The Influence of Gender and Sexual Prejudice on Empathic Responding

by
Seth Winward

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Psychology

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

Although studies of the relationship between empathy and racial prejudice have become increasingly popular in recent years, there is comparatively little research on the potential relationship between empathy and sexual prejudice. A handful of studies exclusively using trait measures indicate that sexual prejudice is negatively related to empathy, and the existing literature on sexual prejudice implies that this relationship may be modulated by participant and target gender. We investigated whether these relationships persist using state measures of empathy, outside of previously targeted social contexts. Participants in two samples (n_1 = 99, n_2 = 113) completed trait measures of sexual prejudice and empathy online. A few weeks later, they completed an online study advertised as a memory study. Participants read brief text vignettes describing gay/lesbian or straight male and female characters in emotionally negative and neutral scenarios and rated their empathy for the target character after each vignette. Empathy ratings were significantly higher for negative than neutral scenarios but there was no interaction with target sexual orientation and no further modulation by participant gender. In line with previous research, trait empathy was positively correlated with empathy ratings, negatively correlated with trait sexual prejudice, and was higher for female than male participants. Although the relationship between sexual prejudice, gender, and empathy was not supported at the state level, the results illustrate a potential disconnect between trait and state measures of empathy. We discuss potential explanations for the seemingly differential influence of various forms of social prejudice on state and trait empathy. Issues with habituation and fatigue, which may confound state measures of empathy, are also discussed.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Roxane Itier, for her guidance and support throughout the research process. Even in the midst of a global pandemic, she has always been willing to adapt her enthusiasm and expertise to new topics and research environments. I am sincerely grateful for the opportunities she has provided me with and the effort she has taken to advance my development as a scientist. I would also like to thank Dr. Hilary Bergsieker and Dr. Richard Eibach for their insight and advice during the preparation of this thesis.

I would like to extend my thanks to all my lab mates: Anna Hudson, James Siklos-Whillans, and Zora Chen, as well as former Honours students Amie Durston and Kaitlynn Meier-Ross. I would also like to thank our former lab member Sarah McCrackin for inspiring this research project and providing support and advice early in the process. I would like to thank Bill Eickmeier, Vid Sharda, and Zuhayr Shaikh for their technical expertise in developing our online platform, as well as Dave McLean for his assistance in using the Mass Testing process.

Finally, I would like to extend my heartfelt gratitude to my family: Annamaria Feltracco, Paul Winward, and Eli Winward. Your unconditional support means more to me than I can express in words.

I was supported by a SSHRC Canada Graduate Scholarship (Master's) and the research was supported by an NSERC Discovery grant awarded to my supervisor RJ Itier.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

Empathy is an essential component of social cognition that exercises great influence on how we think, feel, and reason about others. However, the prosocial influence of empathy can be attenuated by many trait- and state-level factors. In recent years, an expanding literature has consistently demonstrated an own-race bias in empathy such that individuals typically express more empathy towards members of their own racial ingroup using self-report, behavioural, and neuroimaging methods. The degree to which the negative relationship between racial prejudice and empathy is generalizable to other forms of prejudice remains largely unknown. A handful of studies seem to demonstrate that this relationship extends to sexual prejudice as well, but this body of research has only used trait measures of empathy and sexual prejudice in their designs thus far (Burke et al., 2015; Marsden & Barnett, 2020; Poteat et al., 2013). Consequently, it is unclear whether a relationship between sexual prejudice and state-level empathy also exists. The present study was designed to answer this question and determine whether previously observed gender effects involved in sexual prejudice further modulate state-level empathy. This research has implications not only for formal theories of empathy as a construct, but also for the use of empathy-inducing interventions to promote prosocial behaviour and reduce intergroup conflict.

1.1 Defining Empathy

Although a precise definition of empathy remains elusive, several common themes have emerged from the psychological literature on the topic. It is generally accepted that empathy includes both affective and cognitive components, and that these elements are heavily interconnected in neurotypical individuals (Batson, 2009; Cuff et al., 2016; Decety & Lamm, 2006; Decety, 2011; Watt, 2007). Affective empathy is usually discussed in terms of affective sharing; this construct is defined by the automatic mirroring of others' emotional states and

expressions, albeit with at least some degree of distinction between the self and the other (Batson, 2009; Cuff et al., 2016; Decety & Jackson, 2006; Decety & Lamm, 2006; Decety, 2011; Hatfield et al., 2009; van Baaren et al., 2009). Several constructs have been described as part of cognitive empathy, the most prominent of which are empathic concern and perspective-taking (Batson, 2009; Cuff et al., 2016; Davis, 1983; Decety & Jackson, 2006; Decety & Lamm, 2006). Empathic concern is a motivational construct defined by "feeling for" others rather than sharing their affect and is believed to mediate prosocial behaviour (Davis, 1983; Decety & Lamm, 2006; Decety & Lamm, 2009). Perspective-taking describes the metacognitive process by which individuals consciously emulate the emotional states of others, incorporating the circumstances surrounding the target (Cuff et al., 2016; Davis, 1983; Decety & Jackson, 2006; Decety & Lamm, 2006; Decety & Lamm, 2009; Watt, 2007). Although these constructs are generally recognized as foundational to the broader concept of empathy, their relationships with each other and empathy in general are still subject to debate. The present study focused on the empathic concern construct because it has been a primary focus of the existing literature on empathy and prejudice described below.

Empathy research is largely dominated by studies using trait measures of empathy such as the Interpersonal Reactivity Index (IRI; Davis, 1983), which can make it difficult to extrapolate results to the state level. Such studies using trait measures suggest that the ability to empathize with others is relatively stable across time and that some people are more adept empathizers than others; these individual differences may even be reflected at the neuroanatomical level (Banissy et al., 2012; Baron-Cohen & Wheelwright, 2004; Cuff et al., 2016; Davis, 1983; Jolliffe & Farrington, 2006; Spreng et al., 2009; Vachon & Lynam, 2016). Therefore, empathy appears to be at least partially a trait-level construct that is subject to some

degree of individual differences. Indeed, one of the most consistent findings in empathy research is that relative to men, women display more empathy for others (Derntl et al., 2009; Di Tella et al., 2020; Groen et al., 2013; Kobach & Weaver, 2012; Longobardi et al., 2019; Sonnby-Borgström et al., 2008; Stępień-Nycz et al., 2021; Toussaint & Webb, 2005). However, a wide variety of state-level factors also influence empathic responding, such as similarity between observer and target, cognitive load, and moral judgement of the target (Contreras-Huerta et al., 2014; Cuff et al., 2016; Fabi & Leuthold, 2018; Fourie et al., 2017; Han, 2018; Meiring et al., 2014; Molenberghs et al., 2016; Sessa et al., 2014; Suleiman et al., 2018). Although it is clear that empathy can be modulated by a variety of both trait and state influences, the relative strength of and interaction between these influences is not yet fully understood. For instance, there is currently no consensus as to whether a state-level influence can affect a trait-level measure. Despite this lack of a theoretical consensus, many researchers have begun using broadly accepted empathic constructs to investigate how empathy affects and is affected by other psychological phenomena.

1.2 Empathy and Sexual Prejudice

Due to its relevance for real-world social issues, research on the relationship between empathy and prejudice has become increasingly popular in recent years. Most of this literature currently focuses on how empathic processing differs based on the race of the target. Studies utilizing a trait-based approach have demonstrated that individuals who score highly on trait measures of empathy such as the IRI (Davis, 1983) tend to have low scores on trait measures associated with prejudice such as social dominance orientation, and individuals who score highly on measures associated with prejudice tend to have low scores on measures of trait empathy (Bäckström & Björklund, 2007; Hudson et al., 2019; Pettigrew & Tropp, 2008; Sidanius et al.,

2013; Stathi et al., 2021; Vanman, 2016). In research on empathy for physical pain, participants tend to underestimate the pain of other-race targets and report less concern for other-race targets' wellbeing relative to same-race targets (Forgiarini et al., 2011; Kaseweter et al., 2012; Suleiman et al., 2018; Trawalter & Hoffman, 2015). This research clearly indicates that there is a negative relationship between racial prejudice and empathy at the trait and state level.

In contrast to this growing literature on empathy and racial prejudice, the relationship between empathy and sexual prejudice has received little attention. Sexual prejudice is defined as prejudice against others based on their sexual orientation or sexual behaviour and is most commonly used to refer to prejudice against LGBTQ+ people (Herek & McLemore, 2013; Herek, 2000). There is extensive research demonstrating that sexual prejudice is rooted in the emotional disgust response (Crawford et al., 2014; Herek & McLemore, 2013; Herek, 2000; Inbar et al., 2009; Kiebel et al., 2017; Mahaffey et al., 2005; O'Handley et al., 2017; Olatunji, 2008; Shields & Harriman, 1984; Terrizzi et al., 2010). Therefore, sexual prejudice is an intrinsically affective process and should logically have an influence on empathy. Despite this apparent link, only a few studies have begun to explore the relationship between sexual prejudice and empathy.

For example, Burke et al. (2015) sought to determine whether contact with LGBTQ+ people and individual differences in trait empathy moderate sexual prejudice among a very large sample of medical students using measures of implicit and explicit sexual prejudice. Although no association between empathy and implicit bias was observed, both the empathic concern and perspective-taking subscales of the IRI were negative predictors of explicit bias against gays and lesbians (Burke et al., 2015). In another study, Marsden and Barnett (2020) examined trait empathy as a mediator between sociopolitical ideology and sexual prejudice in a large sample of

straight U.S. college students. They also found that trait empathic concern and trait perspective-taking were negatively related to sexual prejudice. In addition, trait empathy significantly mediated the relationship between conservative sociopolitical ideology and sexual prejudice such that conservatives scored lower on the measures of empathic concern and perspective-taking relative to liberals; these lower scores predicted higher levels of sexual prejudice among conservatives (Marsden & Barnett, 2020). In a study of adolescents, Poteat et al. (2013) included trait empathy as part of a broader investigation of factors contributing to anti-LGBTQ+ bullying. Their analysis revealed that in another parallel to research on empathy and racial prejudice, empathic concern and perspective-taking were both negatively correlated with sexual prejudice such that individuals who were high in trait empathic concern and trait perspective-taking tended to have lower scores on trait measures of sexual prejudice (Poteat et al., 2013). However, perspective-taking did not remain a significant predictor when entered alongside the full regression model, which also included media messages, LGB friends, parent attitudes, identity importance and classroom norms as predictors; empathic concern remained a significant and negative predictor of anti-LGBTQ+ bullying in the full model (Poteat et al., 2013). Although these studies paint a fairly consistent picture of negative relationships between empathic constructs and sexual prejudice, some limitations are worth highlighting.

