]. In Emanuel Zagury Tourinho & Sérgio Vasconcelos Luna (Eds.), *Análise do comportamento: investigações históricas, conceituais e aplicadas* (pp. 37-60). São Paulo: Roca.
- Dixon, M. R., Rehfeldt, R. A., Zlomke, K. R., & Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *The Psychological Record*, 56(1), 83-103. doi: 10.1007/BF03395539
- Garrett, R. (1979). Value conflict in a skinnerian analysis. *Behaviorism*, 7(1), 9-16. Retrieved from www.jstor.org/stable/27758926
- Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
- Gleichgerrcht, E., Torralva, T., Rattazzi, A., & Marengo, V. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *SCAN*, 8(7), 780–788. doi: 10.1093/scan/nss067
- Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLOS One*, 8(4), 1–9. doi: 10.1371/journal.pone.0060418
- Glenn, S. S. (1991). Contingencies and metacontingencies: relations between behavioral, cultural, and biological evolution. In P. A. Lamal (ed.) *Behavioral analysis of societies and cultural practices*, (pp.39-73). Washington, DC: Hemisphere Publishing.
- Glenn, S. S., Malott, M. E., Andery, M. A. P. A., Benvenuti, M., Houmanfar, R. A., Sandaker, I.,... & Vasconcelos, L. A. (2016). Toward consistent terminology in a behaviorist approach to cultural analysis. *Behavior and Social Issues*, 25, 11-27. doi: 10.5210/bsi.v.25i0.6634

- Graham, G. (1977). On what is good: a study of B. F. Skinner's operant behaviorist view. *Behaviorism*, 5(2), 97-112. Retrieved from www.jstor.org/stable/27758896
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: the pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55–130. doi:10.1016/B978-0-12-407236-7.00002-4
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., & Ditto, P. H. (2011). Mapping the moral domain. *Perspectives in Social Psychology*, 101(2), 366–385. doi: 10.1037/a0021847
- Haidt, J. (2001). The emotional dog and its rational tail : a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. doi: 10.1037//0033-295X.
- Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, 3(1), 65–72. doi: 10.1111/j.1745-6916.2008.00063.x
- Haidt, J., & Graham, J. (2007). When morality opposes justice : conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1), 98–116. doi: 10.1007/s11211-007-0034-
- Haydu, V. B., Camargo, J., & Bayer, H. (2015). Effects of pre-experimental history on the formation of stimulus equivalence classes: A study with supporters of Brazilian soccer clubs. *Psychology & Neuroscience*, 8(3), 385-396. doi:10.1037/h0101276
- Hayes, S.C., Gifford, E. V., Hayes, G. J. (1998). Moral behavior and the development of verbal regulation. *The Behavior Analyst*. 21(2), 253-279. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2731399/>

- Hayes, S. C., Sanford, B. T. (2014). Cooperation came first: evolution and human cognition. *Journal of The Experimental Analysis of Behavior*, *101*(1), 112-129. doi: 10.1002/jeab.64
- Hocutt, M. (1977). Skinner on the word “good”: a naturalistic semantics for ethics. *Ethics*, *87*(4), 319-338. Retrieved from www.jstor.org/stable/2379901
- Hocutt, M. (2013). A behavioral analysis of morality and value. *The Behavior Analyst*, *36*(2), 239-249. doi: 10.1007/bf03392310
- Hooper, N., Villatte, M., Neofotistou, E., & Mchugh, L. (2010). The effects of mindfulness versus thought suppression on implicit and explicit measures of experiential avoidance. *International Journal of Behavioral Consultation and Therapy*, *6*(3), 233–244. doi: 10.1037/h0100910
- Houmanfar, R., & Rodrigues, N. J. (2006). The metacontingency and the behavioral contingency: points of contact and departure. *Behavior and Social Issues*, *15*(1), 13-30. doi: 10.5210/bsi.v15i1.342
- Hussey, I., Mhaoileoin, D. N., Barnes-holmes, D., & Ohtsuki, T. (2016). The IRAP is nonrelative but not acontextual: changes to the contrast category influence men’s dehumanization of women. *The Psychological Record*, *66*(2), 291–299. doi: 10.1007/s40732-016-0171-6
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus And Giroux.
- Killeen, P. R. (2001). The four causes of behavior. *Current Directions in Psychological Science*, *10*(4), 136–140. doi: 10.1111/1467-8721.00134
- Krispin, J. V. (2016). What is the metacontingency? Deconstructing claims of emergence and cultural-level selection. *Behavior and Social Issues*, *25*, 28-42. doi: 10.5210/bsi.v25i0.6186

- Krispin, J. V. (2017). Positive feedback loops of metacontingencies: a new conceptualization of cultural-level selection. *Behavior and Social Issues*, 26, 95-110. doi: 10.5210/bsi.v26i0.7397
- Leech, A., Barnes-holmes, D., & Madden, L. (2016). The implicit relational assessment procedure (IRAP) as a measure of spider fear, avoidance, and approach. *The Psychological Record*, 66(3), 337–349. doi: 10.1007/s40732-016-0176-1
- Leigland S. (2005). Variables of which values are a function. *The Behavior analyst*, 28(2), 133–142. doi:10.1007/bf03392109
- Malott, M. E., & Glenn, S. S. (2006). Targets of intervention in cultural and behavioral change. *Behavior and Social Issues*, 15(1), 31-56. doi: 10.5210/bsi.v15i1.344
- Mattaini, M. A. (2004). Systems, metacontingencies, and cultural analysis: are we there yet?. *Behavior and Social Issues*, 13(2), doi:10.5210/bsi.v13i2.20
- Mattaini, M. A. (2006). Will cultural analysis become a science?, *Behavior and Social Issues*, 15(1), 68-80. doi: 10.5210/bsi.v15i1.380
- May, J. (2014). Moral judgment and deontology: empirical developments. *Philosophy Compass*, 9(11), 745–755. doi: 10.1111/phc3.12172
- Mizael, T. M., de Almeida, J. H. de, Silveira, C. C., & de Rose, J. C. (2016). Changing racial bias by transfer of functions in equivalence classes. *The Psychological Record*. 66(3), 451-462, doi: 10.1007/s40732-016-0185-0
- Nowak M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560–1563. doi:10.1126/science.1133755

O'Shea, B. (2015). Capitalism versus a new economic model : implicit and explicit attitudes of protesters and bankers. *Social Movement Studies, 14*(3), 311–330. doi:

10.1080/14742837.2014.938732

Perez, W. F., Nico, Y. C., Fidalgo, A. P., Caro, D., Linares, I. M. P., Boavista, R. R. C., ... dos Santos, G. A. R. (2017). Utilizando o implicit relational assessment procedure (IRAP) para avaliar relações verbais implicadas na esquiva experiencial: um estudo piloto. [Using the implicit relational assessment procedure (IRAP) to evaluate verbal relations implied on experiential avoidance]. *Revista Brasileira de Terapia Comportamental E Cognitiva, 19*(3), 77–91. doi:10.31505/rbtcc.v19i3.1056

Power, P., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The implicit relational assessment procedure (IRAP) as a measure of implicit relative preferences : a first study. *The Psychological Record, 59*(4), 621–640. doi: 10.1007/BF03395684

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. doi: 10.1007/s40732-014-0006-2

Rachlin, H. (1989). *Judgment, decision, and choice: a cognitive/behavioral synthesis*. New York: W. H. Freeman and Company.

Rachlin, H. (1992). Teleological behaviorism. *American Psychologist, 47*(11), 1371-1382. doi: 10.1037/0003-066X.47.11.1371

Rand, D., G., Greene, J., D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*, 427-430, doi:10.1038/nature11467

- Rottschaefer, W. (1980). Skinner's science of value. *Behaviorism*, 8(2), 99-112. Retrieved from www.jstor.org/stable/27758957
- Ruiz M. R. (2013). Values and morality: science, faith, and feminist pragmatism. *The Behavior analyst*, 36(2), 251–254. doi:10.1007/bf03392311
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99-118. doi: 10.2307/1884852
- Simon, H.A. (1978). Prize lecture: rational decision-making in business organizations. Retrived from http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1978/simon-lecture.html
- Skinner, B. F. (1938). *The behavior of organisms: an experimental Analysis*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1953). *Science and human behavior*. Oxford: McMillan.
- Skinner, B. F. (1956). *Cumulative record*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1969). *Contingencies of reinforcement: a theoretical analysis*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Penguin Books.
- Skinner, B. F. (1981). Selection by consequences. *Science*, 213(4507), 501-504. doi: 10.1126/science.7244649
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119(3), 454–458. doi: 10.1016/j.cognition.2011.01.018

Tinbergen, N. (1963). On aims and methods of ethology. *Ethology*, 20(4), 410-433.

doi:10.1111/j.1439-0310.1963.tb01161.x

Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge, MA: Harvard University Press.

Capítulo 2

An empirical investigation on moral values: explicit and implicit values of fairness and purity²

² Este capítulo está no formato em que está sendo preparado para submissão como artigo.

Abstract

Moral behavior comprehends a wide set of repertoires encompassing the classification of behaviors in terms of "right" or "wrong" and the use of such classification to adjust one's own behavior and to provide differential consequences for other person's behavior. The literature concerned to the assessment of individual moral values has sought to identify individual moral values using mostly self-report instruments. However, this type of data is highly susceptible to distortions that compromises the use of self-reports as predictors of non-verbal behavior. The present study proposes the use of the implicit relations assessment procedure – IRAP –, an alternative procedure to self-report procedures, for the identification of moral values of fairness and purity. IRAP procedure uses, as basic data, the mean latency of the response to sets of verbal stimuli, consistent or not between them. Consistency is deemed in terms of the personal reinforcement history of relations (or not) among stimuli composing the set. Twenty participants first responded to a self-report questionnaire, the Moral Foundations Questionnaire (MFQ) and then to the IRAP. Results were different depending on the instrument. In questionnaire responses we observed high degrees of agreement with statements about fairness and high degrees of disagreement with the assertions about purity in all participants. In IRAP, the biases of agreement and disagreement measured by mean response latency indicated similar degrees of agreement between the two values. The findings raise important methodological questions about the measurement of moral values in terms of predicting nonverbal moral behavior.

Key words: Moral Values; IRAP; implicit measures; symbolic relations.

Moral behavior is a special and complex set of repertoires encompassing conducts that should improve coexistence by curbing selfish acts (Curry, 2005; DeWaal, 2006; Haidt & Graham, 2007; Skinner, 1953; Skinner, 1971; Tomasello, 2016). A behaviorist account on moral behavior suggests it encompasses repertoires installed and maintained in social contingencies where a morally desired behavior is the one enforced by pairs without the need of a formal control as found in controlling agencies (Hocutt, 1977, 2013).

In terms of function, such repertoire encompasses the classification of behaviors as "right" and "wrong", the adjustment of one's own behavior unto such classification, and the provision of reinforcement or punishment for others person's behaviors based on these classifications (Hocutt, 1977, 2013; Skinner, 1971).

In behavior analysis, discussions on moral behavior are essentially theoretical. They have its start on Skinner (1953, 1961, 1971) and appear in notions as group control (Skinner, 1953), values (Dittrich, 2010; Garret, 1979; Graham, 1977; Hocutt, 1977; Hocutt, 2013; Leigland, 2005; Rottschaefter, 1980; Waller, 1982), and cultural *milieu* (Houmanfar, & Rodrigues, 2006).

Beyond the scope of behavior analysis, recent advances in the empirical research on moral behavior are located in two major branches. The first one is characterized by the search for distal, evolutionary routes of morality (Curry, 2005; DeWaal, 2006; Haidt & Graham, 2007; Tomasello, 2016). Another set of advances come from the search of proximal, contextual variables and learning processes that affect moral decisions or moral judgments (Bartels et al, 2015; Cushman & Young, 2011; Gleichgerricht et al 2013; Suter & Hertwig, 2011). And also, in the description of examples from human and non-human research relating features of context as creating the conditions for the selection of cultural practices (see Harris, 1974, 1979; Laland & Hoppitt, 2003 for examples).

One prolific field of research in the evolutionary branch lead to the proposal of the Moral Foundations Theory-MFT (Haidt & Graham, 2007; Graham & Haidt, 2010; Graham et al, 2011). It states the every social organization maintains, with more or less force, five moral foundations that establish rules (not in the sense of a behavior analytic reference to rules, but as principles) for behavior: care/harm; fairness/reciprocity; loyalty/ingroup; authority/respect; and purity/sanctity (Haidt & Graham, 2007). Those foundations are pointed as organizing principles that turn coexistence possible (Haidt & Graham, 2007; Graham & Haidt, 2010; Graham et al, 2011).

Based on the assumptions, these authors developed the Moral Foundations Questionnaire (MFQ)³. It has already been answered by more than 30,000 participants (mostly people born in the USA) and has been used for predictive analyzes on foundation's clusters (Graham et al, 2011; Haidt, Graham & Joseph, 2009) as the relationship between moral values and judgments and decision-making about party-political position (Haidt & Hersh, 2001; 2012; Haidt & Graham, 2007), abortion, torture, and same-sex marriage (Koleva et al, 2012), to name a few.

Through the MFQ, moral values have been approached as verbal behavior accessed through participants' self-report. However, as any self-report instrument, results may be subject to intervening variables such as social desirability, the tendency to give positive answers in order to present a particularly positive self-image (Paunonen & LeBel, 2012; Ziegler & Buehner, 2009). That is, verbal statements are sensitive to variables that biased verbal behavior such as specific social reinforcement and contextual social cues (Skinner, 1957).

The Implicit Relational Assessment Procedure (IRAP) may be used as an alternative to explicit measures of behavior. IRAP present stimulus sets to participants on a computer

³ Available online at <https://www.yourmorals.org/index.php>

screen. Three types of stimuli make up this combination. At the top of the screen are presented the label stimuli. Usually at least two label stimuli are used that represent opposite categories (right / wrong, pleasant / nasty, white / black, woman / man). The target stimuli are displayed in the center of the screen. The number of target stimuli ranges from a minimum of 6 to a maximum of 12 (Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010). The label and target stimuli can be words, complete sentences or images. In the lower corners of the screen the two relational terms are presented. They are presented in opposite valences (true/false, equal/opposite).

The response requested from the participant is to choose the relational term that best expresses the relation between label and target stated by a previous rule. Trials are presented in two types of blocks, consistent and inconsistent. These blocks are differentiated in terms of the type of rule to respond that must be followed by the participant. If the participant's response is not in the directions of the pre-programmed rule for the current block, it is exposed to a correction procedure that forces the correct response to be issued as a condition to finish the trial. (Carpenter et al, 2012; Dawson et al., 2009; Hussey et al., 2016, McKenna et al., 2007).

Typically, participants are exposed to a minimum of two practice blocks followed by a fixed number of six test blocks. Practice and test have the same stimuli (Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010). Exposure to the early blocks is considered a practice only because the data produced in it is not used in the analysis. Participants are usually asked to produce an accuracy rate on each of the test blocks. This rate may vary from 65% (Drake et al, 2010) to 80% of correct responses (Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010). Participants are requested to respond to these trials as soon as possible. In recent studies it has been common to additionally impose a maximum time interval of 2 seconds for

the response to be considered valid (Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010; Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010).

The whole procedure is supported by the widely empirically verified assumption that the latency of the response will be shorter in trials in which the rule imposes a response that confirms the relationship among the set of stimuli that are related to each other in the participant's history of reinforcement. Thus, the participant's performance in confirming and denying the same relation among stimuli produces a differentiated pattern of latencies between consistent and inconsistent responses called the "IRAP effect" (Power et al, 2009).

The "IRAP effect" is determined by calculating the D-IRAP score, and indicates the biases of probability of response to pairs of stimuli. These biases have been referred as implicit attitudes, defined as traits of past experience that mediate favorable or unfavorable feelings, thoughts and actions toward social objects (Greenwald, McGhee, Schwartz, 1998). In recent advances, the term implicit attitude has been replaced by the notion of arbitrarily applicable relational responding, highlighting the stimulus control as a core aspect of the procedure (Finn, Barnes-Holmes & McEnteggart, 2018).

The role of latency as critical for the identification of implicit relational responses has been strengthened by some experimental findings. It is common that IRAP procedures also involve the comparison of participants' performance between IRAP and explicit measures instruments such as psychological tests and questionnaires. In general, results from explicit and implicit measures differ (Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010; Carpenter et al., 2005; Dawson et al, 2009; Perez et al, 2017, to mention some).

Additionally, IRAP has been shown to be effective as a predictor of nonverbal behaviors related to the content under analysis. In the study performed by Carpenter et al (2012), 25 people who had seek a treatment service for cocaine abuse responded a set of questionnaires and an IRAP before starting treatment. The self-report measures versed about

the effects of cocaine and the personal need for cocaine. The IRAP presented the label terms "with cocaine" and "without cocaine" and, as target stimuli, phrases portraying pleasant experiences (e.g., "I'm sexier", "I'm in control") or unpleasant experiences (e.g., "I'm nervous", "I'm paranoid"). Explicit and implicit measures were correlated with adherence to a six-month treatment. Adherence was measured by the number of relapses (seen in periodic urine tests) and by frequency at clinic sessions. Results in questionnaires did not correlate with adherence to the treatment. In IRAP results, participants who exhibited lower latencies when asked to confirm the combination of "cocaine" + "positive consequences" = "true" and the combination "with cocaine" + "negative consequences" = "false" had the worst indicators of treatment effectiveness lacked face-to-face sessions in the clinic and had a higher frequency of relapse.