1.3 Limitations of the Existing Literature

First and foremost, the above studies exclusively use trait measures of empathy; they do not consider state-based influences in their study design and provide no measures of empathic behaviour. This limited perspective does not do justice to the full scope of empathy, which is clearly influenced by state factors (Contreras-Huerta et al., 2014; Cuff et al., 2016; Fabi & Leuthold, 2018; Fourie et al., 2017; Han, 2018; Meiring et al., 2014; Molenberghs et al., 2016;

Sessa et al., 2014; Suleiman et al., 2018). Therefore, it is somewhat unclear as to whether the results of the above studies are generalizable beyond the trait level. Furthermore, the studies described above all take their measures of empathic concern and perspective-taking from the IRI. Despite being the most commonly used measure of trait empathy, recent psychometric evaluations have revealed troubling issues with its factor structure as well as relatively low convergent validity with other trait and behavioural measures of empathy (De Lima & Osório, 2021; Murphy & Lilienfeld, 2019). Although studies using the IRI are valuable contributions to the literature, their results might not be generalizable to other measures of empathy, especially measures of empathic behaviour. While existing studies appear to demonstrate negative relationships between cognitive empathic constructs and sexual prejudice, their focus on trait measures in general, and the IRI in particular, limits their conclusions.

Another limitation of these existing studies is that they ignore a unique aspect of sexual prejudice. Despite replicating some well-documented gender effects on empathy, none of the studies described above investigated the gender effects typically associated with sexual prejudice. In the broader literature on sexual prejudice, behavioural and self-report studies have consistently demonstrated straight greater sexual prejudice for same-gender targets relative to other-gender targets (Ahrold & Meston, 2008; Herek, 2002; Herek & McLemore, 2013; Herek, 2000; Herek & Gonzalez-Rivera, 2006; Kite, 1996; Mahaffey et al., 2005; Monto & Supinski, 2014), especially among male participants (Herek & McLemore, 2013; Herek, 2000; Parrott & Gallagher, 2008). For instance, if a straight male participant encounters a gay/lesbian target, he will likely express more prejudice and disgust than he would towards a lesbian target, whereas the converse is true with straight female participants. Although this gender bias is well represented in research on sexual prejudice, at present there are no published studies that assess

whether the effect of sexual prejudice on empathy depends on the gender of the perceiver and the target at either the trait or state level.

1.4 The Present Study

To fill the gaps in the literature discussed above, the present study was designed to determine whether sexual prejudice has a negative relationship with empathy at both the trait and state level, as well as whether this relationship is modulated by target and/or participant gender. Instead of relying solely on trait measures of empathy, the present study utilized a series of brief empathy-inducing written vignettes based on sentences validated by a previous study (McCrackin & Itier, 2021) and asked participants to rate their affect and level of empathic concern for the target character in each vignette, which had either negative or neutral emotional valence. The neutral vignettes provided a baseline against which we compared responses to the negatively valenced vignettes. Within these vignettes, we manipulated the sexual orientation of the target character to determine whether participants were less empathetic towards gay/lesbian characters than toward straight characters. We also manipulated the gender of the target in the vignettes to assess whether the same-gender effect associated with sexual prejudice influenced the relationship between sexual prejudice and empathy. We also sought to begin exploring the generalizability of previous trait results by using a different measure of trait empathy, the Toronto Empathy Quotient (TEQ), a unidimensional scale which has greater convergent validity with behavioural measures and is most closely associated with the empathic concern construct that is the focus of the present study (De Lima & Osório, 2021; Spreng et al., 2009). However, because the temporal stability of the TEQ has not yet been tested in any published literature, we also sought to analyze this psychometric property as a secondary objective.

Our main hypothesis was that state-level empathy has a similar relationship to sexual prejudice as trait-level empathy, such that participants would tend to be less empathetic to the misfortunes of gay/lesbian characters relative to straight characters. We also hypothesized that the same-gender effects reported in studies of sexual prejudice would affect this negative relationship such that it would be even stronger when participants and targets were of the same gender. Based on previous literature, we hypothesized that our measure of trait empathy would be positively related to our measures of state empathy and that trait empathy would be negatively related to trait sexual prejudice. Furthermore, we hypothesized that previously observed gender differences in empathic responding would be observed at both the trait and state levels.

1.5 Predictions

In regards to our primary analysis of empathy ratings (our measure of state empathy) using repeated measure ANOVAs, we predicted that there would be a significant main effect of vignette valence such that participants would express more empathy for targets during the negative vignettes relative to the neutral ones; this result was a manipulation check (Prediction 1), based on other studies on state empathy using vignette paradigms (Arditte Hall et al., 2018; Gehenne et al., 2021; Hein et al., 2018; McGrath & Haslam, 2020; Vera Cruz & Mullet, 2019). Because empathic concern would be expressed during the negative condition, we predicted that there would be a significant interaction between vignette valence and our variable of interest (target sexual orientation), such that participants would express less empathy for gay/lesbian targets relative to straight targets within negative vignettes (Prediction 2). This decreased empathy for gay/lesbian targets was expected to be more pronounced for same-gender gay/lesbian targets relative to other-gender gay/lesbian targets, based on the gender effects observed in the literature on sexual prejudice (Ahrold & Meston, 2008; Herek, 2002; Herek &

McLemore, 2013; Herek, 2000; Herek & Gonzalez-Rivera, 2006; Kite, 1996; Mahaffey et al., 2005; Monto & Supinski, 2014); therefore we predicted that there would be a significant four-way interaction between participant gender, target gender, target sexual orientation, and vignette valence (Prediction 3). Furthermore, we predicted that trait sexual prejudice, as measured by the Modern Homonegativity Scale (MHS; (Morrison & Morrison, 2003) would be negatively correlated with participants' empathy ratings for vignettes with gay/lesbian targets, indicating that higher trait sexual prejudice is related to a lesser state empathic response to gay/lesbian targets (Prediction 4).

In regards to our correlational analyses using trait measures, we predicted that scores on the TEQ would be negatively correlated with MHS scores (Prediction 5) because existing research demonstrates a negative relationship between other measures of trait empathy like the IRI and measures of both sexual and racial prejudice (Bäckström & Björklund, 2007; Burke et al., 2015; Marsden & Barnett, 2020; Pettigrew & Tropp, 2008; Poteat et al., 2013; Sidanius et al., 2013; Stathi et al., 2021) Based on previous literature demonstrating the TEQ's high external validity with state measures of empathy (De Lima & Osório, 2021; Spreng et al., 2009), we predicted that scores on the TEQ would be positively correlated with average empathy ratings (across vignettes), such that participants with higher trait empathy scores would tend to express more state empathy for the character regardless of condition (Prediction 6).

In terms of analyses evaluating the TEQ, we predicted that female participants would have significantly greater scores on the TEQ than male participants for both iterations of this scale (Prediction 7), based on the sex differences in empathy reported with other empathy scales (Derntl et al., 2009; Di Tella et al., 2020; Groen et al., 2013; Kobach & Weaver, 2012; Longobardi et al., 2019; Sonnby-Borgström et al., 2008; Stępień-Nycz et al., 2021; Toussaint &

Webb, 2005). Based on the assumption that the TEQ is stable across time, we predicted that scores on the first iteration of the TEQ (hereafter TEQ1) would be significantly correlated with scores on the second iteration of the TEQ (hereafter, TEQ2; Prediction 8) and that the two scores would not be significantly different from one another (Prediction 9). Because we expected scores on the two iterations of the TEQ to be very similar, we also predicted that TEQ2 scores would be significantly and positively correlated with average empathy ratings across conditions (Prediction 10) and significantly and negatively correlated with MHS scores, just like TEQ1 scores (Prediction 11).

2. METHODS

2.1 Participants

The present study recruited two different groups of participants; one was composed of undergraduate students at the University of Waterloo while the other was recruited through Amazon's Mechanical Turk (hereafter MTurk) system. For both samples, participants were required to be 18-30 years of age, living in either the United States or Canada and required to self-identify as fluent English speakers to be eligible for the study. All participants in both samples received informed written consent prior to participation. The study was conducted according to the Declarations of Helsinki and was approved by the UW Research Ethic Board. Although we initially intended to combine the samples for final analysis, inspection of the data revealed that our two samples were different both in terms of demographics (see Table 1) and range of scores on the various scales administered (see and Table 2 and Results). Previous research indicates that demographic factors can have significant effects on empathy (Brewer, 2003; Derntl et al., 2009; Di Tella et al., 2020; Groen et al., 2013; Herek & Gonzalez-Rivera, 2006; Hicks & Lee, 2006; Kobach & Weaver, 2012; Kozloski, 2010; Longobardi et al., 2019; Sonnby-Borgström et al., 2008; Stępień-Nycz et al., 2021; Toussaint & Webb, 2005). Furthermore, our samples differed in terms of ethnicity and this may indicate that levels of sexual prejudice could differ across samples as ethnic differences often reflect cultural differences and sexual prejudice is sometimes influenced by culture (Herek & Capitanio, 1995; Kite et al., 2019; Lee, W. & Cunningham, 2014; Licciardello et al., 2011; Rosenthal et al., 2011). Therefore, both samples were analyzed separately; characteristics of each sample are described below.

2.1.1 SONA Participants

This study recruited 276 undergraduate student participants from the University of Waterloo SONA system in exchange for course credit. After exclusion criteria were applied (see Results), the final SONA sample was composed of 94 participants (age range 18-30; $M = 20.26$, $SD = 2.06$). Other demographic information is presented in Table 1 below.

Sample	Gender	*n*	Ethnicity	*n*	Sexual Orientation	*n*
SONA	Male	38	Black/African	2	Asexual	2
(*n*=94)	Female	56	East Asian	21	Bisexual	10
			Hispanic	2	Heterosexual	72
			Indigenous	1	Homosexual	1
			Middle Eastern	7	Pansexual	5
			Mixed	5	Queer	2
			South Asian	16	Decline	2
			Southeast Asian	4		
			White	35		
			Decline	1		
MTurk	Male	58	Black/African	8	Bisexual	4
(*n*=113)	Female	55	East Asian	6	Heterosexual	108
			Hispanic	9	Decline	1
			Middle Eastern	1		
			Mixed	4		
			South Asian	3		
			Southeast Asian	6		
			White	76		

Table 1. Demographic information of the final SONA sample (top row) and MTurk sample (bottom row).

2.1.2 MTurk Participants

This study initially recruited 211 participants from Amazon's Mechanical Turk system, accessed through the CloudResearch (formerly TurkPrime) platform. These participants were offered $3.00 for completing the prerequisite study and $12.00 for completing the follow-up main study, the details of which are described below. After rejection criteria, the final MTurk

sample consisted of 113 participants (age range 20-30; mean age = 26.95, *SD* = 2.55). Other demographic information is presented in Table 1 above.

2.2 Materials

2.2.1 Mass Testing/Prerequisite Study

Both SONA and MTurk participants completed a preliminary series of scales prior to the main study. SONA participants completed these scales as part of the Mass Testing process completed by all SONA users at the University of Waterloo at the beginning of every semester, while MTurk participants completed a brief prerequisite study that was created to mimic the Mass testing process of SONA participants and was required to be eligible for the main study. All questionnaires were presented using the Qualtrics survey platform and were presented in semi-random (SONA) or random (MTurk) order. Only the TEQ MHS, and sexual orientation questionnaire were analyzed (see Table 2 for score range). The order of materials in the Mass Testing/Prerequisite Study phase is depicted in Panel A of Figure 1 below.

Sample	Scale	Minimum	Maximum	Mean	*SD*
SONA	TEQ1	1.20	2.87	2.30	0.36
	MHS	2.17	3.33	2.70	0.27
	TEQ2	1.27	3.27	2.30	0.45
MTurk	TEQ1	1.13	3.93	2.81	0.65
	MHS	0.92	5.00	2.52	1.10
	TEQ2	1.20	3.47	2.31	0.55

Table 2. Descriptive information for TEQ and MHS scores within each sample.

2.2.1.1 Toronto Empathy Quotient (TEQ)

The TEQ (Spreng et al., 2009) is a measure of trait empathy. This questionnaire consists of 16 statements such as “When someone else is feeling excited, I tend to get excited too” and “It upsets me to see someone being treated disrespectfully”; participants are asked to rate how frequently they feel or act in the manner described by the statement. Possible responses range from 0 (*Never*) to 4 (*Always*”). High scores on this scale indicate that a participant is high in trait

empathy, with possible total scores ranging from 0 to 64 in the interpretation of data described by Spreng et al. (2009). However, in the present study participants' scores on this scale are expressed as the average of each participants' responses such that their average scores range from 0 to 4. We chose averages over summation so that participants with some missing scale items could still provide meaningful data and to make comparisons with the original scale anchors much simpler. Existing literature on the psychometric properties of this scale suggests that it is a sound measure of general empathy, with good construct validity and internal consistency (Kourmousi et al., 2017; Totan et al., 2012; Xu et al., 2020). Note that due to clerical error, the TEQ was initially presented to participants without the final item. All results reported below thus use the 15-item version of the TEQ and exclude the final sixteenth item.

2.2.1.2 Modern Homonegativity Scale (MHS)

The MHS (Morrison & Morrison, 2003) is a measure of trait sexual prejudice consisting of 12 statements such as "many homosexuals use their sexual orientation so that they can obtain special privileges" and "Homosexuals seem to focus on the way in which they differ from heterosexuals, and ignore the ways in which they are the same." Participants are asked to rate their level of agreement with these statements on a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree).* High scores on the MHS indicate high levels of sexual prejudice; as above we expressed participants' results as their average MHS score with a possible range from 1 to 5. The MHS was designed to measure the more subtle forms of sexual prejudice that are common among younger participants and is highly regarded in terms of its construct validity and reliability, having been validated and re-evaluated several times in the years since its original publication (Grey et al., 2013; Morrison & Morrison, 2003; Morrison et al., 2009; Morrison et al., 2005; Peterson et al., 2017; Romero et al., 2015; Rye & Meaney, 2010). We note, however,

that this scale seems to have been developed and validated mainly within Western cultures and that it is now 20 years old.

2.2.1.3 Sexual Orientation Questionnaire

To assess more sensitive demographic information without revealing the purpose of the study, the researchers added a brief sexual orientation questionnaire to be presented alongside the Mass Testing/Prerequisite Study scales. This questionnaire simply asked participants to identify their sexual orientation and gender identity (see Table 1).

2.2.2 Main Study

Participants were invited to take part into a study on memory for social information. This deception was necessary to avoid participants knowing the true aim of the study, which would have modulated their responses. The order of materials in the main study phase is depicted in Panel B of Figure 1 below.

2.2.2.1 Demographics Questionnaire

All participants completed a simple demographics questionnaire at the beginning of the main study. This questionnaire collected information about the participants' gender, age, and ethnicity. Although this information was redundant with that collected during Mass Testing and the Prerequisite Study, the data was used to confirm that we had recruited the same participants for both parts of the study and that they were truthful in their responses.

2.2.2.2 Vignettes

The main study included 125 brief written vignettes based on the emotional and empathy-inducing sentence stimuli developed by (McCrackin & Itier, 2021). Five of these were practice trials, with the target gender, partner gender, and valence selected at random. In addition to the practice trials, a positive valence condition consisting of 24 trials (6 male gay/lesbian vignettes, 6

male straight vignettes, 6 female gay/lesbian vignettes, and 6 female straight vignettes) was included to reduce monotony and prevent the study from having an overwhelmingly negative valence. Responses to trials in this condition were not analyzed. The remaining 96 vignettes were divided evenly into eight conditions according to a factorial design with 2 valence (neutral, negative) x 2 target gender (male, female) x 2 target sexual orientation (gay/lesbian, straight).

The 12 vignettes making up each of the eight main conditions were selected at random from 25 possible generic vignettes to make the likelihood of two conditions being made up of the same combination of vignettes extremely small. For instance, the generic version of one vignette is "(CHARACTER NAME) is a (AGE) year old living in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). Outside of work, (CHARACTER NAME) enjoys (HOBBY) and taking care of [HIS/HER] pets. (CHARACTER NAME)'s pet dog was {FED/KILLED} yesterday afternoon." In each vignette, the character's name was randomly selected from a list of unambiguously masculine or feminine names depending on the target gender for that condition. Similarly, the partner's name was randomly selected from one of the two lists of names depending on target sexual orientation for the condition; the gendered names allowed the participants to infer whether the target character was gay/lesbian or straight. Note that the sexual orientation of the target is only implied through the names of the target and their partner rather than stated explicitly. The valence of the vignette was determined by the final sentence, with a key word or phrase changing to make the scenario either emotionally negative or emotionally neutral based on the condition.

The target's age, location, and hobby were selected at random from lists of 30 options each (see Appendix A for details on the vignettes). Although these details were not strictly relevant to the variables of interest and did not vary according to condition, they were

randomized to create variety. In the Negative Male Gay/Lesbian condition, for example, the generic vignette provided above might be presented as “Michael is a 25 year old living in Toronto with his partner, Matthew. Outside of work, Michael enjoys woodworking and taking care of his pets. Michael’s pet dog was killed yesterday afternoon.” The vignettes making up each condition were sorted at random such that there were always 12 different vignettes from each condition for every participant, all intermixed at random along with the positive trials.

2.2.2.3 Vignette Questions

After reading each vignette, participants were asked to answer four questions. The first two questions appeared on the same screen as the vignette. The first question asked participants to rate their level of empathy for the main character in the vignette on a 9-point Likert scale using the mouse cursor to click and drag a sliding scale from 1 (*Very little empathy*) to 9 (*Extreme empathy*). The second question asked participants to rate their emotional valence after reading the vignette on another 9-point Likert scale, from “1 – Very negative” to “9 – Very positive”. These questions were identical to those used by McCrackin and Itier (2021) to validate the emotional stimuli that form the basis of the vignettes in the present study. The exact text of the questions was “Please rate how much empathy you feel for (CHARACTER NAME), where a rating of 1 means very little empathy and a rating of 9 means extreme empathy” and “Please rate the valence of the emotion elicited by the previous vignette, where a rating of 1 means very negative and a rating of 9 means very positive”, respectively. After providing these ratings, participants were presented with a second screen containing two more questions asking them to recall two of the three following details from the preceding vignette: the target’s age, the target’s location, or the target’s hobby. Participants typed in their answers in the two text boxes provided. These memory questions were included to ensure participants were paying attention to the

vignettes and to give credence to the notion that the present study was examining memory for social information (see Procedure below). Participants' responses to these questions were used only for data screening and not analyzed further.

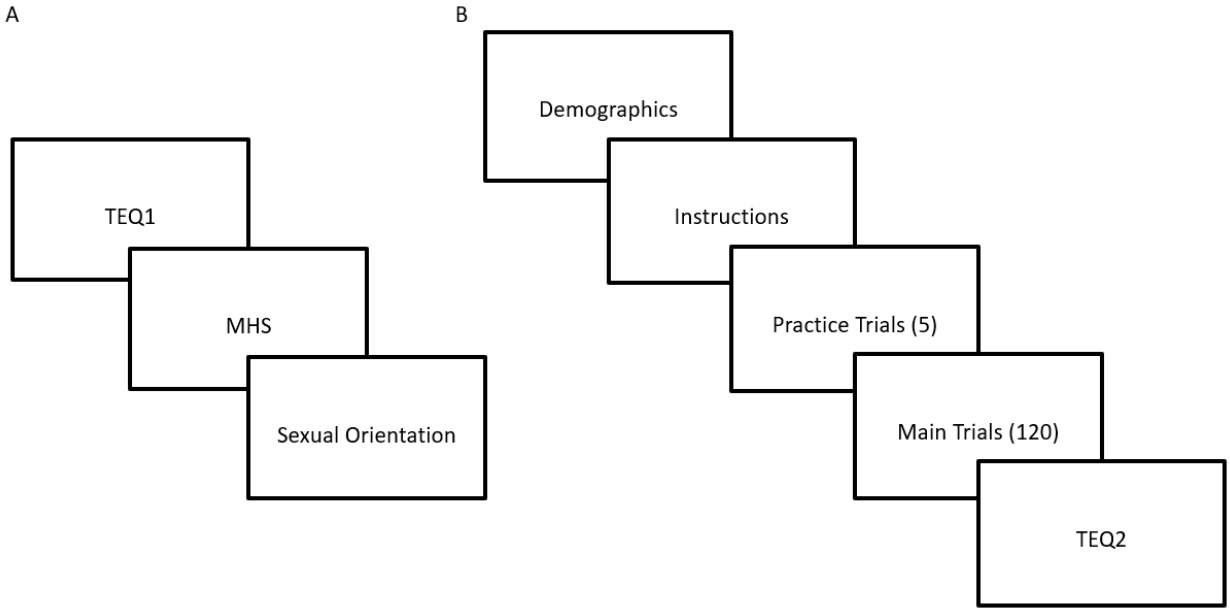

Figure 1. Materials and procedure for both phases of the study. Panel A: Materials for Mass Testing/Prerequisite Study phase. Panel B: Materials for the main study phase, administered a few weeks after the Mass Testing/Prerequisite Study phase.