In Perez et al. (2017), 35 participants were evaluated for the repertoire of experiential avoidance through three stages: response to questionnaires that assessed experiential avoidance, depression, anxiety and stress, response to an IRAP, and an avoidance task. The IRAP used the label stimuli "okay to feel/remember/think" or "avoid to feel/remember/speak", and target stimuli described pleasant or unpleasant private events. After the procedure participants were separated into two groups depending on whether or not they presented avoidance responses during the avoidance task. Both groups were considered equal according to performance in the questionnaires. But IRAP measures showed difference in the trials that asked participants to confirm the relationship "I avoid feeling / remembering / thinking" + unpleasant events = "true". In the non-avoidance group, latencies in these trials were undifferentiated among consistent and inconsistent blocks. In the avoidance group, responses to this trial were systematically shorter in the consistent blocks, producing a pronounced IRAP effect. These patterns suggest that IRAP correctly identified participants prone to exhibited experiential avoidance.

Performance in IRAP was also predictive of non-verbal behaviors even for relationships trained in the laboratory. In the study of Bortoloti and de Rose (2012), 19 participants were assigned to two training conditions of arbitrary relations for the emergence of equivalence classes: simultaneous matching to sample (SMTS) or delayed matching to sample (DMTS). After training and testing, emerging relationships were further tested via IRAP. Results showed that participants trained with the DMTS (but not those trained via SMTS) responded faster in blocks consistent with the emergent relations than in the inconsistent ones, characterizing the IRAP effect. Difference between blocks was statistically significant ($P = 0.0016$). The findings were interpreted as demonstrating that DMTS training produces "stronger" stimulus relationships than those produced through SMTS training.

In another study that also combined stimulus equivalence and IRAP procedures, Mizael et al. (2016) exposed thirteen children (who, at the baseline, presented negative bias toward photos of black men) to an equivalence relations training to produce the emerging relationship between photos of black men and a positive symbol. The results in the matching procedure demonstrated the emergence of the new relationship, which was confirmed by the IRAP. Thus, results pointed out the absence of the anti-black bias presented in the baseline. The absence was demonstrated by the shorter latencies of responses given to confirm the relation black man photo + nice photo = "true" and black man photo + nasty photo = "false".

Some evidence also indicates that responses to IRAP cannot be easily distorted, even when participants are instructed to fake their performance and are instructed on how to do it (Mc Kenna et al, 2007). For opposing results, see Drake, Seymour and Habib (2016).

Taken together, the IRAP findings suggest that the formation of relations between stimuli can be captured via IRAP and correlated with non-verbal behavior not directly measured by the procedure. From this, the present study aims to propose an experimental procedure for the analysis of moral values by extending the use of IRAP to stimuli of moral

content, comparing the performance of participants through implicit and explicit measures on values.

Method

Participants

Participants were recruited at the University of São Paulo. They are undergraduate, graduate students and technical staff. Participants were selected from two inclusion criteria: be portuguese native and have at least high school diploma. These criteria were applied to ensure that participants were competent to read and understand information in Portuguese. Twenty-four participants completed the study, eleven women and thirteen men aged 18 to 55 years (average 28, 1 years). None of them had prior experience in IRAP research. Written informed consent was obtained prior to participation and no remuneration was offered.

Apparatus, instruments and material

The entire procedure was performed on an HP p6660br desktop computer with Windows 7 home Premium operating system and 15-inch monitor with resolution of 1024x768 pixels. The computer was equipped with Open Source software IRAP, version 0.9.10 made available by Ian Hussey in 2015. The software provides written instructions to the participant informing them of the general characteristics of the task and of the response required, introducing the participant to the label, target and relational terms stimuli and recording, for each trial, the chosen relational term and the latency of the response. Latency is defined as the time, in milliseconds between the onset of stimuli on the computer screen and the response required, which consists of pressing the computer key that selects the relational term. The explicit measure used was the Brazilian version of the MFQ (apêndice A) presented via *googleforms*. The original version of the questionnaire was elaborated by Jesse Graham, Jonathan Haidt and Brian Nosek in 2008. The Brazilian version was created by Júlio Frota

Lisbôa Pereira de Souza and approved by the authors of the original version through back translation procedure in 2015. The questionnaire is composed of eleven questions on socioeconomic aspects, and thirty-two questions in likert scales about moral values. The first sixteen questions dealt with the importance of some characteristics of the situation that the person may take into account in a moral judgment, for example, "if anyone cared about a weak or vulnerable person." Before each proposition the participant should choose a number from one to four ranging from "not very relevant" to "very relevant". The second set of questions presents statements based on the five moral foundations to which the person should express her degree of agreement ranging from 0 (strongly disagree) to 5 (strongly agree).

Procedure

The experiment consisted of three phases: (a) response to the explicit measurement questionnaire; (b) IRAP practice on flowers and insects; (c) IRAP test on moral values. The three phases were applied in a single session and participants took 40 minutes on average to complete their participation. There was no counterbalance in the order of the phases or the IRAP blocks.

Response to MFQ. At this stage, participants responded to the Brazilian version of the MFQ. This phase lasted 15 minutes on average.

Practice with IRAP. The practice had the function of introducing the participant to the IRAP scenario. At this stage the participant was exposed to a fixed number of four practice blocks, with 24 trials each, two blocks of the consistent type and two blocks of the inconsistent type. This IRAP was inspired in the Mc Kenna et al (2007) content procedure. The label stimuli used in this phase were positive and negative adjectives, the target stimuli were names of flowers and insects and the relational terms were "true" and "false." Trials were presented in a quasi-random order that guaranteed that all trials would be presented six

times in each block and two equal trials would never be presented in sequence. The types of trials to which participants were exposed in the practice phase are summarized in Table 1:

Table 1

Trial types presented in practice phase.

Type of Block	Rule	Label Stimulus	Target Stimulus	Required Response
Consistent	“Flowers are positive and insects are negative”	Tulip/Rose/Lily/Daisy	Happy/Friendly/Pretty/Good	“true”
		Spider/Wasp/Maggot/Bee	Pain/Mean/Hostile/Hateful	“true”
		Tulip/Rose/Lily/Daisy	Pain/Mean/Hostile/Hateful	“false”
		Spider/Wasp/Maggot/Bee	Happy/Friendly/Pretty/Good	“false”
Inconsistent	“Flowers are negative and insects are positive”	Tulip/Rose/Lily/Daisy	Pain/Mean/Hostile/Hateful	“true”
		Spider/Wasp/Maggot/Bee	Happy/Friendly/Pretty/Good	“true”
		Tulip/Rose/Lily/Daisy	Happy/Friendly/Pretty/Good	“false”
		Spider/Wasp/Maggot/Bee	Pain/Mean/Hostile/Hateful	“false”

The trial was initiated with the simultaneous presentation of the stimuli. The participant should then give a response in accordance with the current rule in each block by pressing the "e" and "i" keys on the computer keyboard in the shortest possible time. The position of the relational terms was always the same with the term "false" positioned to the left and being triggered by the letter "e" and the term "true" positioned to the right and being triggered by the letter "i".

If the participant responded correctly, the trial was finished and after a 400-second intertrial interval (ITI), a new trial began. If the rule was not respected, a correction procedure took place: A red X appeared between the two relational terms and the triggered key was turned inactive. The trial was only terminated after the participant activated the other available button. During the practice phase, the researcher stayed close to the participant and answered questions about the task and its criteria. Participation in this phase lasted an average of 5 minutes.

IRAP test on moral values. In the third phase participants responded to an IRAP on moral values. In this IRAP, the label stimuli used were the terms "I am for" and "I am against"(respectively "Eu sou a favor" e "Eu sou contra" in the original version). The target stimuli used were adapted versions of the six statements about fairness and purity taken from the MFQ. The statements were adapted to maintain the same meaning of the original statement, but with a similar number of words among statements. The relational terms were the words "true" and "false," arranged on the screen just as in the practice phase. The combination among label and target stimuli generated twelve different trial types (see table 2).

We chose to work with only two out of the five moral values addressed in the MFQ, since the required number of trials to evaluate five different categories would make the procedure very extensive. The values of purity and fairness were chosen among the five possibilities, since, according to the MFT, these two values are commonly observed in opposite valences, that is, the greater the agreement with one, the smaller the agreement with the other. Thus, the IRAP procedure would provide a confirmation or otherwise of the theory to the extent that it was capable of capturing opposing biases in these categories.

Participants responded to 8 blocks of 24 trials each, being 4 blocks of the consistent type and 4 blocks of the inconsistent type. No correction procedure was used. Before the consistent blocks, the software presented the following instruction: "This is a consistent

block. Respond it indicating if you agree or not with the information presented”. In the inconsistent blocks, the instruction was: “This is an inconsistent block. Respond the opposite of what you think of the information presented”.

Table 2

Trial types presented in test phase

Type of Block	Trial Type	Label Stimulus	Target Stimulus
Consistent or Inconsistent	Fai 1	“I’m for/against”	"Justice is the most important requirement for a society."
	Fai 2	“I’m for/against”	"Everyone should be treated fairly."
	Fai 3	“I’m for/against”	"Rich children inherit a lot of money while poor children inherit nothing."
	Pur 1	“I’m for/against”	"People should not do things that are disgusting."
	Pur 2	“I’m for/against”	"Some acts are wrong on the grounds that they are unnatural."
	Pur 3	“I’m for/against”	"Chastity is an important and valuable virtue."

The trial randomization and the ITI were identical to those in the practice phase. This phase lasted on average 20 minutes. Because there were no pre-established rules for the

response, each block had 12 types of trials and these were identical in both blocks. Table 2 shows the stimulus combinations presented to the participant during the test.

In the practice phase we opted to use stimuli different from those in the test phase. In addition, the correction procedure is present in practice, but not in the test. These decisions were taken based on pilot studies that showed that a more structured practice phase (rules and correction) and with simpler stimuli (a single word as opposed to the phrases used in the test phase) makes the practice phase shorter and is sufficient for the participant to show a correct performance on the task and present coherence in the responses in the test phase, that is, to show differing patterns according to the type of block. In the test, information on whether the participant agrees or not with a given set of stimuli was verified by the frequency with which she chose a given relational term in relation to a specific pair, for example: "I am for" and "fairness is the most important prerequisite for a society" and whether it systematically varied the choice of this relational term as a function of the type of block.

This procedure was previously approved by the Human Research Ethics Committee of the University of São Paulo in the process number 82303318.5.0000.5561.

Data analysis

MFQ. Only the answers on the degree of agreement with the fifteen statements were analyzed. The degree of agreement could vary between 0 (respond "strongly disagree" on the three items of the same moral basis) and 15 (respond "strongly agree" on the three items of the same moral basis). It was calculated as a group mean for each of the five moral bases. Means above 7.5 were deemed as agreement bias, while means below 7.5 were deemed as disagreement bias.

IRAP. Two types of analyzes were conducted from the IRAP data. In individual analyzes the each participant's mean latencies were calculated for each of the twelve trial types. For the group analysis, D – IRAP score was calculated for four types of trial: Fai +

(agreement with fairness statements), Fai – (disagreement with fairness statements), Pur + (agreement with purity statements) and Pur – (disagreement with purity statements). Only the data from the test phase were analyzed.

The D-IRAP score was obtained by the steps shown in Table 3.

Table 3

Method to convert mean latencies from each participant in D – IRAP scores

-
- (1) Only response-latency data from test blocks are used;
 - (2) Latencies above 10,000ms are eliminated from the analysis;
 - (3) The data are eliminated for a participant for whom more than 10% of test-block trials have latencies less than 300ms;
 - (4) Is one participant produce less than 80% of coherence in the choosing of the relational term during the entire test phase, his data are eliminated from the dataset;
 - (5) Compute mean latencies for each trial type (Fai +, Fai -, Pur + e Pur -);
 - (6) Compute the standard deviation for each trial type;
 - (7) For each category, compute the difference scores by subtracting the mean latency of each consistent trials from the mean latency of the corresponding inconsistent trials;
 - (8) Divide each difference score by its corresponding standard deviation generating four D – IRAP scores.
-

After applying the method to convert the latencies into D – IRAP scores, four participants were eliminated from analyzes. The following results comprises only the twenty remaining participants.

A typical analysis of data produced with IRAP describes only the mean D – IRAP scores of the group and the stimuli of the same category are treated as similar. That is, two scores are calculated: one from the consistent block and the other from the inconsistent block to all statements about fairness, and two others for all statements on purity. However, in this

study, we also performed an analysis of the individual performances from the average latencies of the answers for each of the twelve types of trials. The latencies were chosen as data of analysis of the individual performance because they are data without normalization (different from the score D – IRAP) and as so, express more clearly the individual differences. This is important in this study because no upper limit was set for the response emission, so that the individual mean latencies comprises great variation. Moreover, in the individual analyzes, each statement was considered as a different stimulus, what is not possible by the typical aggregated analyzes.

From the two types of results, t-tests of independent samples were performed to verify if the mean data of the group, expressed by the D-IRAP score, adequately represents the observed trends at the individual level and whether the statements about fairness and purity may or may not be considered similar to each other.

Comparison between explicit and implicit measures. the results obtained with the questionnaire and IRAP were compared using the mean frequency of agreement (questionnaire) and mean agreement bias (IRAP). Answers to the questionnaire corresponding to numbers 0, 1 and 2 were classified as disagreement and the answers corresponding to numbers 3, 4 and 5 were classified as concordance. The IRAP responses were classified as agreement bias when the mean latencies were shorter in the consistent block to confirm the relation "I am for" + affirmation about fairness or purity = "true." Following the same reasoning, bias classification of disagreement was applied when mean latencies were shorter in the consistent block for relations "I am against" + affirmation about fairness or purity = "true." In cases where the difference between mean latencies in consistent and inconsistent blocks was undifferentiated, the applied classification was absence of bias. Responses that demonstrated absence of bias were considered as dissonances in relation to the explicit measure since in the MFQ there were no "non-agree or disagree" responses. From this

classification were calculated the mean of agreement for each category and for each procedure.

Results

Figure 1 presents the mean scores of degree of agreement of the participants for each of the five moral foundations.

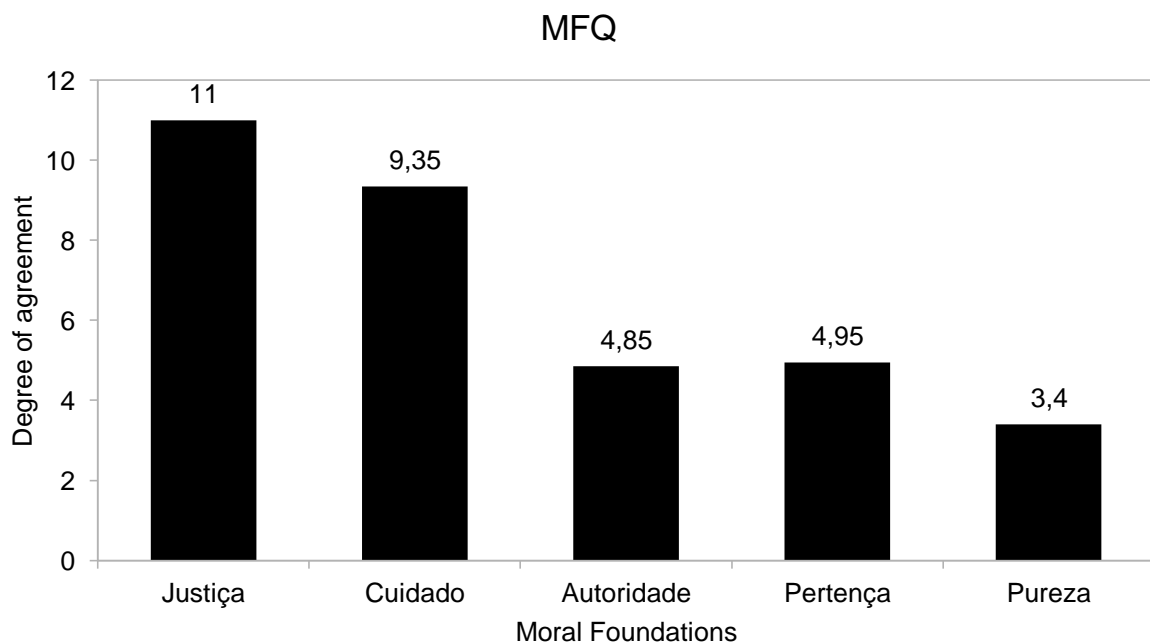


Figure 1. Group means of agreement for each of the five foundations in MFQ.

Among the 20 participants, it was observed that the moral foundation with the highest degree of agreement was fairness, with 11 points and soon after care, with 9.35 points. The moral foundation with the lowest degree of agreement was purity with a mean of 3.4 and then authority with 4.85 and Ingroup with 4.95. The data obtained follow the trend pointed out in the literature based on the MFT (Haidt & Graham, 2007): fairness and care are values that tend to covariate together and the same is valid to the values of authority, ingroup and purity, forming two clusters. In addition, the fairness/care pair and the trio authority/ingroup/purity tend to have negative correlation. If the participant has higher scores in the fairness/care

cluster, she tends to have lower scores on the other cluster and vice versa. Our results showed that only a single participant did not fit this prediction. Participant 20 presented his highest scores on authority (8) and his lowest score was on item care (0). For all participants the agreement with the items of fairness was much greater than the agreement with the items of purity.

Just as in the MFQ, in the IRAP there was a general tendency to chose “true” given the combination of “I’m for + statement or fairness” and to chose false otherwise in consistent blocks. Thus, shorter latencies in the consistent blocks would indicate that explicit and implicit measure captured the same trends.

Figures 2 and 3 show the individual mean latencies for each of the six statements in the two block types. Figure 2 shows the average latencies of each participant for each of the affirmations of the fairness category in the consistent blocks (black color) and inconsistent blocks (gray color). For the first statement, nine participants (2, 3, 4, 7, 8, 9, 12, 13 and 15) presented lower latency in the consistent block. That is, in the combination "I am in favor" + an statement on fairness, participants responded faster for "true" on the consistent block. Five participants (1, 6, 10 17 and 19) presented almost identical latencies in both blocks, indicating absence of bias given this statement. Six participants (11, 12, 14, 16, 18 and 20) presented lower latencies in the inconsistent block, indicating bias of disagreement on the statement.

The aggregate mean was 3.900 milliseconds in the consistent block and 4.270 milliseconds in the inconsistent block, which, on a group level, would be indicative of bias of agreement with the first affirmative statement on fairness.

For the second statement, twelve participants (1, 2, 6, 7, 8, 9, 11, 13, 14, 16, 17 and 20) presented lower mean latency in the consistent block. Participants 3 and 4 presented almost identical latencies, and participants 5, 10, 12, 15, 18 and 19 presented lower mean latency in

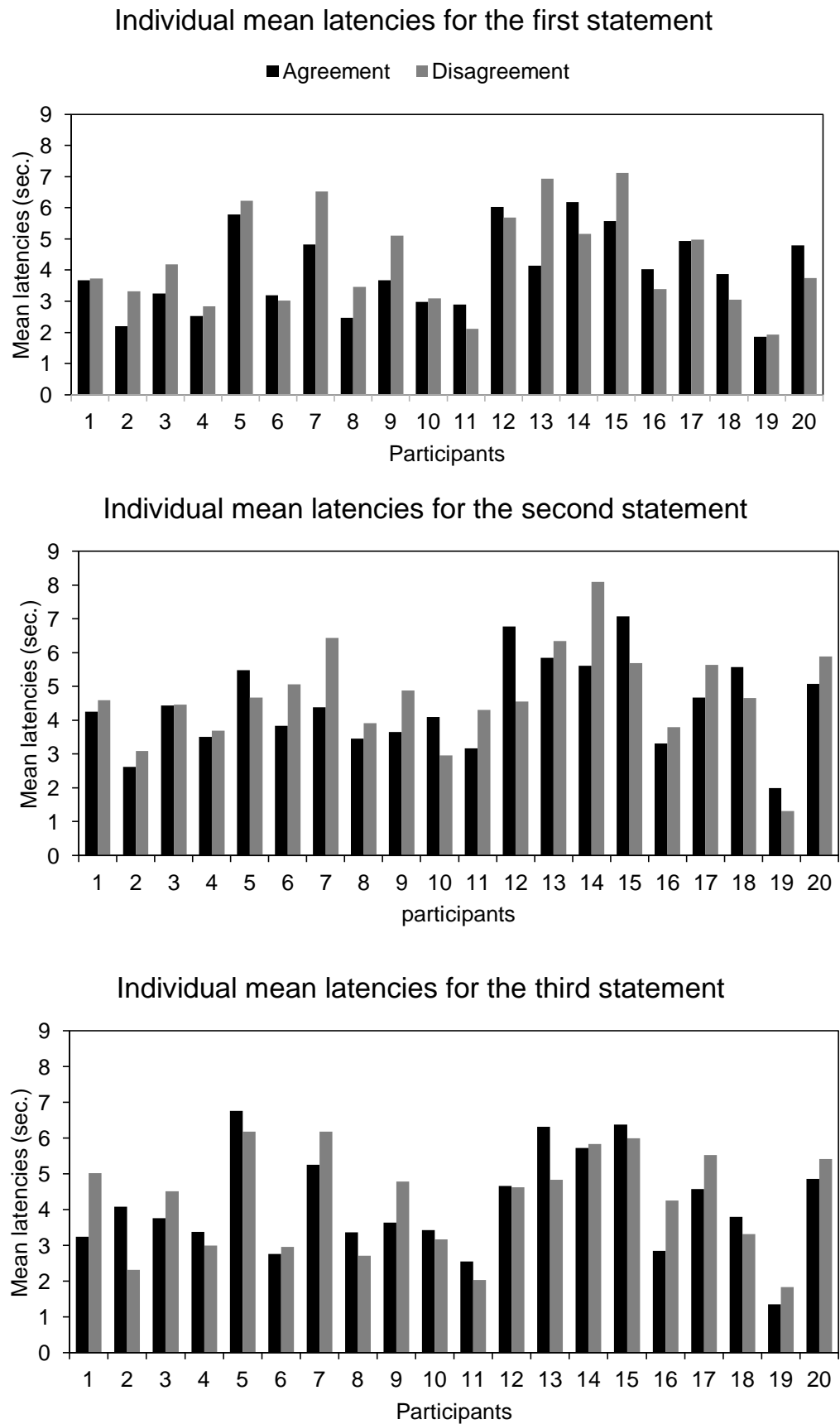


Figure 2. Individual mean latencies for statements on fairness.

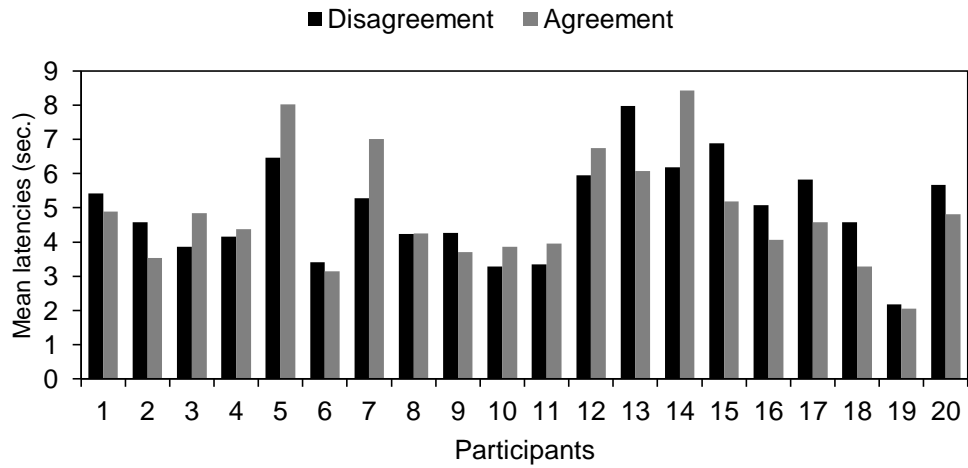
the inconsistent block, indicating disagreement with the affirmative. The aggregate mean was 4.400 milliseconds in the consistent block and 4.700 milliseconds in the inconsistent block, which, on a group level, would be indicative of agreement with the second statement about fairness.

In the third statement, nine participants (1, 3, 6, 7, 9, 16, 17, 19 and 20) presented lower latency in the consistent block. Participants 12 and 14 presented similar latencies in both blocks and the remaining nine participants (2, 4, 5, 8, 10, 11, 13, 15, and 18) presented lower latency in the inconsistent block. The aggregate mean was 4.100 milliseconds in the consistent block and 4.200 in the inconsistent block. At the group level, these averages suggest absence of agreement or disagreement bias.

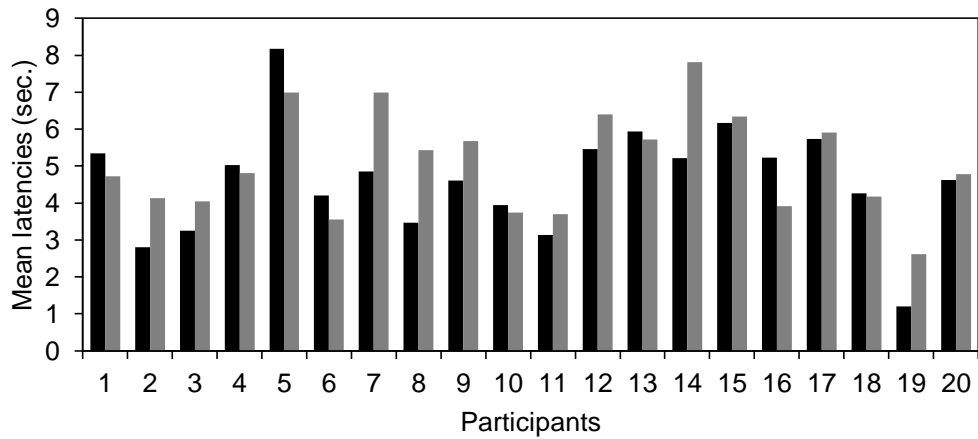
Figure 3 shows the average latencies of each participant for the three statements about purity in the consistent and inconsistent blocks. In the first statement on purity, eight participants (3, 4, 5, 7, 10, 11, 12, and 13) demonstrated disagreement, indicated by the lower mean latency in the consistent block. Participant 8 presented indifferenciation in the latencies between the blocks. Eleven participants (1, 2, 6, 9, 13, 15, 16, 17, 18, 19 and 20) presented lower latency in the inconsistent blocks, suggesting agreement with this item. The aggregate means of latencies were 4.900 for the consistent block and 4.800 for the inconsistent block, which, at group level, would be indicative of indifferenciation.

For the second assertion about purity, 12 participants (2, 3, 7, 8, 9, 11, 12, 14, 15, 17, 19 and 20) presented disagreement, illustrated by the lower mean latency in the consistent block. The remaining eight participants (1, 4, 5, 6, 10, 13, 16, and 18) presented smaller latencies in the consistent block, indicating agreement with the affirmative. The aggregate averages of the latencies were 4.600 milliseconds in the consistent block and 5.000 in the inconsistent block, which would indicate disagreement with the affirmative at the group level.

Individual mean latencies for the fourth statement



Individual mean latencies for the fifth statement



Individual mean latencies for the sixth statement

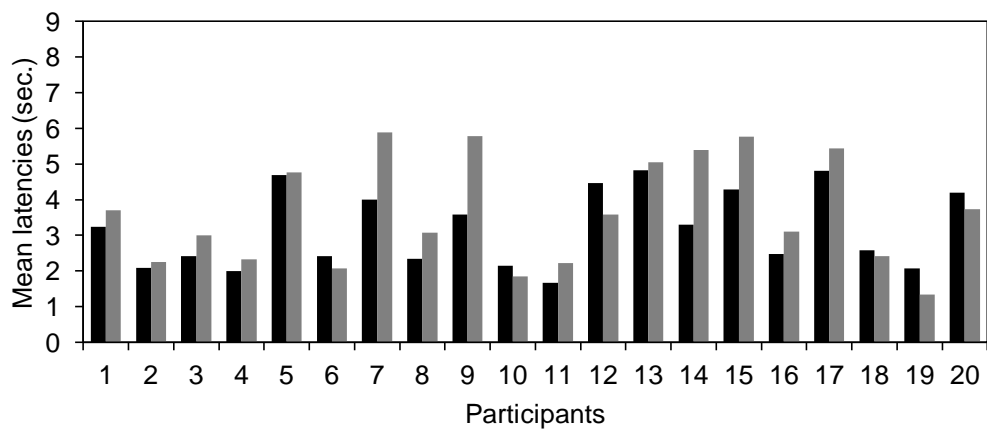


Figure 3. Individual mean latencies for statements on purity

On the Statement three, 13 participants (1, 2, 3, 4, 7, 8, 9, 11, 13, 14, 15, 16 and 17) presented lower mean latency in the consistent block, indicating disagreement with the item. Participant 5 presented undifferentiated means between the blocks. Six participants, 16, 10, 12, 18, 19 and 20 presented smaller latencies in the inconsistent blocks, suggesting agreement with the affirmative. Group mean was 3.100 milliseconds in the consistent block and 3.600 milliseconds in the inconsistent blocks, suggesting disagreement with the affirmative.

Figure 4 presents results based on D-IRAP score. The columns "Fai +" and "Pur +" refer to answers given in the consistent blocks, indicating agreement with fairness and disagreement with purity. The columns "Fai-" and "Pur-" refer to the responses to inconsistent blocks.

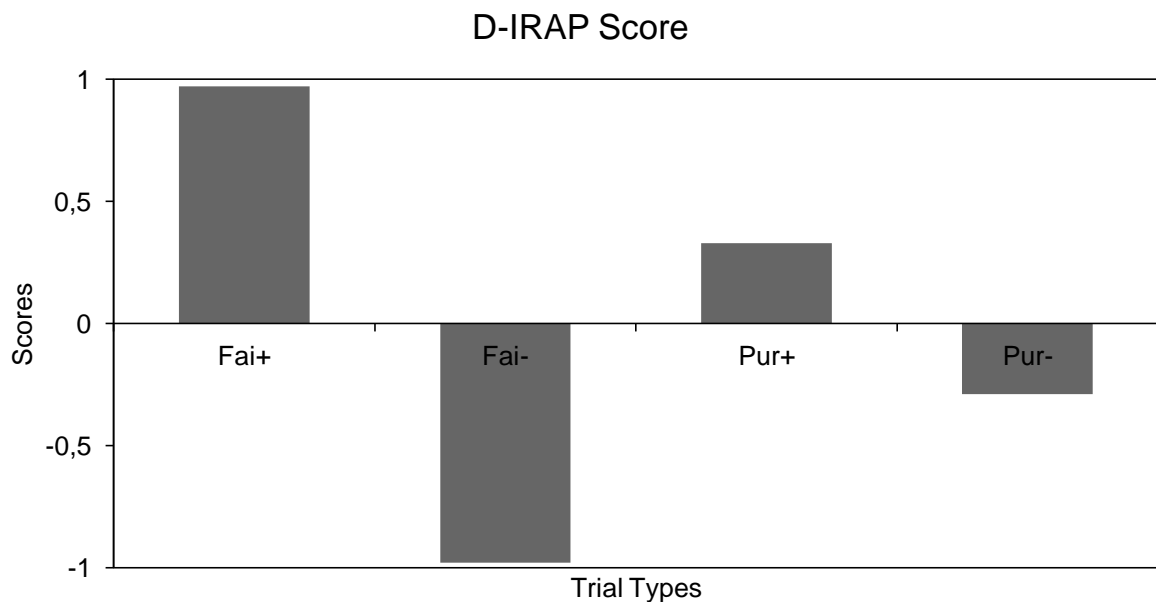


Figure 4. D – IRAP scores for each one of the trial types

The Figure shows that participants demonstrated agreement with fairness (Fai +) and disagreement with purity (Pur +) in both consistent and inconsistent blocks. That is,

participants tended to respond "true" more quickly than "false" when they were introduced to the pair "I am for" + statement about fairness or "I am against" + an statement about purity. On the other hand, when presented to a "I am against" + statement about fairness or "I am for" + affirmation about purity, they tended to respond "false" more quickly. *t* Tests confirmed that the four scores are significantly different from zero ($t(20) = 3.57, p = 0.38$); the scores Fai + and Fai- are statistically equal to each other ($t(20) = 0.492, p = 0.492$), as well as Pur + and Pur - ($t(20) = 0.58, p = 0.811$); and that the difference between the scores of fairness and purity was statistically significant (for the Fai + and Pur + scores, $t(20) = 3.904, p = 0.055$). So, participants were more concordant to fairness statements than discordant to purity statements.

Comparing individual and group results, it can be observed that group result via D – IRAP scores did not capture the variation of the observed trends in the individual mean latencies. While group data expresses general agreement with the items in the fairness category and a general disagreement with items of the purity category, the individual analyzes presented biases of discordance present in all items of the category fairness and biases of agreement in all items in the purity category.

Supporting this finding, a repeated measures ANOVA was performed for each participant comparing trial type (label stimulus + target stimulus) and mean latencies in order to verify if the statements of the same category could be considered different in themselves. Results indicated that stimuli of the same category were significantly different from each other for 18 of the 20 participants analyzed.

Figure 5 show the comparison among frequency of agreement within the six items in the questionnaire and the agreement displayed in IRAP for each of the statements.

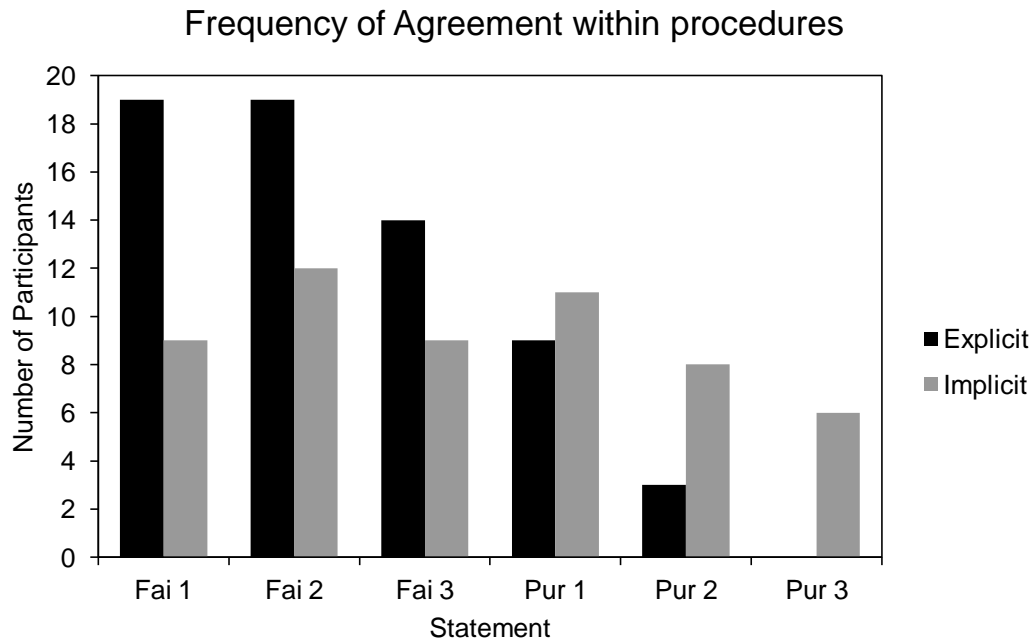


Figure 5. Frequency of agreement within the six statements in the questionnaire (explicit) and IRAP (implicit).