2.3 Procedure

2.3.1 Procedure for SONA Participants

At the beginning of each term and prior to the main study, all participants completed the Mass Testing survey, which included the questionnaires described above. Participants were then recruited to the main study through an online notice on the University of Waterloo's SONA platform, which contained a link to the study on our custom-made online platform.

They were first presented with an information and consent form and were required to give consent before continuing. Next, the participants completed the demographics questionnaire. The participants then received instructions on how to respond to the vignettes and their questions and completed five practice trials for which the responses were not recorded. The

practice vignettes were followed by 120 experimental trials (the 96 main vignettes mixed with the 24 positive vignettes) described above. Vignettes were presented in random order, but there were always 12 vignettes from each of the 8 main conditions.

Participants responded to each vignette by rating their level of empathy and the valence of their emotional state on 9-point Likert scales presented below the vignette. On a separate screen after each vignette and the accompanying scales, participants were asked to answer two of three possible questions about the characters in the vignette they had just read; they were asked to report the target's age, location, or hobby. The two questions that were presented were chosen at random from the three possible options. After answering these questions, the participants continued to the next vignette until they had completed all 120 trials. Every 25 vignettes, participants viewed a brief pop-up that asked them to solve a simple algebra problem as an attention check. At the same time, they received a reminder to take a brief break as needed to prevent fatigue.

After responding to each vignette, participants were asked to complete the TEQ so that a stability analysis could be conducted as described above. Following this questionnaire, participants read a debriefing letter that informed them of the true nature of the study and explained the necessity of the deception. They were then asked to read another, entirely truthful information and consent letter and re-affirm their consent. They were then redirected back to the SONA platform and their credits were granted within 24 hours. The procedure is depicted in Figure 1.

2.3.2 Procedure for MTurk Participants

Participants were initially recruited to the Prerequisite Study through an online notice on the MTurk platform. In the study description, participants were informed that completing this

study would make them eligible for the main study, for which there was greater remuneration. The description also contained a link that directed participants to the Prerequisite Study, a survey hosted on the Qualtrics platform. Participants were first presented with an information and consent form and were required to give consent before continuing. Participants then completed the demographics questionnaire, TEQ, MHS, sexual orientation questionnaire, and 5 more questionnaires[1], all presented in random order. After completing the scales, participants received a feedback letter providing them with more information about the study and were redirected to MTurk. Their remuneration was granted within 24 hours.

One week after completing the Prerequisite Study, all participants received an email notifying them that the main study had become live and that they were eligible to participate. Two weeks after completing the Prerequisite Study, all remaining eligible participants received the same email again. No further attempts to contact eligible participants were made. Participants accessed the main study through the same platform as the Prerequisite Study. Note that the study was listed as "Effects of Emotional Context and Social Memory" and participants were given a false description of the study that did not mention empathy or sexual prejudice to prevent social desirability and experimenter expectation biases. Upon clicking this link, they completed the main study in the same manner as the SONA participants described above. The procedure is depicted in Figure 1.

[1] The Prerequisite Study also included the Attention-Related Cognitive Errors Scale (ARCES), Deep Effortless Concentration-Internal (DEC-I) scale, Deep Effortless Concentration-External (DEC-E) scale, Mind Wandering: Deliberate (MW:D) scale, and Mind Wandering: Spontaneous (MW:S) scale. These scales were not analyzed and had no relevance to the present study; they were simply included to make the Prerequisite Study more similar to the Mass Testing process.

3. RESULTS

3.1 Data Screening and Preprocessing

Prior to analysis, we rejected all participants who did not consent to having their data used after debriefing (see Procedure; SONA n = 7, MTurk n = 1) as well as those who withdrew from or otherwise did not complete the study (SONA n = 66, MTurk n = 54). We also rejected those who repeated the same response to the empathy rating question (see Materials) for more than 10% of trials in a row (SONA n = 30, MTurk n = 4), as we believe such responses indicate that participants were not responding genuinely. Those who failed to answer at least one memory question (see Materials) correctly on 20% or more trials were also rejected (SONA n = 13, MTurk n = 5), as we believe these participants were not paying attention to the vignettes. Based on experimenter experience testing the study, we did not deem reasonable for a single session to take longer than an hour and 45 minutes, and we could not account for what participants might be doing while taking extended breaks. Similarly, any participant who completed the study in less than 45 minutes was most likely not reading all the vignettes and question prompts. We thus rejected those who took either longer (SONA n = 44, MTurk n = 2) than one hour and 45 minutes or less than 45 minutes (SONA n = 22, MTurk n = 32) to complete the study. For the remaining participants not rejected by these criteria, all trials in which a participant did not answer at least one of the memory questions correctly were rejected from analysis. We believe that if participants could not recall the answer for either of the memory questions, they were most likely not paying attention to the vignette during that trial and their response should therefore be discounted. The number of trials rejected per participant ranged from 0 to 14 in the SONA sample (M = 3.69, SD = 3.16) and from 0 to 21 (M = 5.72, SD = 4.81) in the MTurk sample.

For all analyses, a series of data screening techniques were applied to ensure that the statistical tests could be meaningfully interpreted and prevent any violations of the assumptions upon which the tests were based. The following processes were applied to all dependent variables in each data set in the analyses described below. First, univariate outliers were identified as being three standard deviations above or below the mean of a particular variable; these outliers (11 among all analyzed variables in SONA sample and 6 in MTurk sample) were winsorized such that they were replaced with the score corresponding to three SDs above or below the mean. To assess normality, we adopted Kline's (1998) cut-off values for normality at |skew| < 3 and |kurtosis| < 10. Descriptive statistics for all variables of interest in both samples revealed no skew values greater than ±3 and no kurtosis values greater than ±10, indicating approximately normal distributions for all variables (see Appendix C). Visual inspection of Q-Q plots for each variable was consistent with this assessment; the assumption of normality was not violated for any of the present analyses. Since each factor in the ANOVA had no more than two levels, the sphericity of the data is not a cause for concern. Tests of homogeneity of variance were conducted as part of each primary ANOVA analysis and are described below.

3.2 Empathy Ratings Analysis

The empathy ratings analysis sought to test Predictions 1, 2, and 3; it compared participants' mean empathy ratings in each vignette condition according to participant gender, vignette valence, target gender, and target sexual orientation using a 2 × (2 × 2 × 2) mixed factorial repeated-measures ANOVA with $\alpha = .05$. Post-hoc simple effects tests were used to decompose significant interactions; Bonferroni corrections were used for multiple comparisons. To determine the degree to which non-significant results supported the null hypothesis, Bayes factors were generated for each predicted but non-significant interaction by comparing Bayes

factors for the relevant effect across all possible models including that interaction. Bayesian analysis was conducted using JASP version 0.16.3 while all other analyses were conducted using SPSS version 28.0.1.0. Interpretation of the Bayes Factors was conducted according to the criteria described by Lee and Wagenmakers (2013).

3.2.1 SONA Participants

Within the SONA sample, Levene's tests of homogeneity of variance revealed that this assumption was not violated for empathy ratings in any condition (all *p*s < .05). The main ANOVA revealed a significant main effect of vignette valence, $F(1, 92) = 356.237$, $MSE = 852.702$, $p < .001$, $\eta^2 = .795$ (Prediction 1), with larger scores for the negative than the neutral condition, indicating that the valence manipulation check was successful (see Figure 2) . A significant interaction between vignette valence and participant gender, $F(1, 92) = 9.084$, $MSE = 21.744$, $p = .003$, $\eta^2 = .090$, was driven by larger empathy scores among female ($M = 7.415$, $SD = 0.145$) than male ($M = 6.649$, $SD = 0.175$) participants in the negative condition only, $F(1, 92) = 11.356$, $MSE = 13.281$, $p = .001$, $\eta^2 = .110$ (Figure 2). The predicted interaction between vignette valence and target sexual orientation was not significant, $F(1, 92) = 0.006$, $MSE = .001$, $p = .940$, $\eta^2 = .000$ (in contrast to Prediction 2), as was the four-way interaction, $F(1, 92) = 0.178$, $MSE = .029$, $p = .674$, $\eta^2 = .002$ (in contrast to Prediction 3). The Bayes factor for the interaction between target sexual orientation and valence was 0.016, indicating very strong support for the null hypothesis. The Bayes factor for the four-way interaction was < .001, indicating extremely strong support for the null hypothesis. All other main effects and interactions were not significant. Significant effects are displayed in Figure 2 below.

3.2.2 MTurk Participants

Within the MTurk sample, a series of Levene's tests revealed that the assumption of homogeneity of variance was not violated for empathy ratings in any condition (all *p*s > .05). The primary ANOVA revealed a significant main effect of vignette valence (Figure 2), such that participants reported greater empathy for target characters in negative than neutral vignettes, $F(1, 111) = 264.571$, $MSE = 948.161$, $p < .001$, $\eta^2 = .704$ (supporting Prediction 1). A significant main effect of participant gender was also found, driven by women reporting more empathy than men, $F(1, 111) = 4.092$, $MSE = 31.687$, $p = .045$, $\eta^2 = .036$, which was qualified by a significant interaction between vignette valence and participant gender, $F(1, 111) = 4.661$, $MSE = 16.702$, $p = .033$, $\eta^2 = .040$. Simple effect analyses revealed that female participants reported higher empathy ratings ($M = 7.409$, $SD = 0.146$) than male participants ($M = 6.782$, $SD = 0.142$) when the vignette valence was negative, $F(1, 111) = 10.118$, $MSE = 11.800$, $p = .002$, $\eta^2 = .084$, but no gender difference was seen for the neutral empathy condition (Figure 2). Finally, there was no significant interaction between target sexual orientation and vignette valence, $F(1, 111) = 2.010$, $MSE = 0.299$, $p = .159$, $\eta^2 = .018$ (contrary to Prediction 2), and no significant four-way interaction, $F(1, 111) = 0.061$, $MSE = 0.012$, $p = .805$, $\eta^2 = .001$ (contrary to Prediction 3). The Bayes factor for the interaction between target sexual orientation and valence was 0.038, indicating strong evidence for the null hypothesis. The Bayes factor for the four-way interaction was < .001, indicating extremely strong support for the null hypothesis. All other main effects and interactions were not significant.