It can be observed that the group tended to agree more with the category of fairness and disagree more with the category of purity in the questionnaire than in IRAP. For fairness statements, the frequency of agreement was 95% in questionnaire, and 51% in IRAP. For purity statements, the frequency of disagreement was 80% in questionnaire and 55% in IRAP.

Individual comparisons showed that none of the participants presented complete similarity among responses to the questionnaire and responses to IRAP. For two participants (3 and 8) there was dissonance between responses in only one item. For seven participants (2, 6, 9, 13, 14, 15 and 17), two of their answers expressed in the questionnaire presented opposite patterns of agreement with the responses to IRAP. For four participants (1, 4, 11 and 12), three responses were dissonant. For four other participants (4, 16, 19 and 20), four responses were dissonant. For two participants (5 and 18), five of the six responses to the questionnaire are opposite to those obtained in IRAP. And for one participant (10), all six of his responses to the questionnaire are opposed to the biases of agreement observed in IRAP.

Discussion

The purpose of the present study was to use different procedures to verify how revealing the moral values of fairness and purity they are. Explicit and implicit measures pointed to different results. Results obtained via questionnaire showed a trend already observed in the results obtained with the North American population: high agreement with items on fairness correlated with high disagreement of the items on purity. IRAP, on the other hand, showed biases of agreement with fairness and disagreement with purity close to 50% of the universe of responses, which contradicts the predicted negative correlation between these two values in the MFQ literature.

One way of analyzing this result is to consider that different procedures, despite having the same content, capture different phenomena. The answer to the questionnaire is purely verbal. It can be analyzed in terms of control by the verbal community which reinforces the verbalization of certain values and also with respect to the reactivity to the verbal procedure, since in questionnaires responses can be easily distorted. The response to IRAP is measured by latency, used as an indication of the degree of relationship among stimuli, and cannot be easily distorted (Power et al, 2009) even when the participant is directly instructed to do so (Mc Kenna et al, 2007).

Another important concern is whether the results obtained are in fact informative about the moral behavior of these participants beyond what was observed in this study. The explicit measures are recognized in the literature as instruments of great limitation when it comes to predicting behavior, especially when dealing with socially controversial issues (Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2012). A growing number of studies comparing explicit and implicit measures with a third measure of nonverbal behavior have consistently demonstrated that performance on IRAP (but not in verbal reports) has been predictive of related non-verbal behavior (Carpenter et al, 2012; Nicholson & Barnes-Holmes, 2012; Perez

et al., 2017). In this study, however, it is not possible to conclude that one or another instrument would have greater predictive value on the values of fairness and purity.

The different results suggest that the type of procedure used in the measurement of moral values can be a critical variable in the investigation of this phenomenon. It is possible to draw a parallel between the differences captured by the two instruments and the findings of research based on the dual model of cognitive processing, which classifies the decision-making process into intuition and deliberation (Kahneman, 2011). Studies that manipulate time pressure to respond in social situations consistently find a pattern of higher frequency of intuitive responses when there is temporal pressure and opposition to the greater frequency of deliberate responses in the absence of temporal pressure (Suter & Hertwig, 2011; Rand, Greene & Nowak, 2012).

The IRAP performed in this study presented some methodological changes in comparison to the traditional application of this procedure. The stimuli of the practice phase were different from those of the test phase and no correction procedure was used in the test phase. The data obtained demonstrate that these changes did not adversely affect participants' performance during the test. The practice phase was sufficient to familiarize the participant with the task, since during the test 22 of the 24 participants produced coherence in the choosing of relational terms in each block type above 80%

Results also indicate that in this IRAP it was accurate to consider target stimuli of the same category as being different. In the individual analyzes, both through visual and statistical inspection, it was demonstrated that the target stimuli were different from each other so that even items of the same category correlated with sometimes opposite biases of agreement. Therefore, an analysis of the $D - IRAP$ score with group means was uninformative of individual performances.

A possible explanation for the difference observed between individual analyzes and group analyzes would be precisely that the moral statements on fairness and purity presented here comprise different stimuli between them, so that a bias of agreement with one statement in a category does not imply an identical agreement for all items included in the category. In studies with IRAP, traditionally, the selection of stimuli from each category presupposes that they can be treated as if they were the same stimulus. In this study, this was not done, but, nevertheless, the results obtained seem to indicate that this does not preclude the use of IRAP, it will only require independent analyzes for each stimulus.

The procedure applied here did not follow the recommendation of Barnes-Holmes, Barnes-Holmes, Stewart and Boles (2012) to establish a maximum time limit for the response. The definition of the appropriate time interval would be difficult, since the number of words presented as stimuli in the present study was higher than that commonly observed in studies with IRAP, so that it does not seem appropriate to apply the interval of 2 or 3 seconds as observed in other procedures with IRAP. Future studies using the same stimuli applied here may consider the average range of participants presented here as a basis for establishing this limit.

The understanding of moral behavior in terms of relational responding may be a useful alternative to verbal procedures. Yet, to reliably evidence differences between procedures it is still necessary to compare results to non-verbal behavior related to moral contents. Studies with IRAP and the comparison between different verbal and nonverbal measures of behavior can bring important contributions in the identification of controlling variables for moral behavior.

References

- Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A Sketch of the implicit relational assessment procedure (IRAP) and the relational elaboration and coherence (REC) model. *The Psychological Record*, 60(3), 527-542. doi: 10.1007/BF03395726
- Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y. & Stewart, I. (2010). The implicit relational assessment procedure: exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white irish individuals. *The Psychological Record*, 60, 43-66. doi:10.1007/BF03395694
- Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, P. (2015). Moral judgment and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of judgment and decision making*. Chichester, UK: Wiley. doi: 10.1002/9781118468333.ch17
- Bortoloti, R., & de Rose, J. (2012). Equivalent stimuli are more strogly related after training with delayed matching than after simultaneous matching: a study using implicit relational assessment procedure (IRAP). *The Psychological Record*, 62(1), 41-54. doi: 10.1007/BF03395785
- Carpenter, K. M., Martinez, D., Vadhan, N. P., Barnes-Holmes, D., & Nunes, E. V. (2012). Measures of attentional bias and relational responding are associated with behavioral treatment outcome for cocaine dependence. *The American Journal of Drug and Alcohol Abuse*, 38(2), 146-154. doi: 10.3109/00952990.2011.643986
- Curry, O. S. (2004). *Morality as natural history: an adaptationist account of ethics* (doctoral dissertation). Retirado de core.ac.uk/download/pdf/40091549.pdf

- Cushman, F. A., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, *35*(6), 1052–1075. doi: 10.1111/j.1551-6709.2010.01167.x
- Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J., & Gore, N. J. (2009). Assessing the implicit beliefs of sexual offenders using the implicit relational assessment procedure. *Sexual abuse: a journal of research and treatment*, *21* (1), 57-75. doi: 10.1177/1079063208326928.
- de Waal, F. (2006). *Primates and philosophers: how morality evolved*. Princeton (NJ): Princeton University Press.
- Dittrich, A. (2010). Ética e comportamento. In E. Zagury Tourinho e S. Vasconcelos Luna (eds.) *Análise do Comportamento: investigações históricas, conceituais e aplicadas*, (pp. 37-59), São Paulo: Roca.
- Drake, C. E., Kellum, K. K., Wilson, K. G., Luoma, J. B., Weinstein, J. H., & Adams, C. H. (2010). Examining the implicit relational assessment procedure: four preliminary studies. *The Psychological Record*, *60*(1), 60-81. doi:10.1007/BF03395695
- Drake, C. E., Seymour, K. H., & Habib, R. (2016). Testing the IRAP: exploring the reliability and fakability of an idiographic approach to interpersonal attitudes. *The Psychological Record*, *66*, 153-163. doi: 10.1007/s40732-015-0160-1
- Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2018). Exploring the single-trial-type-dominance-effect in the IRAP: developing a differential arbitrarily applicable relational responding effects (DAARRE) model. *The Psychological Record*, *68*(1), 11-25 doi:10.1007/s40732-017-0262-z

- Garret, R. (1979). Value conflict in a skinnerian analysis. *Behaviorism*, 7(1), 9-16. Retrieved from www.jstor.org/stable/27758926
- Gleicherricht, E., Torralva, T., Rattazzi, A., & Marengo, V. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *SCAN*, 8(7), 780–788. doi: 10.1093/scan/nss067
- Graham, G. (1977). On what is good: a study of B. F. Skinner's operant behaviorist view. *Behaviorism*, 5(2), 97-112. Retrieved from www.jstor.org/stable/27758896
- Graham, J. & Haidt, J. (2010). Beyond beliefs: religions bind individuals into moral communities. *Personality and Social Psychology Review*, 14, 140-150. doi: 10.1177/1088868309353415
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., & Ditto, P. H. (2011). Mapping the moral domain. *Perspectives in Social Psychology*, 101(2), 366–385. doi: 10.1037/a0021847
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480. doi: 10.1037/0022-3514.74.6.1464
- Haidt, J., & Graham, J. (2007). When Morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Social justice research*, 20 (1), 98-116. doi: 10.1007/s11211-007-0034-z
- Haidt, J., Graham, J., & Joseph, C. (2009). Above and below left-right: ideological narratives and moral foundations. *Psychological inquiry*, 20(2-3), 110-119. doi: 10.1080/10478400903028573

- Haidt, J. and Hersh, M. A. (2001), Sexual Morality: The Cultures and Emotions of Conservatives and Liberals. *Journal of Applied Social Psychology*, 31, 191–221. doi:10.1111/j.1559-1816.2001.tb02489.x
- Harris, M. (1974). *Cows, pigs, wars and witches: the riddles of culture*. New York: Random House.
- Harris, M. (1979). *Cultural materialism: the struggle for a science of culture*. New York: Random House.
- Hocutt, M. (1977). Skinner on the word “good”: a naturalistic semantics for ethics. *Ethics*, 87(4), 319-338. Retrieved from www.jstor.org/stable/2379901
- Hocutt, M. (2013). A behavioral analysis of morality and value. *The Behavior Analyst*, 36(2), 239-249. doi: 10.1007/bf03392310
- Houmanfar, R., & Rodrigues, N. J. (2006). The metacontingency and the behavioral contingency: points of contact and departure. *Behavior and Social Issues*, 15(1), 13-30. doi: 10.5210/bsi.v15i1.342
- Hussey, I., Mhaoileoin, D. N., Barnes-Holmes, D., Ohtsuki, T., Kishita, N., Hughes, S., & Murphy, C. (2016). The IRAP is nonrelative but not acontextual: changes to the contrast category influence men’s dehumanization of women. *The Psychological Record*, 66(2), 291–299. doi:10.1007/s40732-016-0171-6
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus And Giroux.
- Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: how five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality*, 46(2), 184-194. doi:10.1016/j.jrp.2012.01.006

- Laland, K. N., & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology*, *12*, 150-159. doi: 10.1002/evan.10111
- Leigland S. (2005). Variables of which values are a function. *The Behavior analyst*, *28*(2), 133–142. doi:10.1007/bf03392109
- Mizael, T. M., de Almeida, J. H., Silveira, C. C., & de Rose, J. C. (2016). Changing racial bias by transfer of functions in equivalence classes. *The Psychological Record*, *66*(3), 451-462. doi: 10.1007/s40732-016-0185-0
- McKenna, I. M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology & Psychological Therapy*, *7*(2), 253–268. Retrieved from: www.redalyc.org/pdf/560/56070209.pdf
- Paunonen, S. V., & LeBel, E. P. (2012). Socially desirable responding and its effects on the validity of personality assessments. *Journal of personality and social psychology*, *103* (1), 158-175. doi: 10.1037/a0028165
- Perez, W. F., Kovac, R., Linares, I. M. P., Barbosa, S. F. U., Gomes, C. T., dos Santos, G. A. R., & de Almeida, J. H. (2017). Utilizando o AAQ-II e o IRAP para prever a ocorrência de esquiva experiencial. *Revista Brasileira de Terapia Comportamental e Cognitiva*. *19*(3), 34-48. doi: 10.31505/rbtcc.v19i3.1052
- Power, P., Barnes-Holmes, D., Barnes-Holmes, Y. & Stewart, I. (2009). The implicit relational assessment procedure (IRAP) as a measure of implicit relative preferences: a first study. *The Psychological Record*. *59*, 621-640. doi:10.1007/BF03395684

- Rand, D., G., Greene, J., D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489, 427-430, doi:10.1038/nature11467
- Rottschaefer, W. (1980). Skinner's science of value. *Behaviorism*, 8(2), 99-112. Retrieved from www.jstor.org/stable/27758957
- Skinner, B. F. (1953). *Science and human behavior*. Oxford: McMillan.
- Skinner, B. F. (1957). *Century psychology series. Verbal behavior*. Appleton-Century-Crofts. doi:10.1037/11256-000
- Skinner, B. F. (1961). Freedom and the control of men. In R. M. Elliott (ed) *Cumulative Record*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Penguin Books.
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119(3), 454–458. doi: 10.1016/j.cognition.2011.01.018
- Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge, MA: Harvard University Press.
- Waller, B. (1982). Skinner's two stage value theory. *Behaviorism*, 10(1), 25-44. Retrieved from www.jstor.org/stable/27758993
- Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and psychological measurement*, 69(4), 548-565. doi: 10.1177/0013164408324469

Capítulo 3

Religiosidade como preditora de valores morais de justiça e pureza em medidas explícitas e implícitas

No capítulo anterior os resultados apresentados demonstram que medidas implícitas e explícitas sobre os valores morais de justiça e pureza apontam resultados diferentes. A medida explícita indicou grande concordância para itens sobre justiça e grande discordância para itens sobre pureza entre os participantes testados. Este resultado foi consonante com previsões da literatura da MFT de que valores de justiça e pureza tendem a ser negativamente correlacionados (Haidt & Graham, 2007; Graham et al, 2011; Graham et al, 2012). A medida implícita, por sua vez, apontou médias de latência indicativas de concordância semelhantes para os dois valores.

Todavia, o procedimento realizado não permite garantir qual das medidas apresenta resultados mais preditivos do comportamento moral dos participantes. As conclusões sobre o procedimento são baseadas nas limitações conhecidas sobre as medidas explícitas (Paunonen & LeBel, 2012; Power et al, 2009; Ziegler & Buehner, 2009) e na premissa de que tempos de reação menores são indicadores de relação entre estímulos (Finn et al, 2016).

Estudos com o IRAP têm utilizado algumas estratégias metodológicas a fim de verificar o valor preditivo de medidas explícitas e implícitas (Carpenter et al, 2012; Dawson et al, 2009; Hooper et al, 2010; Leech et al, 2016; Rönspies et al, 2015; Vahey et al, 2009; Perez et al, 2017). Uma dessas estratégias é realizar comparações entre grupos cujas histórias ou características de interesse sejam conhecidas do experimentador. Dawson et al (2009), por exemplo, compararam os desempenhos de homens condenados por ofensa sexual contra crianças e estudantes universitários em medidas explícitas e implícitas sobre sexualidade e infância. Todos os participantes responderam a um questionário que visava a mensurar distorções cognitivas sobre sexualidade e infância e um IRAP no qual o estímulo rótulo eram as palavras “adulto” e “criança” e os estímulos alvo foram: “sexual”, “não sexual”, “glamour”, “puro”, “sedutor”, “inocente”, “excitante”, “não excitante”, “erótico”, “não erótico”, “sexualmente consciente” e “sexualmente não consciente”. Os resultados mostraram

que de acordo com o questionário os dois grupos estatisticamente idênticos, apresentando pequeno grau de distorção cognitiva. Já nos resultados do IRAP, o grupo de estudantes universitários mostrou alto efeito IRAP ao confirmar as combinações “adulto” e “sexual” e “criança” e “não sexual” enquanto o grupo de agressores sexuais mostrou alto efeito IRAP para a combinação “adulto” e “sexual” mas teve um efeito IRAP próximo a zero na combinação “criança” e “sexual” indicando que não há diferença entre a criança ser considerada sexual ou não sexual. Estes achados mostram que o desempenho no IRAP, mas não a resposta ao questionário, permitia a diferenciação entre os grupos.

Outros estudos que manipularam a variável história pré-experimental têm produzido achados semelhantes. Quando grupos pré-experimentalmente diferentes são expostos a medidas explícitas e implícitas sobre um determinado tema, o IRAP foi capaz de captar diferenças entre os grupos enquanto as medidas explícitas não o fazem (Hooper et al, 2010; Rönspies et al, 2015; Vahey et al, 2009).

Outra estratégia metodológica utilizada com esta finalidade é a comparação entre medida explícita e implícita e o desempenho em uma tarefa não-verbal relacionada ao conteúdo. No estudo de Nicholson e Barnes-Holmes (2012), os participantes responderam a dois IRAPs baseados em dois constructos distintos sobre nojo. Os resultados foram comparados com escalas sobre nojo, depressão, ansiedade, estresse e obsessão-compulsão e com uma tarefa não verbal que consistia a mensuração do grau de aproximação física a estímulos que evocam nojo. Os resultados mostraram que apenas um dos IRAPs se correlacionou com o desempenho na tarefa de aproximação. Além disso, o desempenho nos questionários não foi preditivo sobre o desempenho na tarefa. Outros estudos com este delineamento têm obtido resultados semelhantes, apontando a medida implícita como preditora mais acurada do desempenho na tarefa comportamental (Carpenter et al, 2012; Hooper et al, 2010; Leech et al, 2016; Perez et al, 2017).

Estudos dedicados à mensuração de valores morais, com destaque para aqueles orientados pela MFT (Haidt & Graham, 2007; Haidt & Joseph, 2008; Haidt et al, 2009), têm investigado correlações entre uma ou algumas das cinco bases morais e opinião sobre questões culturais. Uma das relações mais exploradas pelos pesquisadores embasados na MFT é entre bases morais e posição político-partidária (Haidt & Graham, 2007; Graham et al, 2011; Koleva et al, 2012).

As conclusões de Haidt & Graham (2007) sobre as diferentes bases morais que orientam a conduta humana se baseiam nos estudos Shweder (1990) que aponta três grandes tendências da regulação moral. De acordo com Shweder, valores de justiça e cuidado estão relacionados a uma ética da autonomia, respaldando condutas voltadas para a proteção da pessoa e de seu direito de exercer preferências pessoais. Os valores de autoridade e pertença ao grupo estariam relacionados a uma ética da comunidade, que valoriza condutas voltadas à proteção de instituições sociais. Os valores de pureza estariam relacionados a uma ética da divindade. Esta ética de baseia na existência de Deus ou deuses e de uma alma que habita os corpos humanos. O objetivo de uma regulação moral seria proteger a alma por meio da regulação dos corpos.

A partir disso, Haidt e Hersh (2001) e Haidt e Graham (2007) propõem uma relação entre tipos de ética e posicionamento político. O *cluster* justiça e cuidado, base da ética da autonomia, seria subjacente a visões políticas progressistas liberais que visam à garantia de direitos e proteções individuais. Uma postura política conservadora combina os três tipos de ética estaria apoiada na valorização das cinco bases morais. Esta predição foi confirmada por *surveys* posteriores que, adicionalmente, atrelaram valores conservadores à vivência de religiosidade cristã (Graham et al, 2011).

Outros estudos buscaram estabelecer correlações entre concordância com alguma ou algumas bases morais, posição político partidária e a opinião sobre temas polêmicos. Em

Koleva et al (2012), o MFQ foi aplicado junto a questionários que mensuravam a concordância com temas sociais polêmicos. Os resultados apontaram que os participantes que se declararam como religiosos praticantes de alguma fé cristã apresentaram altos escores de concordância com a base moral da pureza e que esta concordância foi preditora de desaprovação de condutas como sexo casual, uso de pornografia, relacionamentos entre pessoas do mesmo sexo, ter filhos fora do casamento, eutanásia e clonagem.

Os achados mencionados até aqui indicam que: o IRAP tem apresentado maior capacidade preditiva sobre comportamentos relacionados ao conteúdo do procedimento quando comparado com procedimentos baseados no relato verbal para uma variedade de temas; a religiosidade, especialmente a cristã, é uma variável pré-experimental altamente correlacionada com valores de pureza.

Os dados sobre a correlação entre religiosidade cristã e valores de pureza foram obtidos com a população dos EUA. Entretanto, a comunidade religiosa cristã no Brasil apresenta posturas semelhantes às observadas em Koleva et al (2012) sobre temas sociais polêmicos (Busin, 2011; Kalsing, 2002; Rosado-Nunes, 2015). No cenário político, religiosos cristãos encabeçam projetos políticos de cunho moral como a liberação da oferta de psicoterapia de reorientação sexual; a manutenção da proibição do aborto em qualquer circunstância; a manutenção da criminalização do uso da maconha; o embate sobre educação sexual nas escolas, apenas para citar alguns.

A partir disso, o presente estudo tem por objetivo comparar as respostas de participantes religiosos cristãos e não religiosos a medidas explícitas e implícitas de justiça e pureza. O conjunto dos achados a partir de amostras do EUA (Haidt & Graham, 2007; Graham et al, 2011; Koleva et al, 2012) sugere que não há diferenças significativas entre religiosos e não religiosos no que diz respeito a valores de justiça, mas que há diferenças significativas entre esses grupos no que toca valores de pureza, expressa pela concordância de participante de

religiosos diante desses valores. Este estudo buscou verificar se a mesma tendência também seria observada junto a participantes brasileiros e se haveria consistência entre procedimentos baseados em autorrelato, como o MFQ e um procedimento baseado na medida de responder relacional, o IRAP.

Método

Participantes

20 participantes completaram o estudo. Eles foram selecionados a partir de um questionário de elegibilidade (apêndice C) para a pesquisa que visou a determinar a inclusão no grupo religiosidade ou sem religiosidade ou a exclusão do estudo.

Questionário de elegibilidade. 49 pessoas responderam a este questionário. Estas pessoas foram convidadas via cartazes afixados em locais públicos e convites orais em salas de aula de disciplinas introdutórias de psicologia e em reuniões de grupos de cristãos universitários. Para ter acesso ao questionário, era preciso possuir no mínimo segundo grau completo e ser nativo do idioma português. Tais critérios foram aplicados a fim de garantir que os participantes eram competentes para ler e compreender informações em português. O questionário de elegibilidade foi inspirado no estudo de Cardoso, Leite e Nogueira (2002) que tece discussões e oferece alternativas para se questionar sobre crença religiosa em um país caracterizado pelo sincretismo religioso, como é o caso do Brasil. O questionário é composto de 23 itens em escala likert e mais 12 perguntas de respostas curtas ou múltipla escolha. Os itens da escala likert apresentavam elementos religiosos presentes em diferentes matrizes de religiosidade e o participante era solicitado a indicar seu grau de crença em cada item indo de 0 (não acredito nem um pouco) a 4 (acredito completamente). As 12 questões restantes visavam a identificar a rotina religiosa do participante ao longo da vida.

Para ser incluído no grupo religiosidade, o participante precisava responder, na escala likert, que acreditava completamente nos itens “Deus”, “Jesus Cristo”, “Anjos”, “Vida após a morte”, “oração” e “céu/inferno” e que não acreditava nem um pouco nos outros itens. Os itens “Maria” e “santos” também foram relacionados como elementos da crença cristã, mas não precisavam obrigatoriamente ser respondidos, porque a crença nesses elementos varia em função de diferentes matrizes cristãs. Nas questões sobre rotina religiosa, esses participantes precisavam indicar que frequentam estabelecimentos da mesma matriz religiosa desde a infância e que passam no mínimo duas horas semanais dedicados a atividades organizadas pelo estabelecimento religioso que frequentam.

14 participantes se encaixaram neste perfil e completaram o procedimento. Entre estes, 6 se identificaram como cristãos, sem especificar a vertente religiosa; 5 se identificaram como evangélicos; 2 se identificaram como católicos; e 1 se identificou como adventista do sétimo dia. Estas pessoas passam entre 2 e 10 horas semanais envolvidas em atividades ligadas a estabelecimentos religiosos como grupos de estudo bíblico, trabalho voluntário, clube de jovens ou trabalho missionário. 10 participantes são membros da CRU, a Cruzada de cristãos na ciência, grupo missionário que se reúne em universidades. A faixa de idade desses participantes foi de 18 a 21 anos, com a média de idade do grupo de 19,06 anos.

Para ser incluído no grupo sem religiosidade, o participante precisava responder que não acreditava nem um pouco nos elementos identificados com a crença cristã; e que não pratica qualquer religião de matriz cristã há pelo menos 10 anos. 9 participantes atingiram estes critérios, e foram convidados a participar do procedimento, mas apenas 6 deles atingiram os critérios de consistência na resposta ao IRAP e foram incluídos na análise. 3 desses participantes indicaram que frequentaram estabelecimentos religiosos cristãos na infância e nenhum deles frequentou esses estabelecimentos desde a adolescência e até o momento da pesquisa. A faixa de idade do grupo foi de 19 a 40 anos com uma média 22, 12 anos.

Todos os participantes confirmaram seu interesse em participar por meio da assinatura Termo de Consentimento Livre e Esclarecido (TCLE, apêndice B). Este procedimento foi previamente aprovado pelo comitê de ética em pesquisa com seres humanos no processo de número 82303318.5.0000.5561.

Aparato, material e procedimento

Aparato, material e procedimento foram idênticos aos do Estudo 1 descrito no capítulo anterior.

Análise dos Dados

Análises Intra-sujeito. As análises realizadas foram as mesmas descritas no Estudo 1 descrito no capítulo anterior.

Análises entre grupos.

Questionário. Foi calculada média de concordância com as cinco bases morais para cada grupo, tal como no Estudo 1. Adicionalmente, foi realizado um teste *t* de amostras independentes para os itens de cada base moral comparando os grupos com e sem religiosidade.

IRAP. Foram realizados testes *t* de amostras independentes comparando as latências grupais médias nos blocos consistentes e inconsistentes para as tentativas de cada uma das seis afirmações e entre os grupos. O teste permite verificar se houve diferença estatística entre as latências médias dos blocos consistentes e inconsistentes a fim de indicar se a diferença de médias tem relevância estatística. Além disso, foram realizados testes *t* de amostras independentes para as latências médias entre os grupos a fim de verificar se eles podem ser considerados iguais ou diferentes entre si para as tentativas de cada uma das afirmações.

Resultados

A Figura 6 apresenta os dados do MFQ referentes à média de concordância com cada uma das cinco bases morais para os grupos com e sem religiosidade.

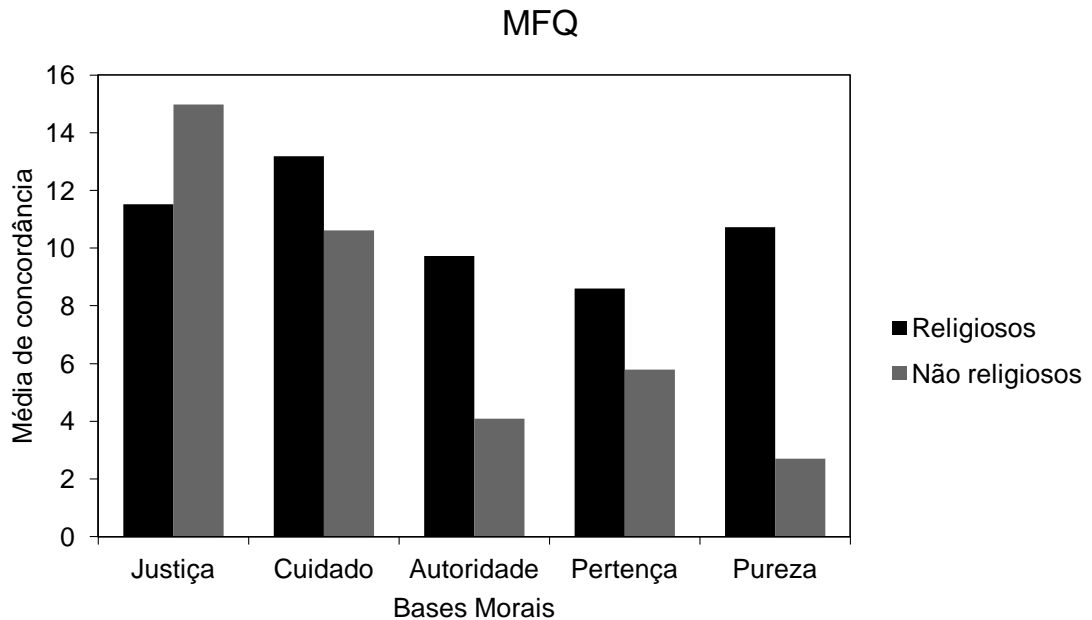


Figura 6. Média de concordância para cada uma das cinco bases morais por grupo.

A média de concordância grupal para cada valor foi verificada a partir do número escolhido por cada participante na escala likert de concordância para cada afirmativa, tal como descrito na sessão de análise de dados no Estudo 1. Uma vez que o valor poderia variar entre 0 e 15, médias a acima de 7,5 são consideradas indicativas de concordância com a afirmativa e abaixo deste valor, o indicativo é de discordância.

Como pode ser observado na Figura 6, ambos os grupos apresentaram média indicativa de concordância com os valores de justiça e cuidado, e diferiram nos valores de autoridade, pertença ao grupo e pureza com valor médio indicativo de concordância para os três itens apenas no grupo religiosidade. Testes *t* de amostras independentes apontaram os dois grupos como estatisticamente iguais no item justiça ($p=0,1$) e diferentes nos demais itens ($p=0,003$; $p=0$; $p=0,01$; e $p=0$, respectivamente).

A figura 7 apresenta as latências médias individuais para cada uma das afirmativas sobre justiça. Os participantes 1 a 14 compõem o grupo religioso. Participantes 15 a 20 compõem o grupo não religioso. Assim como no MFQ, houve uma tendência geral a escolher o termo relacional “verdadeiro” diante da combinação “Sou a favor” + afirmativa sobre justiça. Diante das afirmações sobre pureza houve padrões diferentes entre os grupos: o grupo de religiosos escolheu “verdadeiro” também diante das afirmativas sobre pureza; e o grupo não religioso escolhendo o termo oposto, “falso”.

A fim de facilitar a apresentação dos dados, as latências médias de todos os participantes estão apresentadas conforme a apresentação do Estudo 1, com respostas no bloco consistente apresentando a latência média da concordância com itens de justiça e discordância com itens de pureza e as respostas dos blocos inconsistentes apresentando as latências das respostas opostas. Dessa forma, latências médias mais curtas nos blocos consistentes indicam concordância com itens sobre justiça e discordância com valores de pureza. Latências médias mais curtas nos blocos inconsistentes indicam a quebra dessa tendência observada no Estudo 1.

Diante da primeira afirmativa sobre justiça, 8 de 14 participantes religiosos e 5 dos 6 participantes não religiosos expressaram concordância. As exceções foram os participantes 2, 3, 11, 12, 13, 14 e 15 que apresentaram latências médias mais curtas no bloco inconsistente ou indiferenciação das latências, o que é interpretado como ausência de responder relacional.

A latência média no grupo religioso foi de 3,4 segundos no bloco consistente e 3,66 segundos no bloco inconsistente. No grupo de não religiosos a latência média foi de 3 segundos para os blocos consistentes e 4,84 nos blocos inconsistentes. O teste *t* de amostras independentes indicou que para o grupo religioso, não houve diferença entre blocos consistentes e inconsistentes ($p=0,141$), sugerindo que não há concordância ou discordância com o item. Para o grupo não religioso, os blocos foram estatisticamente diferentes entre si

($p=0,001$) indicando concordância com a alternativa. Religiosos e não religiosos foram grupos estatisticamente diferentes para o item ($p=0,006$).

Na segunda afirmativa, 9 dos 14 participantes religiosos e 5 dos 6 participantes não religiosos apresentaram latência média mais curta no bloco consistente, sugerindo concordância com o item. Discordaram da afirmativa os participantes 2, 7, 8, 16, 12 e 15. A média grupal de latências no grupo religioso foi de 3,6 segundos nos blocos consistentes e 4,3 segundos nos blocos inconsistentes. Para o grupo não religioso, as médias foram de 3,8 segundos nos blocos consistentes e 5,1 para os blocos inconsistentes.

O teste t de amostras independentes confirmou as médias entre blocos consistentes e inconsistentes como estatisticamente diferentes para os dois grupos ($p=0,003$ para o grupo religioso e $p=0,005$ para o grupo não religioso) confirmando a concordância com a segunda afirmativa. Para esta afirmativa, religiosos e não religiosos foram considerados estatisticamente iguais ($p=0,01$).

Diante da terceira afirmativa, 3 de 14 participantes religiosos e 3 de 6 participantes não religiosos expressaram concordância com a afirmativa. Os participantes concordantes foram 10, 12, 13, 15, 17, e 20. A média grupal de latências para o grupo de religiosos foi de 4 segundos nos blocos consistentes e 3,5 segundos nos blocos inconsistentes. Para os participantes não religiosos, a média grupal foi de 4 segundos nos blocos consistentes e 4,2 segundos nos blocos inconsistentes. O teste t apontou que as médias para cada bloco do grupo religioso são diferentes ($p=0,02$) indicando discordância do grupo para com esta afirmativa. Já para o grupo não religioso, as médias foram consideradas iguais ($p=0,5$) sugerindo que não houve concordância ou discordância para com a afirmativa. Para a terceira afirmativa, os grupos foram apontados como estatisticamente iguais ($p=0,9$).