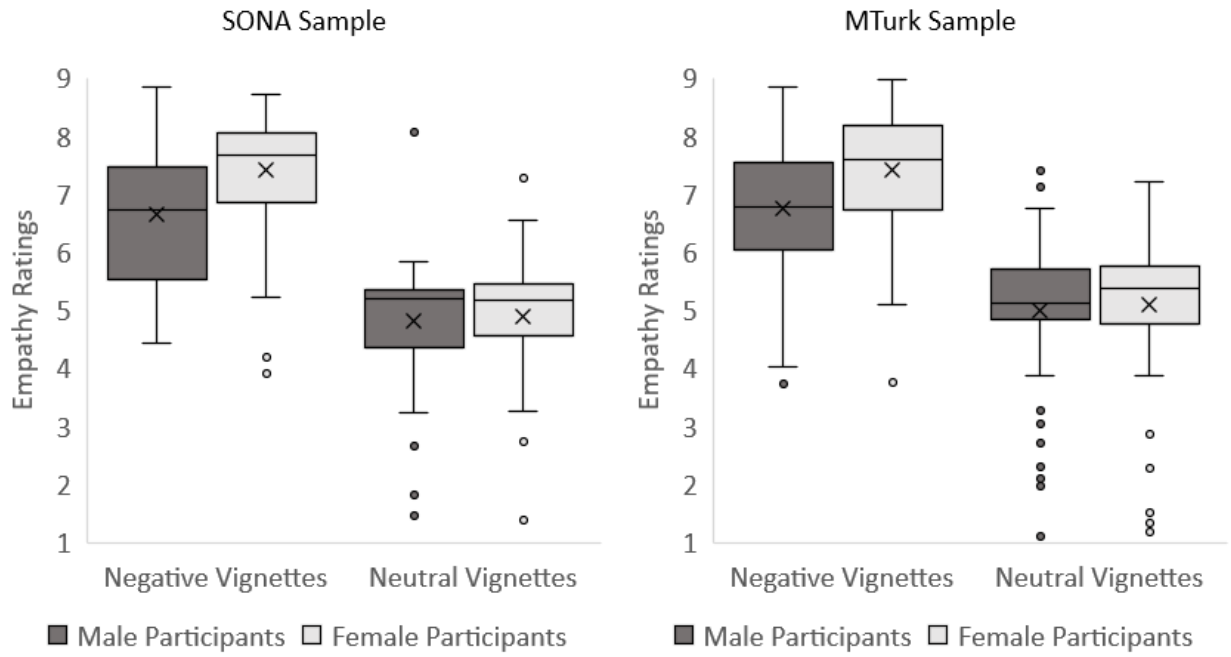

Figure 2. Mean empathy ratings (state empathy) in response to negative and neutral vignettes for male and female participants. Note the larger empathy scores for the negative compared to the neutral conditions, and for female compared to male participants for the negative empathy condition only. Left: SONA participants. Right: MTurk participants.

3.3 Correlational Analyses

A series of Pearson correlations were conducted to determine the strength of the linear relationships between participants' mean (state) empathy ratings and average scores on the MHS (trait sexual prejudice) and TEQ (trait empathy) in accordance with Predictions 4-6. To represent the difference between the baseline level of state empathy (conceptualized by empathy ratings in the neutral conditions) and the actual empathy-eliciting negative condition, participants' empathy rating mean difference scores between the negative and neutral versions of each condition were created and used in all correlations. We believe these different scores better represent state empathy as they take into account the baseline level of empathy elicited during the neutral vignettes; they will be referred to hereafter as state empathy scores. Two aggregate variables were created. First, we computed the mean of each participant's state empathy scores across all conditions (averaging male gay, male straight, female lesbian, and female straight conditions).

We also calculated the mean of each participants' state empathy scores across the conditions with gay/lesbian targets (average of male and female gay/lesbian conditions). According to Kline's (1998) normality criteria, all state empathy scores and aggregate scores were normally distributed. All Pearson correlations were calculated using SPSS version 28.0.1.0 with $\alpha = .05$ and Bayes factors were obtained for each predicted but non-significant correlation using JASP version 0.16.3 and interpreted as described above (Lee & Wagenmakers, 2013).

3.3.1 SONA Participants

The correlation between MHS scores and mean state empathy scores in the average gay/lesbian condition was not significant, $r = -.076$, $p = .468$, contrary to Prediction 4. The Bayes factor obtained for this correlation was 0.141, indicating moderate support for the null hypothesis. However, MHS scores did correlate significantly with TEQ1 scores (i.e. scores on the first/mass testing iteration of the TEQ), $r = -.417$, $p < .001$, supporting Prediction 5. A significant positive correlation was also observed between TEQ1 scores and average state empathy scores across conditions, $r = .207$, $p = .046$, supporting Prediction 6. The reported correlations within the SONA sample are depicted in Figures 3, 4, and 5 below.

3.3.2 MTurk Participants

No significant correlation was observed between MHS and mean state empathy scores averaged across the gay/lesbian conditions, $r = -.140$, $p = .140$ (contrary to Prediction 4). However, as expected, a significant and negative correlation was detected between MHS and TEQ1 scores, $r = -.227$, $p = .016$ (Prediction 5). The Bayes factor obtained for the first non-significant correlation was 0.418, indicating anecdotal evidence in favour of the null hypothesis. In accordance with our predictions, there was a significant positive correlation between TEQ1

scores and mean state empathy scores averaged across all conditions, $r = .364$, $p < .001$

(Prediction 6). All correlations reported above are illustrated in Figures 3, 4, and 5 below.

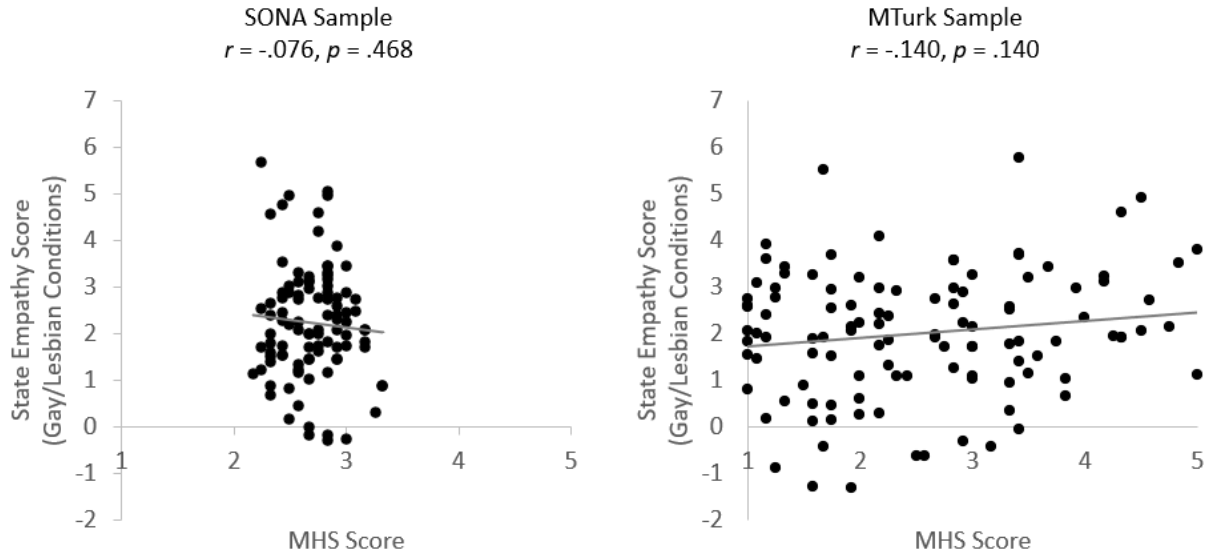

Figure 3. Non-significant correlations between MHS scores (trait sexual prejudice) and mean state empathy difference scores averaged across gay/lesbian conditions. Left: SONA participants. Right: MTurk participants.

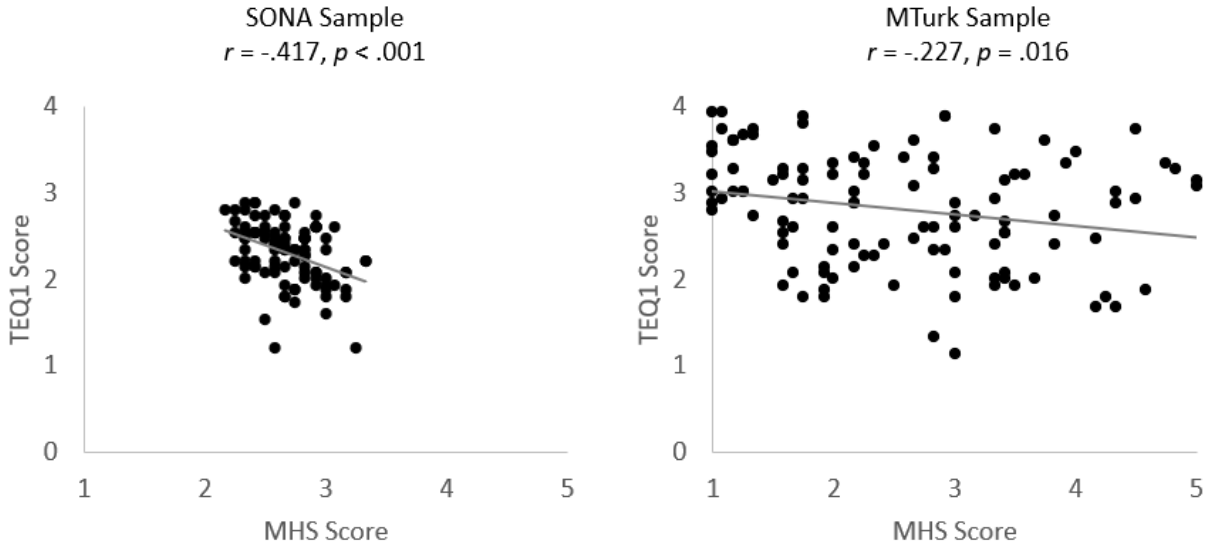

Figure 4. Correlations between TEQ1 scores (trait empathy measured at mass testing) and MHS scores (trait sexual prejudice). Left: SONA participants. Right: MTurk participants.