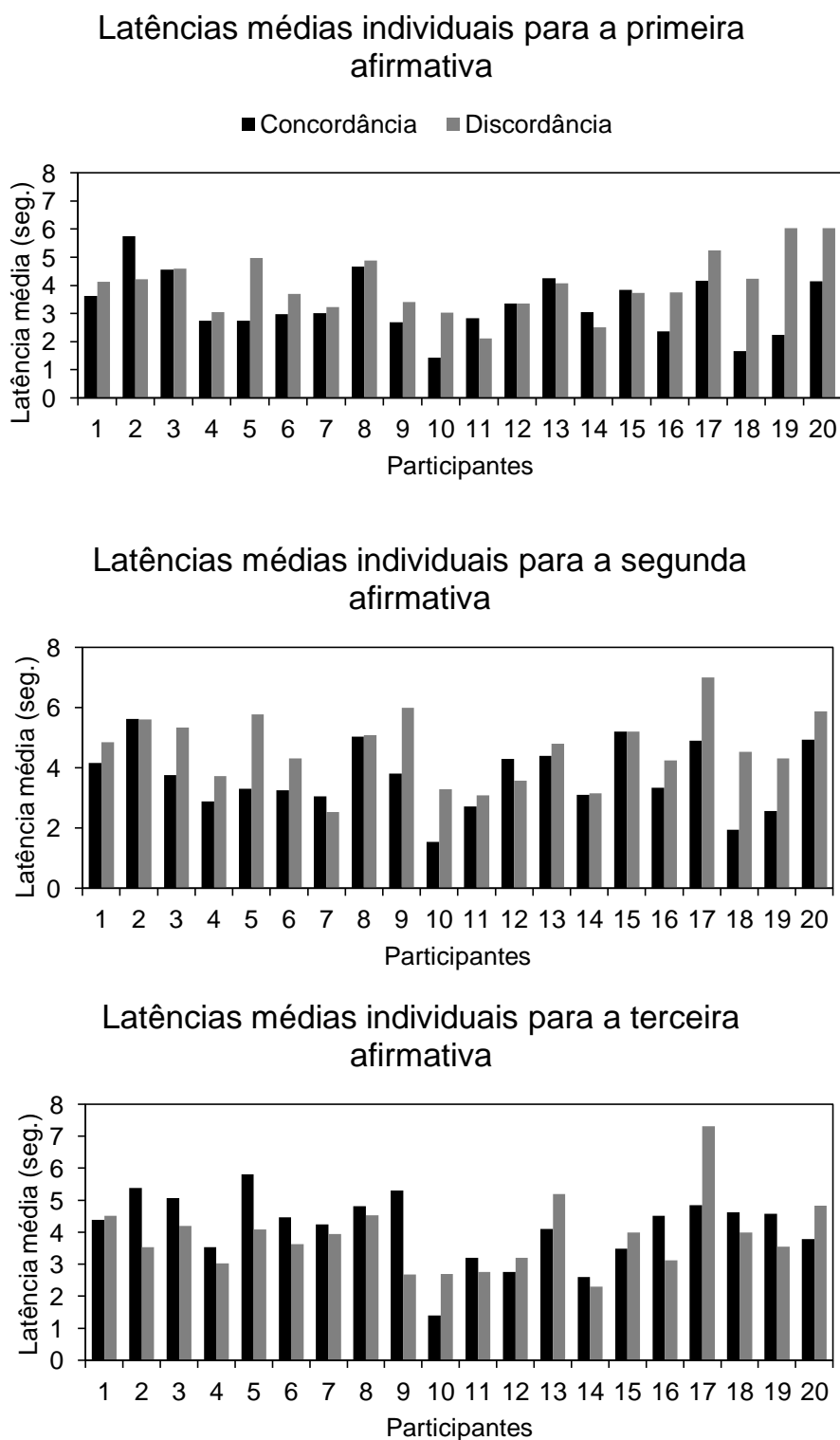


Figura 7. Latências médias individuais nos blocos consistentes e inconsistentes para cada uma das três afirmações sobre justiça.

A Figura 8 apresenta as latências médias individuais diante das afirmações sobre pureza. Diante da primeira afirmativa sobre pureza, 5 de 14 participantes religiosos e 5 de 6 participantes não religiosos apresentaram latência média mais curta no bloco consistente, indicativo de discordância com o item. Os participantes discordantes foram 1, 8, 9, 13, 14, 16, 17, 18, 19 e 20. A Latência média grupal para os participantes religiosos foi de 4,74 segundos nos blocos consistentes e 4,43 segundos nos blocos inconsistentes. O teste *t* apontou que as médias dos blocos não são diferentes entre si ($p=0,07$), sugerindo ausência de concordância ou discordância com o item.

Entre os participantes não religiosos as médias foram de 4,47 segundos nos blocos consistentes e 5,15 segundos nos blocos inconsistentes. Os blocos foram apontados como estatisticamente diferentes entre si, confirmando a discordância para com a afirmativa. Entretanto, a comparação entre grupos apontou que religiosos e não religiosos são estatisticamente semelhantes nesta alternativa ($p=0,5$).

Para a segunda afirmativa, 7 de 14 religiosos e 5 dos 6 participantes não religiosos apresentaram latências mais curtas no bloco consistente, expressando discordância com a afirmativa. Os participantes que expressaram concordância com a afirmativa foram 2, 4, 8, 10, 11, 12, 14 e 15. A latência média do grupo de religiosos foi de 4,33 segundos em blocos consistentes e 4,19 segundos em blocos inconsistentes.

Entre os participantes não religiosos as médias grupais foram de 4,13 segundos nos blocos consistentes e 4,82 nos blocos inconsistentes. Não houve diferença estatisticamente significativa entre os blocos do grupo religioso ($p=0,3$) nem do não religioso ($p=0,07$). Esses valores sugerem que não foi observada concordância ou discordância diante da afirmativa para nenhum dos grupos. A comparação entre grupos aponta religiosos e não religiosos como grupos diferentes nesta afirmativa ($p=0,12$).

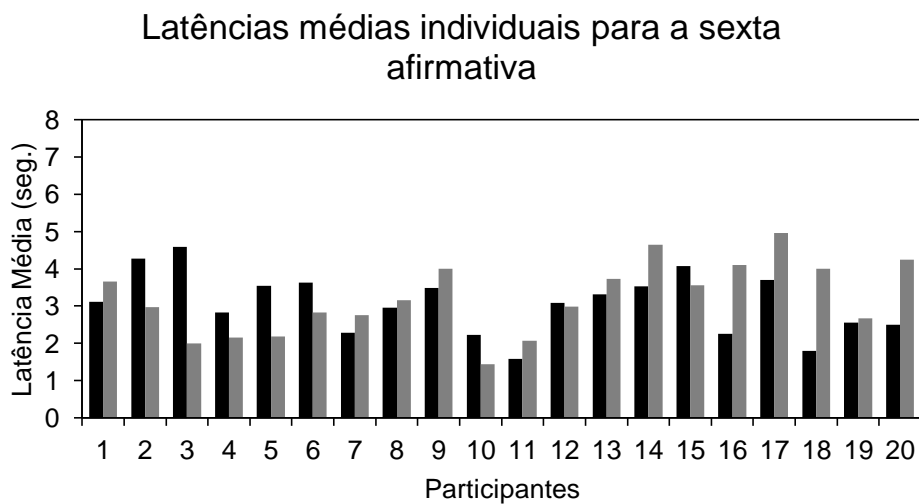
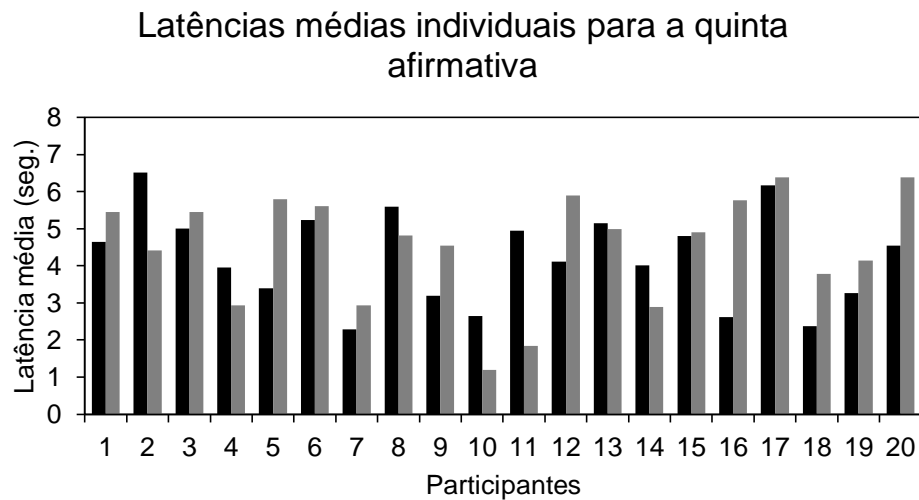
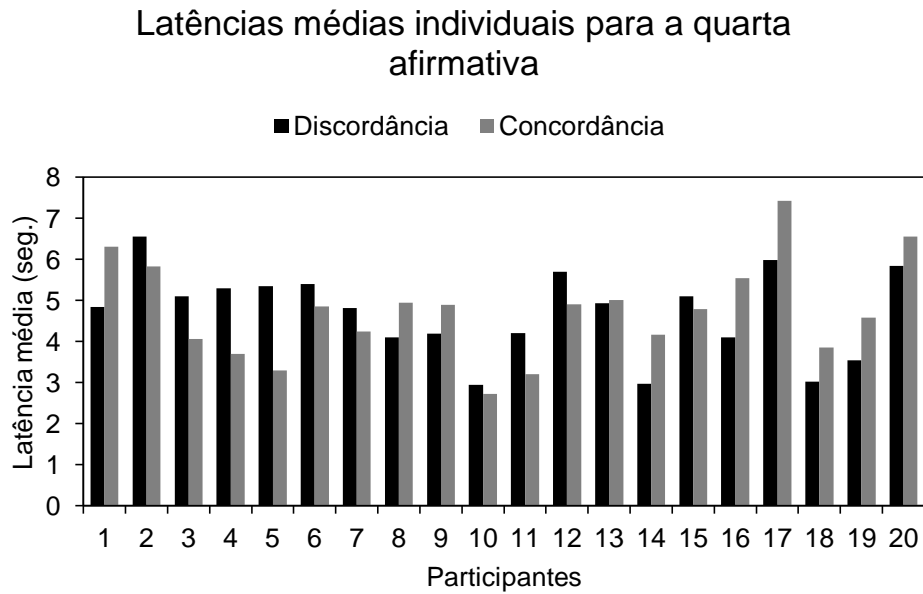


Figura 8. Latências médias individuais nos blocos consistentes e inconsistentes para cada uma das três afirmações sobre pureza.

Diante da terceira afirmativa, 7 de 14 participantes religiosos e 4 dos 6 não religiosos apresentaram latência média mais curta nos blocos consistentes. Os participantes que expressaram concordância ou ausência de preferência diante a afirmativa foram 2, 3, 4, 5, 6, 10, 12, 15 e 19. A latência média grupal para os participantes religiosos foi de 3,17 segundos nos blocos consistentes e 2,89 segundos nos blocos inconsistentes. Não houve diferença estatisticamente significativa entre os blocos, sugerindo ausência de concordância ou discordância para a afirmativa ($P=0,2$).

Para o grupo de não religiosos as médias foram de 2,9 segundos em blocos consistentes e 3,8 segundos em blocos inconsistentes. A diferença entre os blocos é estatisticamente significativa, confirmando a discordância diante da afirmativa ($P=0,004$). A comparação entre grupos aponta religiosos e não religiosos como grupos diferentes nesta afirmativa ($p=0$).

Em suma, participantes não religiosos apresentaram concordância em diante das afirmações 1 e 2 de justiça; ausência de preferência diante das afirmações 3 de justiça e 2 de pureza; discordância diante das afirmações 1 e 3 e pureza. Participantes religiosos apresentaram concordância diante da afirmação 2 de justiça; ausência de preferência diante das afirmações 1 de justiça, e 1, 2 e 3 de pureza; e discordância diante da afirmação 3 de justiça.

Testes *t* de amostras independentes indicam que os grupos são estatisticamente diferentes para a primeira afirmativa sobre justiça (concordância entre não religiosos e ausência de preferência entre os religiosos, $p=0,006$); a segunda sobre pureza (ausência de preferência em ambos os grupos, $p=0,01$); e a terceira afirmativa sobre pureza (discordância entre não religiosos e ausência de preferência entre religiosos, $p=0$), sendo estatisticamente indiferenciados para os demais itens.

Discussão

Os resultados deste procedimento indicam divergência entre as respostas fornecidas pelos participantes em função do tipo de procedimento. Os resultados do procedimento baseado em relato verbal indicam que ambos os grupos são semelhantes quanto à média da concordância com itens de justiça e diferentes na média de concordância para valores de pureza. Esse dado confirma os resultados de Koleva et al (2012) que apresentam uma correlação entre participantes de religião cristã e maior concordância com valores de pureza.

Os resultados confirmam também o indicado em de Haidt e Graham (2007) e Graham et al (2011) sobre a formação de um *cluster* entre valores de justiça e cuidado e outro entre valores de pureza, autoridade e pertença ao grupo. Participantes não religiosos exibiram um padrão de concordância apenas com valores de justiça e cuidado, indicadores de uma ética da autonomia. Os participantes religiosos, por sua vez, exibiram um padrão de concordância com as cinco bases morais, condizente com a combinação entre os três tipos de ética e correlacionada a uma visão política conservadora.

Os resultados do IRAP repetiram a tendência do Estudo 1 com níveis de concordância e discordância menos acentuados comparativamente com o procedimento de autorrelato. Para o grupo de não religiosos, enquanto o MFT sugeriu concordância quase absoluta diante dos itens de justiça (14,98 de um máximo de 15), indicando alta média de concordância para com as três afirmativas, o IRAP sugeriu que houve concordância nas tentativas referentes à primeira e segunda afirmativas e ausência de preferência diante da terceira. Para os valores de pureza, o MFT sugeriu claro padrão de discordância diante das três afirmativas. O IRAP, por sua vez, indicou ausência de preferência diante das tentativas referentes à primeira afirmativa e discordância diante das tentativas referentes à segunda e à terceira afirmativa sobre pureza. Dessa forma, O IRAP sugere padrão de concordância com valores de justiça, presente em duas de três afirmativas e de discordância com pureza, presente em duas de três afirmativas.

São os mesmos vieses observados no MFT, porém menos acentuados, pois não ocorreram nas três afirmativas.

Entre os participantes religiosos, O MFT sugeriu concordância diante de todas as afirmativas com valores próximos entre justiça e pureza (11,52 e 10,73 respectivamente). O IRAP, por sua vez indicou concordância com o valor de justiça presente em apenas uma de três afirmativas e ausência de concordância em todos os itens sobre pureza.

Dessa forma, a partir do IRAP, não é possível prever se o participante apresenta uma história pré-experimental de vivência religiosa cristã com base em um padrão de concordância com valores de pureza, tal como proposto em Koleva et al (2012). Ainda assim é possível diferenciar os dois grupos. Essa diferenciação se dá pelo padrão de respostas de concordância com justiça e discordância com pureza mais pronunciados no grupo de não religiosos.

O conjunto dos resultados indica que ambos os procedimentos permitem diferenciar os participantes em função do *background* religioso ou ausência deste. Portanto, a variável religiosidade foi preditora padrões diferenciais de concordância ou discordância com valores de justiça e pureza. Todavia, a forma como cada procedimento caracteriza esses vieses foi bastante diferente.

Os distintos padrões de resposta observados entre religiosos e não religiosos sugerem, em conformidade com os achados de Koleva et al, (2012), que a presença ou ausência de vivência religiosa cristã pode ser um importante preditor dos valores morais. Adicionalmente, estes resultados seriam indicativos de que essas pessoas seriam expostas a contingências de seleção cultural distintas.

Não foram calculados os escores D-IRAP neste estudo. Os resultados do Estudo 1 indicaram que as afirmativas para cada um dos valores são diferentes entre si. Dessa forma, não há valor informativo no agrupamento das latências médias grupais de um mesmo valor moral.

Discussão geral

Qual a origem da moralidade humana? Como decidimos se uma ação deve ser obrigatória, proibida ou permitida? Como essas decisões afetam a nossa conduta? Essas são questões de grande interesse na filosofia há milênios (Abbagnano, 2007; Brugger, 1977; Hocutt, 2013; Japiassu & Marcondes, 1996, Skinner, 1971).

Mais recentemente elas têm atraído o interesse de pesquisadores dedicados à psicologia experimental. A ciência tem sido historicamente considerada incapaz de se pronunciar sobre moralidade ou valores, pois se entendia que esses temas não poderiam ser abordados nos moldes de uma ciência natural (Moore, 1903/1922; Pimentel-Pequeno, 2003). Essa acusação é forte sobretudo diante da possibilidade de uma ética prescritiva pautada pela ciência. Ela tem sido rebatida por aqueles que defendem que qualquer questão humana pode ser abordada empiricamente e não necessariamente num caráter prescrito, mas em vez disso, em caráter descritivo (para uma revisão desse debate, sugerimos Hocutt, 2013). Uma vez que a moral é uma dimensão importante do comportamento humano, a moralidade é uma questão para exame da análise do comportamento (Skinner, 1971).

Este trabalho apresentou uma abordagem metaética da moralidade. A incursão neste vasto campo do conhecimento foi orientada pela abordagem da análise do comportamento, na medida em que se apoiou em pressupostos caros a esta, como a noção de multideterminação do comportamento orientada pelo princípio de seleção por consequências e o uso de procedimentos experimentais (Chiesa, 1994; Glenn & Malagodi, 1991; Killeen, 2001; Tinbergen, 1963).

Para atender ao pressuposto da multideterminação do comportamento, incluímos pesquisas sobre moralidade de outras disciplinas que também são consistentes com o modelo de seleção por consequências e utilizam o método experimental como ferramenta principal na produção de evidências. Entendemos que a busca por evidências proximais e distais e que

compreendem as quatro causas do comportamento aumenta a nossa compreensão do fenômeno em análise na medida em que abarcam diferentes níveis explicativos desse fenômeno.

O Capítulo 1 contribui para a proposta de intersecção entre diferentes epistemologias no estudo experimental da moral ao apresentar um panorama dos principais temas de interesse a achados em três grandes áreas de pesquisa. A investigação das bases evolutivas da moralidade nos mostra que o comportamento moral tem origem nas pressões seletivas que levam os seres humanos e outros primatas a constituir coletividades e estabelecer regras para compartilhar recursos e benefícios (Curry, 2004; de Waal, 2006; Tomasello, 2016). Algumas bases morais, regras fundacionais da convivência, se destacam por sua presença em diversas culturas ao longo da história (Graham et al, 2012; Haidt & Graham, 2007). Entre as heranças filogenéticas que subjazem o comportamento moral estariam os distintos meios de processamento da informação que vão lidar com as informações ambientais imediatas (Cushman, Kumar & Railton, 2017).

Essas pesquisas oferecem descrições da moralidade que unem os níveis de explicações distais e proximais. Elas se debruçam sobre elementos que teriam sido selecionados ao longo da história evolutiva dos seres humanos e como esses elementos se expressam diante de variáveis proximais que compõem a diversidade cultural e a história de reforçamento dos organismos.

Também no nível proximal de análise, sustentamos que a diversidade de convenções culturais, dos temas que compõem normas de obrigações, permissões e proibições num grupo, pode ser adequadamente descrita por meio de processos de controle de estímulos e responder relacional. Aqui, a busca não é pela possível origem dessas convenções, mas por sua descrição em termos de processos básicos. Foi sobre este aspecto da moralidade que nos debruçamos empiricamente. A compreensão de sistemas de valores como produtos de

responder relacional permite a descrição de uma série de fenômenos culturais de difícil acesso via comportamento verbal. Essa perspectiva demanda procedimentos que produzam dados sobre linguagem que prescindam do autorrelato.

Entre os procedimentos disponíveis para a investigação de relações entre estímulos, optamos pelo IRAP pelo seu extensivo uso com temas sociais controversos e pouca suscetibilidade a fatores intervenientes como a “desejabilidade social”. Os dois experimentos realizados demonstraram que medidas de responder relacional e autorrelato apresentaram resultados divergentes para valores de justiça e pureza.

O Capítulo 3 introduziu a história pré-experimental do participante como uma variável adicional para verificar a confiabilidade dos dois procedimentos em corretamente distinguir participantes religiosos de não religiosos diante de suas respostas para itens sobre justiça e pureza. Ambos os procedimentos diferenciaram corretamente participantes religiosos de não religiosos apontando diferenças nos vieses de concordância com valores de pureza e justiça como fatores preditores de diferenças entre esses grupos. Todavia, a caracterização dos dois grupos é distinta em função do procedimento.

Os resultados do MFQ corroboraram as predições da literatura da MFT. O grupo de participantes religiosos apresenta concordância com os cinco valores morais, o que, na literatura é correlacionado ao padrão de respostas de pessoas religiosas e conservadoras. O grupo de participantes não religiosos apresentou concordância apenas diante de valores de justiça e cuidado, apresentando assim um padrão de respostas alinhado a uma moral da autonomia, correlacionada a uma visão dita progressista de como deve ser pautada a convivência. Dessa forma, a diferença entre os dois grupos se caracteriza pela concordância com valores de pureza, pertença ao grupo e respeito a autoridade expressa pelo grupo de religiosos e ausente no outro grupo.

Os resultados do IRAP não corroboraram essas predições. O grupo de religiosos se caracterizou por leve concordância com valores de justiça e ausência de concordância com valores de pureza. O grupo de não religiosos se caracteriza por concordância com valores de justiça e discordância com valores de pureza. Dessa forma, a diferença entre os dois grupos é caracterizada pela discordância diante de valores de pureza exibida pelo grupo de não religiosos.

Apesar de ambos os procedimentos terem se mostrado eficazes em corretamente diferenciar os dois grupos no Estudo 2, a diferença na caracterização necessita de investigação adicional a fim de se verificar se alguma das caracterizações tem maior poder explicativo da variável religiosidade. O uso do IRAP para a investigação de conteúdos morais ainda pode ser melhor explorado a partir de mudanças em parâmetros do procedimento como o controle da quantidade de palavras nos estímulos alvo; ou ainda pela introdução de um limite de tempo para a emissão da resposta. O conteúdo moral também pode ser investigado de forma mais abrangente por meio da utilização de outras teorias morais como base para a composição dos estímulos.

O caráter preditivo do IRAP pode ser adicionalmente explorado por meio da introdução de tarefas experimentais não verbais. A moralidade é apontada como o substrato para uma série de repertórios como as várias categorias de atos altruístas (Tomasello, 2016; Curry, 2004), e trocas sociais (Cosmides, Barrett & Tooby, 2010), entre outros. O estudo experimental desses repertórios é extensivo e abrange diversos procedimentos como dilema do prisioneiro; aversão à iniquidade; detecção de trapaceiros por meio do dilema de Wason; jogo dos bens públicos e tragédia dos comuns, apenas para citar alguns. O desempenho nesses procedimentos poderia ser utilizado como elemento de comparação com o desempenho no IRAP sobre valores morais.

Os experimentos apresentados aqui se alinham à produção em metaética que tem a linguagem como dado privilegiado de análise dos valores. As pessoas são extensamente treinadas a parabenizar ou censurar comportamentos os quais muitas vezes não as afetam diretamente. Este fornecimento generalizado de consequências sociais para o comportamento é um poderoso meio de transmissão de práticas culturais (Skinner, 1953). Dessa forma, o estudo dos valores morais passa pela classificação de condutas como certas ou erradas e a aplicação dessa classificação em termos de consequências sociais para o comportamento.

Uma análise comportamental do repertório de classificar condutas passa pela descrição de contingências nas quais o comportamento de classificar produz reforçamento ou punição socialmente mediados. Esta classificação está também intimamente ligada a quais comportamentos não verbais produzirão consequências reforçadoras ou punidoras. O comportamento a ser classificado é o objeto sobre o qual se descreve o efeito de reforço, o seu valor. Ou, como afirmou Skinner (1971) “the social contingencies, or the behaviors they generate, are the ‘ideas’ of a culture; the reinforcers that appear in the contingencies are its values.” (p.126).

Dessa forma, é possível definir o comportamento moral como o repertório de classificar comportamentos em certo e errado, adequar o próprio comportamento a esta classificação e utilizá-la como parâmetro para o fornecimento de reforçadores e punidores a outrem como forma de estabelecer regulações sobre a convivência em termos de obrigações, proibições e permissões a fim de regular o constante conflito entre preferências individuais e bem-estar coletivo. Uma consequência dessa definição é que o estudo dos valores pode prescindir do relato verbal sobre classificações de certo e errado e se basear na descrição das práticas culturais que caracterizam a convivência em um dado grupo.

A abordagem à moralidade apresentada aqui não se confunde com tentativas de planejamento cultural baseadas em processos comportamentais. Nas palavras de Skinner “if a

scientific analysis can tell us how to change behavior, can it tell us what changes to make?” (p. 102). A resposta daqueles que veem tentativas de planejar a cultura a partir da ciência como uma tentativa de imperialismo científico (Staddon 2009, 2013) é um categórico *não*.

Algumas prescrições morais costumam ser invocadas para justificar direções para a cultura como o consequencialismo, com a máxima do maior bem possível para o maior número de pessoas, ou o aumento total da felicidade. Porém, não existe um critério primordial que selecione alguma dessas direções como a melhor ou que prescreva de forma clara e pacífica, quais contingências devem ser estabelecidas para o alcance desses objetivos (Skinner, 1953). Para Staddon (2009) o determinante básico em qualquer tentativa de planejamento cultural estará nas crenças no planejador.

Todavia, para Skinner (1971), as coletividades humanas não estão à deriva no que tange a fundamentos morais que guiam a convivência. Ele propõe um valor primordial que guia a conduta humana, sem que tenha sido intencionalmente escolhido para tal: a sobrevivência. Uma cultura seria selecionada na medida em que seus membros sobrevivem e a transmitem seu conjunto de práticas à próxima geração (Skinner, 1971). Se a moralidade surge quando os primeiros homens se organizam em grupos, a manutenção da convivência humana em coletividades seria um atestado de que a moralidade foi selecionada? Os três trabalhos que compõem esta tese caminham na direção dos estudos que podem ajudar a discutir essas questões no contexto da análise do comportamento.

Referências

- Abbagnano, N. (ed.). (2007). Ética. In *Dicionário de filosofia* (A. Bossi & I. Castilho Benedetti, trad.) (pp. 442-451), 5 ed. São Paulo: Martins Fontes.
- Abib, J. A. D. (2002). Ética de Skinner e metaética. In (H. J. Guilhardi, eds.) *Sobre Comportamento e Cognição: contribuições para a construção da teoria do comportamento vol.10* (pp.125-137), 1 ed. Santo André (SP): ESETec Editores Associados.
- Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the implicit relational assessment procedure (IRAP) as a direct measure of implicit beliefs, *The Irish Psychologist*, (32)7, 169-177. Retirado de pdfs.semanticscholar.org/5a62/bd895b2ebf0eeeb11b70a31a3b386ef7061b.pdf?_ga=2.74154394.986342552.1579558814-1477854841.1579558814
- Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y. & Stewart, I. (2010). The implicit relational assessment procedure: exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white irish individuals. *The Psychological Record*, 60, 43-66. doi:10.1007/BF03395694
- Barnes, D., & Holmes, Y. (1991). Radical Behaviorism, Stimulus Equivalence, and Human Cognition. *The Psychological Record*, 41(1), 19–31. doi:10.1007/bf03395091
- Baum, W. M. (1995). Rules, culture, and fitness. *The Behavior Analyst*, 18(1), 1–21. doi: 10.1007/BF03392688

- Bortoloti, R., & de Rose, J. (2012). Equivalent stimuli are more strongly related after training with delayed matching than after simultaneous matching: a study using implicit relational assessment procedure (IRAP). *The Psychological Record*, 62(1), 41-54. doi: 10.1007/BF03395785
- Brugger, W. (ed.). (1977). Ética. In *Dicionário de filosofia* (A. Pinto de Carvalho, trad.) (pp.169-172), 3 ed. São Paulo: EPU.
- Busin, V. M. (2011). Religião, sexualidades e gênero. *Revista de estudos da religião*, 11(1), 105-124. Retirado de: <https://revistas.pucsp.br/rever/article/view/6032/4378>
- Cardoso, A. A., Leite, C. A. C., Nogueira, R. F. A. (2002). Pesquisa quantitativa no campo religioso: reflexões ulteriores sobre a experiência de participação de um grupo acadêmico de estudos da religião em duas pesquisas quantitativas. *Revista de estudos da religião*, 3(2), 86-99. Retirada de: www.pucsp.br/rever/rv3_2002/p_carlei.pdf
- Carpenter, K. M., Martinez, D., Vadhan, N. P., Barnes-Holmes, D., & Nunes, E. V. (2012). Measures of attentional bias and relational responding are associated with behavioral treatment outcome for cocaine dependence. *The American Journal of Drug and Alcohol Abuse*, 38(2), 146-154. doi: 10.3109/00952990.2011.643986
- Chiesa, M. (1994). *Radical Behaviorism: the philosophy and the science*. Boston (MA): Authors Cooperative, Inc., Publishers.
- Chiesa, M. (2003). Sobre la meta-ética, la ética normativa y el conductismo. *Revista latinoamericana de psicología*, 35(3), 289-297. Retirado de www.redalyc.org/pdf/805/80535305.pdf

- Cihon, T. C., & Mattaini, M. A. (2019). Emerging cultural and behavioral systems science [Editorial]. *Perspectives on behavior science*, 42, 699-711. doi: 10.1007/s40614-019-00237-8
- Cosmides, L., Barrett, C., & Tooby, J. (2010). Adaptive specializations, social exchange, and the evolution of human intelligence. *PNAS*, 107(2), 9007-9014. doi: 10.1073/pnas.0914623107
- Curry, O. S. (2004). *Morality as natural history: an adaptationist account of ethics* (doctoral dissertation). Retirado de core.ac.uk/download/pdf/40091549.pdf
- Cushman, F., Kumar, V., & Railton, P. (2017). Moral learning: current and future directions. *Cognition*. doi:10.1016/j.cognition.2017.06.008
- Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J., & Gore, N. J. (2009). Assessing the implicit beliefs of sexual offenders using the implicit relational assessment procedure. *Sexual abuse: a journal of research and treatment*, 21 (1), 57-75. doi: 10.1177/1079063208326928.
- de Carvalho, M. P., & de Rose, J. (2014). Understanding racial attitudes through the stimulus equivalence paradigm. *The Psychological Record*, 64(3), 527-536. doi: 10.1007/s40732-014-0049-4.
- de Rose, J. C. (1993). Classes de estímulos: implicações para uma análise comportamental da cognição. *Psicologia: teoria e pesquisa*, 9, 283-303. Retirado de: <https://pdfs.semanticscholar.org/df5e/582c9ce9722fa22f8bb2bbb9208c0df77e8d.pdf>
- de Waal, F. (2006). *Primates and philosophers: how morality evolved*. Princeton (NJ): Princeton University Press.

- Dittrich, A. (2010). Ética e comportamento. In E. Zagury Tourinho e S. Vasconcelos Luna (eds.) *Análise do Comportamento: investigações históricas, conceituais e aplicadas*, (pp. 37-59), São Paulo: Roca.
- Dixon, M. R., Rehfeldt, R. A., Zlomke, K. R., & Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *The Psychological Record*, 56(1), 83-103. doi:10.1007/BF03395539
- Drake, C. E., Kellum, K. K., Wilson, K. G., Luoma, J. B., Weinstein, J. H., & Adams, C. H. (2010). Examining the implicit relational assessment procedure: four preliminary studies. *The Psychological Record*, 60(1), 60-81. doi:10.1007/BF03395695
- Ferrater-Mora, J. (1982). Moral. In *Dicionário de filosofia abreviado* (A. José Massano & M. J. Palmeirim, trad.) (pp.270-271) 5 ed. Lisboa: Publicações Dom Quixote.
- Ferrater-Mora, J. (2001). Ética. In *Dicionário de filosofia – Tomo II (E-J)* (M. Stela Gonçalves et al., Trad.) (pp.931-935), São Paulo: Edições Loyola.
- Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the implicit relational assessment procedure: the impact of three types of introductory rules. *The psychological record*. 66(2), 309-321. doi: 10.1007/s40732-016-0173-4
- Flores, T. (2016). Genocídio da juventude negra no Brasil: as novas formas de guerra, raça e colonialidade do poder. In A. Macedo et al (eds) *Direitos Humanos: diversas abordagens*, (pp. 109-126), Rio de Janeiro: Câmara brasileira de jovens escritores.
- Glenn, S. S. (1989). Verbal behavior and cultural practices. *Behavior Analysis & Social Action*, 7(1-2), 10–15. doi:10.1007/BF03406102

- Glenn, S. S. & Malagodi, E. F. (1991). Process and content in behavioral and cultural phenomena. *Behavior and social Issues*, *1*(2), 1-14. doi: 10.5210/bsi.v1i2.163
- Glenn, S. S., Malott, M. E., Andery, M. A. P. A., Benvenuti, M., Houmanfar, R. A., Sandaker, I.,... & Vasconcelos, L. A. (2016). Toward consistent terminology in a behaviorist approach to cultural analysis. *Behavior and social Issues*, *25*, 11-27. doi: 10.5210/bsi.v.25i0.6634
- Glenn, S.S. (1988). Contingencies and Metacontingencies: Toward a Synthesis of Behavior Analysis and Cultural Materialism. *The Behavior Analyst*, *11*, 161–179. doi:10.1007/BF03392470
- Gomes, N. L., & Laborne, A. A. de P. (2018). Pedagogia da crueldade: racismo e extermínio da juventude negra. *Educação em Revista*, *34*, e197406. doi: 10.1590/0102-4698197406.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2012). Moral foundations theory: the pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, Forthcoming, *47*, 55-130. Retirado de <https://ssrn.com/abstract=2184440>
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., & Ditto, P. H. (2011). Mapping the moral domain. *Perspectives in Social Psychology*, *101*(2), 366–385. doi: 10.1037/a0021847
- Haidt, J., & Graham, J. (2007). When Morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Social justice research*, *20* (1), 98-116. doi: 10.1007/s11211-007-0034-z

- Haidt, J. and Hersh, M. A. (2001), Sexual Morality: The Cultures and Emotions of Conservatives and Liberals. *Journal of Applied Social Psychology*, 31, 191–221. doi:10.1111/j.1559-1816.2001.tb02489.x
- Haidt, J., Graham, J., & Joseph, C. (2009). Above and below left-right: ideological narratives and moral foundations. *Psychological inquiry*, 20(2-3), 110-119. doi: 10.1080/10478400903028573
- Haidt, J., & Joseph, C. (2008). The moral mind: how five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *Evolution and cognition. The innate mind Vol. 3. Foundations and the future* (pp. 367-391). New York, NY, US: Oxford University Press. doi:10.1093/acprof:oso/9780195332834.003.0019
- Hocutt, M. (2013). A behavioral analysis of morality and value. *The Behavior Analyst*, 36(2), 239-249. doi: 10.1007/bf03392310
- Hooper, N., Villatte, M., Neofotistou, E., & Mc Hugh, L. (2010). The effects of mindfulness versus thought suppression on implicit and explicit measures of experiential avoidance. *International Journal of Behavioral Consultation and Therapy*, 6(3), 233-244. doi: 10.1037/h0100910
- Houmanfar, R., & Rodrigues, N. J. (2006). The metacontingency and the behavioral contingency: points of contact and departure. *Behavior and Social Issues*, 15(1), 13-30. doi: 10.5210/bsi.v15i1.342
- Hussey, I., Mhaoileoin, D. N., Barnes-Holmes, D., Ohtsuki, T., Kishita, N., Hughes, S., & Murphy, C. (2016). The IRAP is nonrelative but not acontextual: changes to the contrast

category influence men's dehumanization of women. *The Psychological Record*, 66(2), 291–299. doi:10.1007/s40732-016-0171-6

Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: a guide to interpreting IRAP data. *Journal of Contextual Behavioral Science*, 4(3), 157-162. doi:10.1016/j.jcbs.2015.05.001.