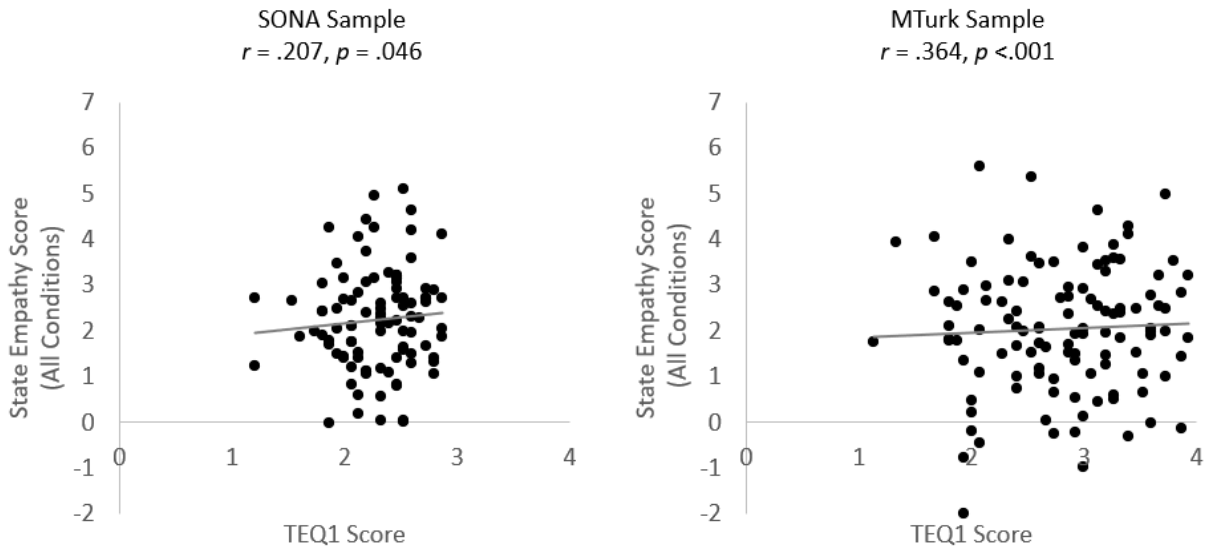

Figure 5. Significant correlations between TEQ1 scores (trait empathy) and mean state empathy difference scores averaged across all conditions. Left: SONA participants. Right: MTurk participants.

3.4 TEQ Analysis

Since the stability of the TEQ has not yet been evaluated within the published psychometric literature on this scale, a series of *t*-tests and Pearson correlations were conducted. In line with Prediction 7, these analyses sought to replicate previously observed sex differences in trait empathy. We also sought to determine whether participants' TEQ scores were different based on whether the scale was administered during Prerequisite Study/Mass Testing (TEQ1) or after the main study procedure (second iteration or TEQ2) in accordance with Predictions 8 and 9. Furthermore, the correlational analyses involving TEQ1 reported above were replicated using TEQ2 to test Predictions 10 and 11 and determine whether different iterations of the same scale had the same relationships with state empathy and sexual prejudice. Note that since some participants entered the same response for every item on the TEQ2, these participants (1 in SONA sample, 3 in MTurk sample) were removed from the following analyses involving TEQ2 scores. All analyses were conducted in SPSS version 28.0.1.0 with $\alpha = .05$ and Bayes factors

were obtained for any predicted but non-significant *t*-test or correlation using JASP version 0.16.3.

3.4.1 SONA Analysis

Within the SONA sample, TEQ1 scores were significantly greater for female compared to male participants, $t(91) = 4.701$, $p < .001$, $d = .996$ (Prediction 7). However, the same difference was not observed for TEQ2 scores, $t(92) = 0.568$, $p = .286$, $d = .119$ (Figure 6). The Bayes factor obtained for this test was 0.350, indicating anecdotal evidence in favour of the null hypothesis. In contrast to Prediction 8, there was no significant correlation between TEQ1 and TEQ2 scores, $r = .163$, $p = .118$ (Figure 7). The Bayes factor for this correlation was 0.510, indicating anecdotal evidence for the null hypothesis. However, and as expected, a paired-sample *t*-test revealed that there was no significant difference between TEQ1 and TEQ2 scores, $t(92) = 0.185$, $p = .854$, $d = .019$ (Prediction 9).

TEQ2 scores were not significantly correlated with state empathy scores averaged across conditions, $r = .173$, $p = .095$ (Prediction 10; Figure 8). The Bayes factor for this correlation was 0.133, indicating moderate evidence in favour of the null hypothesis. Finally, TEQ2 scores were not significantly correlated with scores on the MHS, $r = .151$, $p = .147$ (Prediction 11; Figure 9). The Bayes factor for this correlation was 0.207, indicating moderate support for the null hypothesis.

3.4.2 MTurk Participants

An independent-samples *t*-test demonstrated that female participants exhibited significantly higher TEQ1 scores than male participants, $t(111) = 3.259$, $p < .001$, $d = .613$ (supporting Prediction 7). This difference was also observed for TEQ2, albeit weaker, $t(110) = 1.770$, $p = .040$, $d = .335$ (Figure 6). TEQ1 scores were not positively correlated with TEQ2

scores, r = .158, p = .096 (contrary to Prediction 8; Figure 7); the Bayes factor for this test was 0.466 and indicated anecdotal support for the null hypothesis. Unexpectedly, a dependent-samples t-test revealed that TEQ1 scores were significantly greater than TEQ2 scores, $t(111)$ = 6.941, p < .001, d = .656 (contrary to Prediction 9), as seen on Figure 5.

TEQ2 scores were not significantly correlated with state empathy scores across conditions, r = -.038, p = .690 (contrary to Prediction 10; Figure 8). The Bayes factor for this correlation was 0.119, indicating moderate support for the null hypothesis. Furthermore, there was no significant correlation between TEQ2 scores and scores on the MHS, r = .031, p = .743 (contrary to Prediction 11; Figure 9). The Bayes factor for this correlation was 0.125, indicating moderate evidence for the null hypothesis.

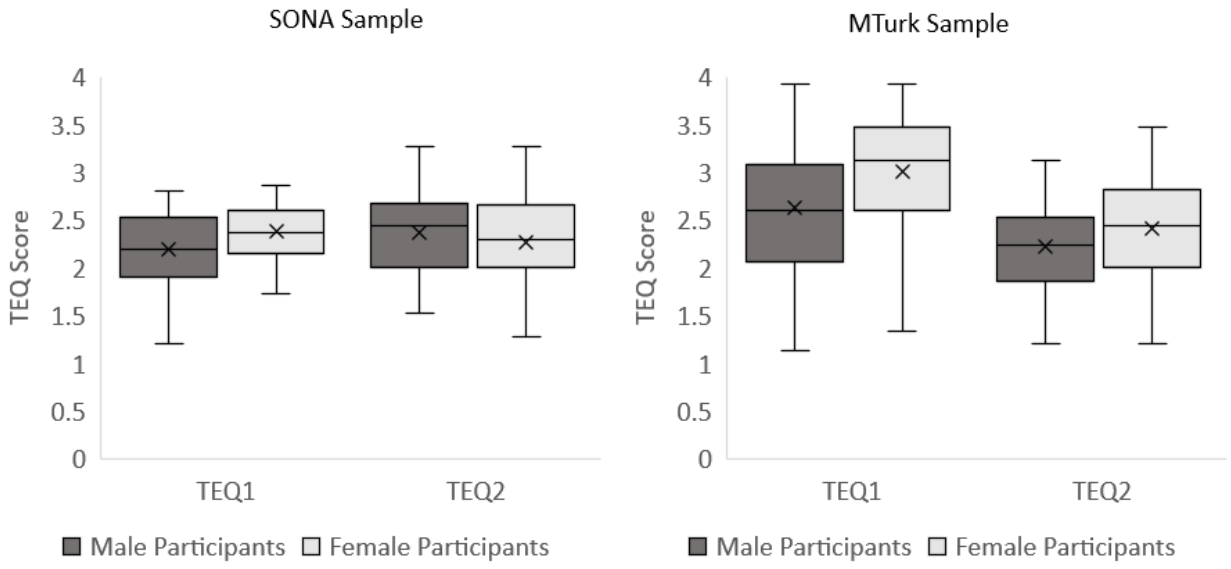

Figure 6. Male and female participants' scores on the first and second iterations of the TEQ. Left: SONA participants. Right: MTurk participants.

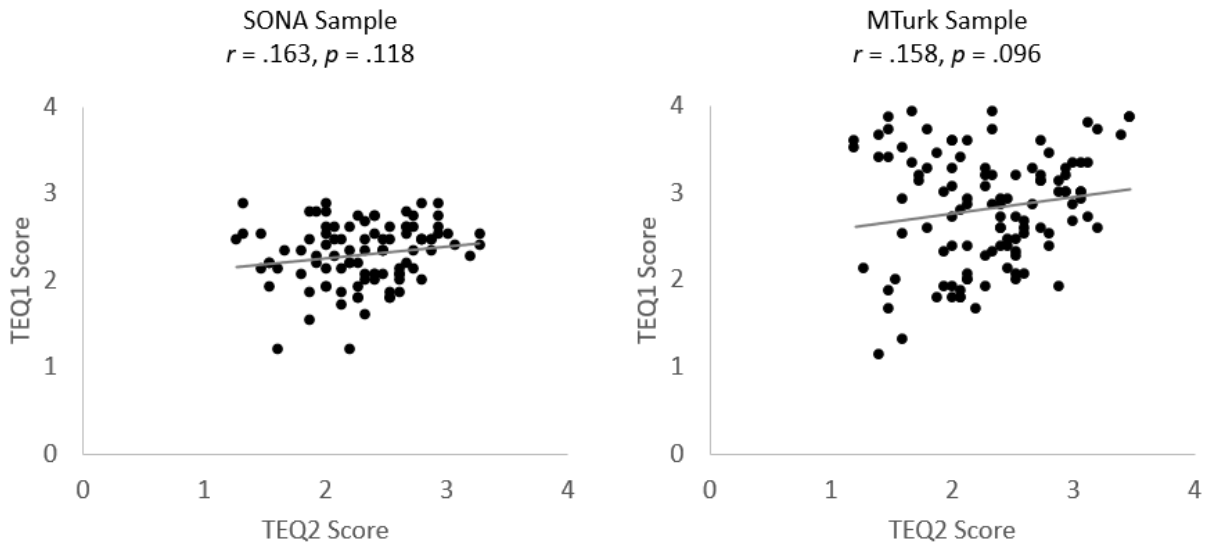

Figure 7. Correlations between TEQ1 scores (trait empathy measured at mass testing) and TEQ2 scores (trait empathy measured after the state experimental empathy manipulation). Left: SONA participants. Right: MTurk Participants.