Japiassu, H. & Marcondes, D. (1996). *Dicionário básico de filosofia*. 3 ed. Rio de Janeiro: Zahar.

Kalsing, V. S. S. (2002). O debate do aborto: a votação do aborto legal no Rio Grande do Sul. *Cadernos pagu*, 19, 279-314. doi:10.1590/S0104-83332002000200011.

Killeen, P. R. (2001). The four causes of behavior. *Current Directions in Psychological Science*, 10(4), 136–140. doi: 10.1111/1467-8721.00134

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: how five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality*, 46(2), 184-194. doi:10.1016/j.jrp.2012.01.006

Lalande, A. (1999). Moral. In *Vocabulário técnico e crítico de filosofia* (F. Sá Correia, M. E. V. Aguiar, J. E. Torres e M. G. de Souza, trad.) (pp.703-705), 3 ed. São Paulo: Martins Fontes.

Leech, A., Barnes-Holmes, D. & Madden, L. (2016). The implicit relational assessment procedure (IRAP) as a measure of spider fear, avoidance, and approach. *The psychological record*, 66(3), 337-349. doi: 10.1007/s40732-016-0176-1

- McKenna, I. M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology & Psychological Therapy*, 7(2), 253–268. Retirado de: www.redalyc.org/pdf/560/56070209.pdf
- Mizael, T. M., de Almeida, J. H., Silveira, C. C., & de Rose, J. C. (2016). Changing racial bias by transfer of functions in equivalence classes. *The Psychological Record*, 66(3), 451-462. doi: 10.1007/s40732-016-0185-0
- Moore, G. E. (1903/1922). *Principia Ethica*. 2 ed. Cambridge at the University Press.
- Nicholson, E., & Barnes-Holmes, D. (2012). Developing an implicit measure of disgust propensity and disgust sensitivity: examining the role of implicit disgust propensity and sensitivity in obsessive-compulsive tendencies. *Journal of Behavior Therapy and Experimental Psychiatry*. 43(3), 922-930. doi: 10.1016/j.jbtep.2012.02.001
- O'Reilly, A., Roche, B., Ruiz, M. R., Tyndall, I. T., & Gavin, A. (2012). The Function Acquisition Speed Test (FAST): a Behavior Analytic Implicit Test for Assessing Stimulus Relations. *The Psychological Record*, 62(3). pp. 507-528. doi:10.1007/BF03395817
- O'Reilly, A., Roche, B., Gavin, A., Ruiz, M. R., Ryan, A., & Campion, G. (2013). A Function acquisition speed test for equivalence relations (Faster). *The Psychological Record*, 63(4), 707–724 doi:10.11133/j.tpr.2013.63.4.001
- Paunonen, S. V., & LeBel, E. P. (2012). Socially desirable responding and its effects on the validity of personality assessments. *Journal of personality and social psychology*, 103 (1), 158-175. doi: 10.1037/a0028165

- Perez, W. F., Kovac, R., Linares, I. M. P., Barbosa, S. F. U., Gomes, C. T., dos Santos, G. A. R., & de Almeida, J. H. (2017). Utilizando o AAQ-II e o IRAP para prever a ocorrência de esquiva experiencial. *Revista Brasileira de Terapia Comportamental e Cognitiva*, 19(3), 34-48. doi: 10.31505/rbtcc.v19i3.1052
- Pimentel-Pequeno, M. (2003). Moore e os pressupostos da metaética. *Étic@*, 2(1), 21-34. doi: 10.5007/%25x
- Power, P., Barnes-Holmes, D., Barnes-Holmes, Y. & Stewart, I. (2009). The implicit relational assessment procedure (IRAP) as a measure of implicit relative preferences: a first study. *The Psychological Record*, 59, 621-640. doi:10.1007/BF03395684
- Rabelo, L., 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. doi: 10.1007/s40732-014-0006-2
- Rönspies, J., Schmidt, A. F., Melnikova, A., Krumova, R., Zolfagari, A., & Banse, R. (2015). Indirect measurement of sexual orientation: comparison of the implicit relational assessment procedure, viewing time, and choice reaction time tasks. *Archives of Sexual Behavior*, 44(5), 1483-1492. doi: 10.1007/s10508-014-0473-1
- Rosado-Nunes, M. J. F. (2015). A “ideologia de gênero” na discussão do PNE. A intervenção da hierarquia católica. *Horizonte*, 13(39), 1237-1260. doi:10.5752/P.2175-5841.2015v13n39p1237

- Sampaio, A. A. S., & Andery, M. A. P. A. (2010). Comportamento social, produção agregada e prática cultural: uma Análise Comportamental de fenômenos sociais. *Psicologia: Teoria e Pesquisa*, 26(1), 183-192. doi: 10.1590/S0102-37722010000100020
- Santos, J. G. de F. (2019). *A discriminação racial e de gênero no mercado de trabalho de Alagoas*. Trabalho de Conclusão de Curso, Universidade Federal de Alagoas, Santana do Ipanema, AL, Brasil.
- Shweder, R. A. (1990). In defense of moral realism: reply to Gabennesch. *Child Development*, 61(6), 2060–2067. doi:10.2307/1130859
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston (MA): Authors Cooperative.
- 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. doi: 10.1901/jeab.1982.37-5
- Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, 52(5), 270–277. doi:10.1037/h0062535
- Skinner, B. F. (1953). *Science and human behavior*. Oxford: McMillan.
- Skinner, B. F. (1957). *Century psychology series. Verbal behavior*. Appleton-Century-Crofts.
Doi:10.1037/11256-000
- Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Peguin Books.
- Skinner, B. F. (1981). Selection by consequences. *Science*, 213(4507), 501-504. doi: 10.1126/science.7244649

- Skinner, B. F. (1991). Genes e comportamento. In *Questões recentes na Análise do Comportamento* (pp.71-82) 5 ed, Campinas (SP): Papyrus editora.
- Staddon, J. E. R. (2009). Scientific imperialism and behaviorist epistemology. *Behavior and Philosophy*, 32, 231-242. Retirado de pdfs.semanticscholar.org/b204/6a178be74c46db33e8796efa224bf6afa386.pdf
- Staddon, J. E. R. (2013). Faith, fact, and behaviorism. *The Behavior Analyst*, 36, 229–238 (2013) doi:10.1007/BF03392309
- Terrace, H. S. (1966). Stimulus control. In W. K. Honig (Ed.) *Operant Behavior: Areas of research and application*. New York: Appleton-Century-Crofts.
- Tinbergen, N. (1963). On aims and methods of ethology. *Ethology*, 20(4), 410-433. doi:10.1111/j.1439-0310.1963.tb01161.x
- Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge, MA: Harvard University Press.
- Vahey, N. A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure as a measure of self-esteem: irish prisoner groups and university students. *The Psychological Record*, 59(3), 371–388. doi:10.1007/BF03395670
- Watt, A., Keena, M., Barnes, D., & Cairns, E. (1991). Social Categorization and stimulus equivalence. *The Psychological Record*, 41(1), 33-50. doi:10.1007/BF03395092

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and psychological measurement*, 69(4), 548-565. doi: 10.1177/0013164408324469

APÊNDICE A - Questionário de Bases Morais

Dados demográficos

Gênero:

Feminino Masculino Outro: _____

Idade: _____

Cidade de nascimento: _____

Cidade onde reside atualmente: _____

Grau de escolaridade: _____

Se estiver cursando ou já concluiu o ensino superior ou técnico, informe sua área de formação: _____

Profissão: _____

Pratica alguma religião?

Sim. Qual? _____ Não

Por favor, marque o número que melhor expressa o grau de importância da religião na sua vida:

- 0 Nada importante
 1 Pouco importante
 2 importante
 3 Completamente importante

Renda familiar mensal:

- 1 a 3 salários mínimos (R\$ 937,00 a R\$ 2.811,00)
 4 a 6 salários mínimos (R\$ 3.748,00 a R\$ 5.622,00)
 7 a 9 salários mínimos (R\$ 6.559,00 a R\$ 8.433,00)
 10 a 12 salários mínimos (R\$ 9.370,00 a R\$ 11.244,00)
 Acima de 12 salários mínimos (acima de R\$ 11.244,00)

Por favor, marque o número que melhor expressa o seu grau de concordância com cada uma das afirmações abaixo:

Compaixão por aqueles que sofrem é a virtude mais crucial.
--

- | |
|---|
| 0 <input type="checkbox"/> discordo fortemente |
| 1 <input type="checkbox"/> discordo moderadamente |
| 2 <input type="checkbox"/> discordo um pouco |

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

Quando o governo faz leis, o princípio número um deve ser garantir que todos são tratados de forma igual.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

Eu tenho orgulho da história do meu país.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

O respeito à autoridade é algo que todas as crianças precisam aprender.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

As pessoas não devem fazer coisas repugnantes, mesmo que ninguém seja prejudicado.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

É melhor fazer o bem do que fazer o mal.

- 0 () discordo fortemente
- 1 () discordo moderadamente
- 2 () discordo um pouco
- 3 () concordo um pouco
- 4 () concordo moderadamente
- 5 () concordo fortemente

Uma das piores coisas que alguém poderia fazer é machucar um animal indefeso.

- 0 () discordo fortemente
- 1 () discordo moderadamente
- 2 () discordo um pouco
- 3 () concordo um pouco
- 4 () concordo moderadamente
- 5 () concordo fortemente

Justiça é o pré-requisito mais importante para uma sociedade.

- 0 () discordo fortemente
- 1 () discordo moderadamente
- 2 () discordo um pouco
- 3 () concordo um pouco
- 4 () concordo moderadamente
- 5 () concordo fortemente

As pessoas devem ser leais aos seus familiares, mesmo quando eles fizerem algo errado.

- 0 () discordo fortemente
- 1 () discordo moderadamente
- 2 () discordo um pouco
- 3 () concordo um pouco
- 4 () concordo moderadamente
- 5 () concordo fortemente

Homens e mulheres tem papéis diferentes para desempenhar na sociedade.

- 0 () discordo fortemente
- 1 () discordo moderadamente
- 2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

Eu considero que alguns atos são errados por não serem naturais.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

Em nenhuma situação é correto matar um ser humano.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

Acredito que é moralmente errado que crianças ricas herdem um monte de dinheiro enquanto crianças pobres não herdam nada.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

É mais importante agir pelo sucesso do grupo do que expressar minhas preferências individuais.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

Se eu fosse um soldado e discordasse das ordens do meu superior, eu obedeceria assim mesmo porque é o meu dever.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

Castidade é uma virtude importante e preciosa.

0 () discordo fortemente

1 () discordo moderadamente

2 () discordo um pouco

3 () concordo um pouco

4 () concordo moderadamente

5 () concordo fortemente

APÊNDICE B - Termo de Consentimento Livre e Esclarecido - TCLE

São Paulo, ____ de _____ de 201__

Por favor, considere essas informações cuidadosamente antes de decidir participar desta pesquisa.

Estamos realizando um conjunto de pesquisas, propostas no projeto intitulado “Valores Morais como Resultado de Responder Relacional: um estudo com o procedimento de avaliação de relações implícitas” A pesquisa é coordenada pela estudante de doutorado em Psicologia Experimental Ísis Gomes Vasconcelos, sob orientação do prof. Dr. Marcelo Frota Lobato Benvenuti (Universidade de São Paulo – Instituto de Psicologia). O propósito desta pesquisa é conhecer como as pessoas julgam situações sociais hipotéticas. Você será solicitado a ler algumas afirmações e expressar suas opiniões sobre elas. Não há respostas certas ou erradas diante das afirmações, e também não é possível acessar fatores de personalidade ou inteligência por meio da sua participação. Você não terá prejuízos físicos, cognitivos ou emocionais nem quaisquer vantagens decorrentes de sua participação. Os efeitos de participar desta pesquisa são comparáveis à experiência cotidiana de sentar-se diante de uma tela de computador e utilizar o teclado ou o touchpad. Ao final do estudo, você receberá explicações sobre as questões que motivaram esta pesquisa, e sobre as potenciais implicações dela.

Sua participação neste estudo é completamente voluntária e você pode se recusar a participar ou se retirar da atividade a qualquer momento sem qualquer prejuízo ou penalidade. Sua participação neste estudo é confidencial e nenhuma informação que possa identificá-lo (la) será associada aos seus dados.

Se você tiver quaisquer questões referentes à sua participação ou deseja receber um sumário dos resultados, por favor, entre em contato com a responsável pela pesquisa, Ísis Gomes Vasconcelos, isisgvasconcelos@gmail.com, (11) 95898-5236. Para quaisquer outros problemas relacionados a este estudo, você pode contatar também o orientador responsável, Marcelo Benvenuti, mbenvenuti@yahoo.com.

Para questões, sugestões ou reclamações que não possam ser endereçadas à pesquisadora, ou para relatar algum dano decorrente da participação na pesquisa, por favor, entre em contato com o Comitê de pesquisas com seres humanos do instituto de psicologia da Universidade de São Paulo, localizado na avenida Professor Melo Morais, 1721, bloco G, segundo andar, sala 27. CEP 05508-030, Cidade Universitária. (11) 3091-4182. ceph.ip@usp.br. Informações complementares referentes à assinatura do TCLE ou os direitos do participante da pesquisa podem ser obtidos junto a este comitê.

Este documento foi elaborado em duas vias. Ambas devem ser assinadas. Uma ficará com a pesquisadora responsável pela pesquisa e a outra com o senhor(a).

Termo de concordância:

A natureza e proposta da pesquisa foi suficientemente explicada e eu concordo em participar deste estudo. Eu compreendo que sou livre para me retirar a qualquer momento.

Ísis Gomes Vasconcelos

Assinatura do Participante

APÊNDICE C - Questionário de elegibilidade de participantes no Estudo 2**Questionário de pré-participação em pesquisa**

Nome:

Idade:

Como você se identifica?

Mulher ()

Homem ()

Outro: _____

Você acredita em... (Você pode marcar mais de uma opção):

a) Deus

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

b) Jesus Cristo

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

c) Virgem Maria

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

d) santos

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

e) anjos

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

f) espíritos

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

g) duendes

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

h) gnomos

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

i) demônios

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

j) orixás

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

k) Karma

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

l) vida após a morte

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

m) reencarnação/vidas passadas

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

n) orações

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

o) benzeção

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

p) céu/inferno

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

q) energias/aura

- () 0 = não acredito nem um pouco
- () 1 = acredito muito pouco
- () 2 = acredito moderadamente
- () 3 = acredito completamente

r) astrologia/horóscopo

- () 0 = não acredito nem um pouco
- () 1 = acredito muito pouco
- () 2 = acredito moderadamente
- () 3 = acredito completamente

s) magia

- () 0 = não acredito nem um pouco
- () 1 = acredito muito pouco
- () 2 = acredito moderadamente
- () 3 = acredito completamente

t) destino

- () 0 = não acredito nem um pouco
- () 1 = acredito muito pouco
- () 2 = acredito moderadamente
- () 3 = acredito completamente

u) tarot

- () 0 = não acredito nem um pouco
- () 1 = acredito muito pouco
- () 2 = acredito moderadamente
- () 3 = acredito completamente

v) mantras

- () 0 = não acredito nem um pouco
- () 1 = acredito muito pouco
- () 2 = acredito moderadamente
- () 3 = acredito completamente

w)reiki

() 0 = não acredito nem um pouco

() 1 = acredito muito pouco

() 2 = acredito moderadamente

() 3 = acredito completamente

Vivências religiosas: nesta seção queremos saber mais sobre suas experiências com religiosidade.

Pratica uma religião atualmente? () Sim () Não

Se marcou sim, qual?

Se marcou sim, qual o nome do estabelecimento religioso?

Você frequenta grupos relacionados a uma instituição religiosa? (Ex. Pastorais, grupos de estudo bíblico, apostolados, trabalho voluntário...) () Sim () Não

Se marcou sim, quais?

Se marcou sim, quantas horas em média você passa em atividades relacionadas à religião semanalmente? (Missas, cultos, frequência a outros grupos)

Praticava alguma na infância? () Sim () Não

Se marcou sim, qual?

Praticava alguma na adolescência? () Sim () Não

Se marcou sim, qual?

As pessoas da sua família que moram com você praticam alguma religião?

Mãe

pai

outros_____

Se você marcou algum item da questão anterior, por favor, especifique qual (quais) religião (ões) essas pessoas praticam.

Por favor, descreva brevemente as atividades, práticas ou rituais relacionados à sua vivência religiosa atual.