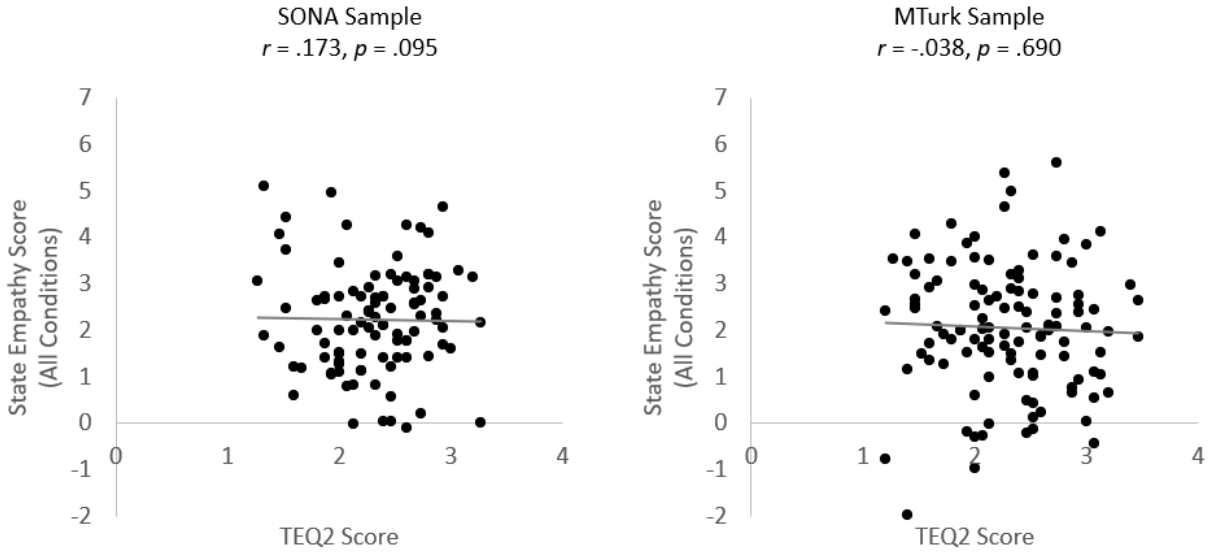

Figure 8. Correlations between TEQ2 scores (trait empathy after the main experiment) and mean state empathy scores averaged across conditions. Left: SONA participants. Right: MTurk participants.

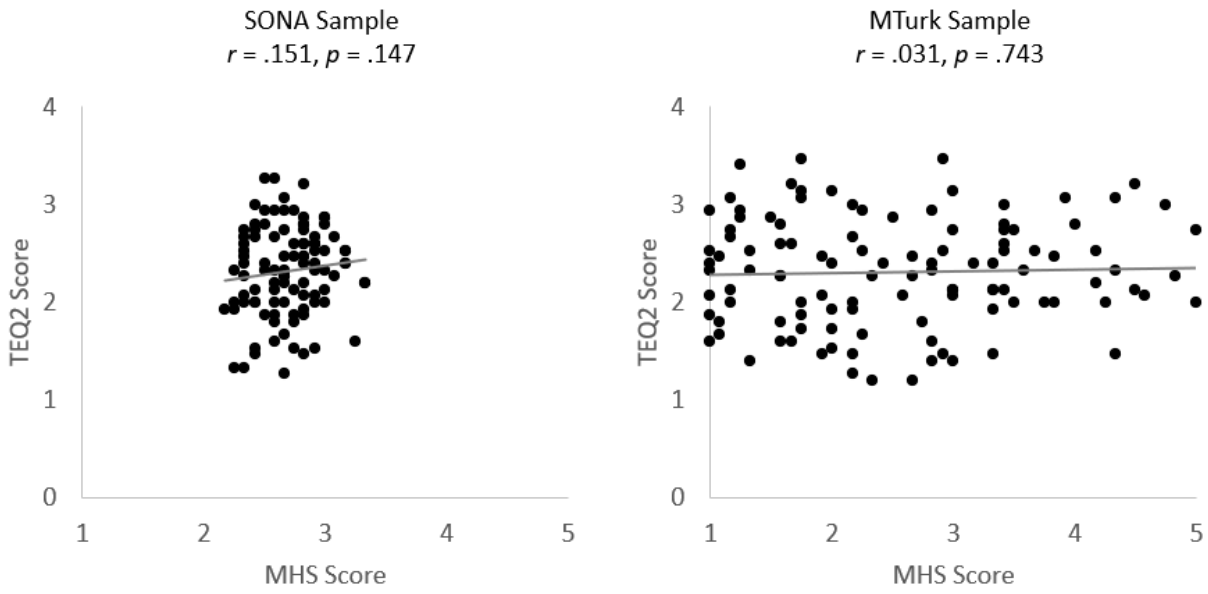

Figure 9. Correlations between TEQ2 scores (trait empathy measured after the experiment) and MHS scores (trait sexual prejudice). Left: SONA participants. Right: MTurk participants.

3.5 Post-Hoc Temporal Analysis of Empathy Ratings

While analyzing the results above, we became concerned that participants' empathy rating scores may have decreased over the course of the procedure. Participants in the SONA and MTurk samples respectively took a median time to complete the study of approximately 70 and 61 minutes. Considering the length of the study and the extremely high exclusion rate in both samples (65.94% in SONA sample, 46.45% in MTurk sample), it seemed likely that participants became fatigued over the course of the study and therefore experienced weaker affective and empathic responses to the vignettes as time went on. Alternatively, the sheer number of trials in the present study may have generated a habituation effect by which participants became increasingly desensitized to the emotional content of the vignettes.

To determine whether this fatigue-based explanation seemed plausible, we first calculated average empathy ratings and state empathy scores for the first three trials (first quarter out of the total 12 trials) and the last three trials (fourth quarter) of each condition. First, we conducted a series of paired-samples *t*-tests comparing participants' average empathy ratings

during the first three trials of each condition with their average empathy ratings in the last three trials of each condition. These tests revealed an inconsistent pattern of results displayed in Tables 4 (SONA) and 5 (MTurk) below: empathy ratings in some conditions decreased over time, while others did not change and some even increased. Additional exploratory analysis was performed by conducting the empathy ratings analysis and relevant correlational analyses on the first quarter of participants' responses. Results reflected the inconsistent changes displayed below; see Appendix B for details.

Condition	**Q1 Mean**	**Q4 Mean**	***t***	***p***	***d***
Negative Male Gay/Lesbian	7.262	6.942	2.63	.010	0.27
Negative Male Straight	7.192	7.047	1.30	.196	0.13
Negative Female Gay/Lesbian	7.094	7.148	-0.34	.738	-0.04
Negative Female Straight	7.298	7.127	1.31	.195	0.14
Neutral Male Gay/Lesbian	5.103	4.784	2.22	.029	0.23
Neutral Male Straight	5.046	4.695	2.82	.006	0.29
Neutral Female Gay/Lesbian	5.075	4.755	2.35	.021	0.24
Neutral Female Straight	5.137	6.029	-6.31	<.001	-0.65
Difference Male Gay/Lesbian	2.160	2.138	0.11	.910	0.01
Difference Male Straight	2.138	2.333	-1.21	.228	-0.13
Difference Female Gay/Lesbian	2.012	2.367	-2.00	.049	-0.21
Difference Female Straight	2.164	1.057	5.82	<.001	-0.60

Table 3. Comparison of the average empathy ratings for the first (Q1) and last (Q4) three trials in every condition within the SONA sample.

Condition	**Q1 Mean**	**Q4 Mean**	***t***	***p***	***d***
Negative Male Gay/Lesbian	6.886	7.167	-2.46	.015	-0.23
Negative Male Straight	7.197	7.038	1.26	.210	0.12
Negative Female Gay/Lesbian	7.121	7.096	0.21	.834	0.02
Negative Female Straight	7.027	6.988	0.28	.783	0.03
Neutral Male Gay/Lesbian	5.038	4.863	1.36	.176	0.13
Neutral Male Straight	5.130	4.832	2.25	.026	0.21
Neutral Female Gay/Lesbian	5.159	4.920	1.85	.067	0.17
Neutral Female Straight	5.100	6.044	-7.60	<.001	-0.72
Difference Male Gay/Lesbian	1.841	2.304	-2.91	.004	1.69
Difference Male Straight	2.045	2.198	-0.90	3.68	1.80
Difference Female Gay/Lesbian	1.948	2.175	-1.45	.150	1.81
Difference Female Straight	1.915	0.937	5.59	<.001	1.86

Table 4. Comparison of the average empathy ratings for the first (Q1) and last (Q4) three trials in every condition within the MTurk sample.

4. DISCUSSION

4.1 Results Summary

The purpose of the present study was to determine whether the apparent relationship between trait empathy and sexual prejudice can be generalized to state empathy and investigate whether previously observed gender effects on sexual prejudice modulate such a state-level relationship. To measure state empathy, we used participants' self-reported empathy ratings in response to empathy-inducing text vignettes; trait empathy was measured with the TEQ questionnaire. Our empathy ratings analysis revealed that our valence manipulation was effective; participants reported significantly more empathy for negative vignettes compared to neutral vignettes (supporting Prediction 1). This finding is congruent with previous studies using valenced vignettes in the broader empathy literature and supports the notion that vignette-based paradigms can be used to effectively manipulate participants' state empathy (Arditte Hall et al., 2018; Gehenne et al., 2021; Hein et al., 2018; McCrackin & Itier, 2021; McGrath & Haslam, 2020; Vera Cruz & Mullet, 2019). However, this analysis also suggests that the trait-level relationship between empathy and sexual prejudice identified by previous studies (Burke et al., 2015; Marsden & Barnett, 2020; Poteat et al., 2013) may not generalize to state empathy. Indeed, there were no significant differences in empathy ratings based on target sexual orientation within the negative vignettes, contrary to Prediction 2. Furthermore, although existing literature on gender effects in sexual prejudice suggested that participants might display even less empathy towards gay/lesbian targets of the same gender as the participant (Ahrold & Meston, 2008; Herek, 2002; Herek & McLemore, 2013; Herek, 2000; Herek & Gonzalez-Rivera, 2006; Kite, 1996; Mahaffey et al., 2005; Monto & Supinski, 2014), no such effect was observed with the present state empathy data, contrary to Prediction 3.

Previous research using the IRI indicates that there is a negative correlation between trait empathy and measures of both (trait) racial and sexual prejudice such that participants who are more empathetic tend to demonstrate less prejudice toward others and vice versa (Bäckström & Björklund, 2007; Burke et al., 2015; Marsden & Barnett, 2020; Pettigrew & Tropp, 2008; Poteat et al., 2013; Sidanius et al., 2013; Stathi et al., 2021). This relationship did not generalize to our measure of state empathy; participants who scored highly on the MHS were not significantly more likely to express less empathy towards gay/lesbian targets (contrary to Prediction 4). However, the present study revealed that the trait-level relationship is still observed when using the TEQ as the measure of trait empathy and thus more likely to be a legitimate phenomenon (supporting Prediction 5). Some previous literature suggests that the TEQ is more strongly related to behavioural measures of empathy than other trait empathy scales (De Lima & Osório, 2021; Spreng et al., 2009). The present study demonstrated a significant and positive correlation between TEQ1 scores and average state empathy scores (supporting Prediction 6), thus providing additional support for the link between TEQ and state empathy when these measures are administered separately.

The present findings also support the consensus that women tend to exhibit greater trait empathy than men (Prediction 7), which is well attested in existing studies using scales like the IRI, Empathy Quotient (EQ), Balanced Emotional Empathy Scale (BEES), as well as various behavioural measures (Derntl et al., 2009; Di Tella et al., 2020; Groen et al., 2013; Kobach & Weaver, 2012; Longobardi et al., 2019; Sonnby-Borgström et al., 2008; Stępień-Nycz et al., 2021; Toussaint & Webb, 2005). However, only the MTurk sample displayed a significant gender difference during both iterations of the TEQ, and it is unclear why the gender effect was not found in the SONA sample for TEQ2. Interestingly, the state empathy analyses revealed that

women's state empathy scores were higher than those of men and this was found in both samples (Figure 2). Although not predicted, this finding supports the idea that women exhibit more state empathy than men, complementing the gender effect on trait empathy.

Although previous studies evaluating the TEQ have assumed it is stable across time (De Lima & Osório, 2021; Kourmousi et al., 2017; Spreng et al., 2009; Totan et al., 2012; Xu et al., 2020), the present study revealed that scores on the first and second iterations of the TEQ were not significantly correlated in either sample, suggesting low test-retest reliability for this scale (contrary to Prediction 8). TEQ2 scores were significantly lower than TEQ1 scores within the MTurk sample (contrary to Prediction 9) but not within the SONA sample, although the implications of this difference across samples are unclear. TEQ2 scores were not significantly correlated with state empathy scores, contrary to Prediction 10. Because TEQ1 scores were correlated with average state empathy scores this finding supports the notion that the TEQ is not stable over time. TEQ2 scores were also not significantly correlated with MHS scores (contrary to Prediction 11), which further supports the instability of the TEQ as these correlations were significant for the first iteration of the TEQ in both samples.

4.2 Fatigue and Habituation

The first potential explanation for the non-significant novel findings discussed above involves the related confounds of fatigue and habituation. The procedure of the present study was lengthy and complex compared to existing studies of sexual prejudice and trait empathy, which simply asked participants to complete a series of questionnaires (Burke et al., 2015; Marsden & Barnett, 2020; Poteat et al., 2013). Indeed, research on the effects of different forms of emotional exhaustion such as burnout and compassion fatigue demonstrates that individuals who are consistently exposed to the suffering of others indeed tend to exhibit blunted affect and

empathic responses over time (Cavanagh et al., 2020; Coetzee & Laschinger, 2018; Hunt et al., 2017; Wilkinson et al., 2017). However, this research is usually conducted on healthcare professionals experiencing chronic emotional exhaustion and thus may not generalize to a relatively mild empathy-eliciting procedure. Alternatively, because identifying the sexual orientation of the target required inference on the part of the participant, it is possible that as participants became fatigued or lost interest in the study, they stopped paying attention to such details and therefore failed to notice the sexual orientation of the targets in some trials. Indeed, they may have chosen to conserve their attentional resources to better answer the memory questions rather than identify the sexual orientation of the targets, especially as they became exhausted. Although we attempted to address our concerns about fatigue and habituation by embedding regular breaks, attention checks, and positively valenced vignettes throughout the study, the influence of these effects cannot be ruled out.

To determine whether fatigue and/or habituation affected the present results, we conducted post-hoc analyses that evaluated participants' responses across the course of the study. Results of these re-analyses revealed that there were inconsistent changes in participants' empathy ratings across time; ratings in some conditions decreased across time while others stayed the same and some even increased. When the dataset was restricted to the first quarter of trials in each condition and the empathy rating analysis was performed again (see Appendix B), a few previously significant effects became insignificant and vice versa. However, none of these newly significant effects were very strong, and they did not include either of the predicted interactions in either sample. Conducting the correlational analyses on the first quarter of trials revealed no newly significant correlations; the relationships between trait scale scores and state empathy scores only became less significant. While the findings of these re-analyses demonstrate

that there were at least some changes in participants' state empathy across time, these changes do not appear systematic and had only minor impacts on the results of the original analyses. Therefore, fatigue and habituation may have contributed to the lack of support for our novel predictions, but it is unlikely that they were the main culprits.

4.3 Power

Another factor that may have affected the present findings is a lack of statistical power due to relatively small sample sizes. Previous studies on sexual prejudice on trait empathy used very large samples, all of which included more than 500 participants (Burke et al., 2015; Marsden & Barnett, 2020; Poteat et al., 2013). Although we initially intended to reach around 500 participants in total, due to our high exclusion rate and rigorous data screening process our final dataset included less than half that number across two samples. However, it is important to note that even within these smaller samples, and despite using a different trait empathy scale, we replicated some of the trait-based findings of the previous studies. Using the observed power in our repeated measure ANOVA calculated by SPSS, we found that the probability of detecting the predicted interaction between target sexual orientation and vignette valence was 29.00% in the MTurk sample and only 5.30% in the SONA sample. The probability of detecting the predicted four-way interaction was 5.70% in the MTurk sample and 5.40% in the SONA sample. To assess the degree to which our non-significant correlational analysis was also underpowered, we performed post-hoc power analyses using G*Power version 3.1.9.7. We found that the probability of detecting the predicted correlation between MHS scores and state empathy scores in the average gay/lesbian condition was 44.00% in the MTurk sample and just 14.21% in the SONA sample.

These analyses indicate that if the hypothesized state-level effects exist, they must be incredibly subtle such that they would require hundreds if not thousands more participants to be reliably detectable. Such effects would likely be so small as to be irrelevant in terms of day-to-day behaviour; indeed the effect sizes for our predicted but non-significant effects were either extremely small or nonexistent. Although our analyses were underpowered, it seems highly unlikely that such minuscule effects are meaningful. Therefore, it remains entirely possible that the null hypothesis is true. This possibility is supported by the Bayes factors calculated for each predicted but non-significant analysis, which indicate strong to extremely strong support for the null hypothesis among the predictions tested by the empathy ratings ANOVA and anecdotal to moderate support for the null hypothesis across correlational analyses. This suggests that the lack of observed power described above is most likely driven by a true lack of effects rather than low sample size. Therefore, it appears unlikely that increasing the sample size would reveal many meaningful results, especially in the case of the empathy ratings analysis. Having ruled out sample size, fatigue, and habituation as the primary causes behind the present non-significant results, we are left with the possibility that the null hypothesis may be supported and that no meaningful relationship exists between trait sexual prejudice, gender, and state empathy[2].

4.4 Potential Role of Specific Emotional Characteristics

There are two possible ways to interpret the lack of a relationship between sexual prejudice and state empathy, each of which has unique and interesting implications for the study of empathy. The first possibility is that such a relationship does not exist at all, across all possible samples. This idea seems unlikely considering that studies focusing on racial prejudice

[2] Note that additional post-hoc correlation analyses independent of any predictions also revealed no significant correlation between MHS scores and state empathy scores across all conditions in either the SONA (r = -.121, p = .904) or MTurk samples (r = -.091, p = .336), with respective Bayes factors of 0.233 and 0.337. This further supports the idea that there is no meaningful relationship between trait sexual prejudice and state-level empathy.

have consistently demonstrated effects on multiple state-level measures of empathy (Craig et al., 2012; Craig et al., 2022; Forgiarini et al., 2011; Halberstadt et al., 2022; Hutchings & Haddock, 2008; Kaseweter et al., 2012; Suleiman et al., 2018; Trawalter & Hoffman, 2015). However, it is important to recall that racial and sexual prejudice are rooted in different emotional responses from the participants; racial prejudice is characterized mostly by fear and anger while sexual prejudice is characterized primarily by disgust toward the target (Banks & Hicks, 2016; Conger et al., 2012; Crawford et al., 2014; Herek & McLemore, 2013; Herek, 2000; Inbar et al., 2009; Mahaffey et al., 2005; O'Handley et al., 2017; Olatunji, 2008; Outten et al., 2012; Shields & Harriman, 1984; Terbeck et al., 2016; Terrizzi et al., 2010). A real lack of relationship between state empathy and sexual prejudice would imply that different forms of prejudice may have different impacts on trait and state empathy, which could be linked to, among other differences, the emotional response elicited by these different types of prejudice. This would explain why prejudice based on fear and anger can modulate empathy response at the state level while prejudice based on disgust seemingly cannot, despite both types of prejudice being linked to trait empathy levels.

However, it is impossible to accurately assess the likelihood of this explanation due to limitations of the present study. For instance, the present study was designed to optimize experimental control at the expense of realism; a comparable real-world social interaction would include direct communication with the target and with far more contextual information. Since existing research on racial and sexual prejudice often includes more ecologically valid methods, it is possible that the present results may reflect differences in study designs rather than differences in underlying affective characteristics. Furthermore, the vignettes in the present study displayed distressing situations designed to elicit concern from participants, but we did not

include any direct measurement of our participants' specific emotional responses to the vignettes. The incongruence between the emotional content of sexual prejudice and the emotional state of the participants may also have contributed to the lack of significant results. We do not know what our participants were feeling *towards* the target characters (e.g. disgust) because our design was focused on ensuring they felt empathic concern *for* the target characters. Also complicating interpretation of our results is the fact that we do not know whether participants actually noticed the sexual orientation of the target characters at all. We did not ask participants to identify characters' sexual orientation as part of the memory questions because we were concerned that explicit mention of sexual orientation might produce a social desirability bias, but such a manipulation check may be worth the risk to be surer of our results. It also seems logical that differences in participants' affective states would be most relevant for the affective components of empathy such as affective sharing rather than the cognitive components such as empathic concern, which was the focus of the present study. At time of writing there is also no published research testing the possibility that different forms of prejudice (sexual or racial) may have differential impact on state-level empathy constructs according to the specific affective characteristics each prejudice elicits (disgust and fear/anger, respectively). Future studies, especially those on the differences between trait and state empathy, should therefore take care to elicit different emotional responses, include manipulation checks for all independent variables, and include measures of different empathic constructs.

4.5 Potential Role of Sample Characteristics

The second possibility is that a relationship between sexual prejudice and state empathy exists, but only within the context of samples that are higher in sexual prejudice. Therefore, the novel predictions of the present study (Predictions 2, 3, and 4) might actually be supported

within other samples drawn from different populations. For instance, research consistently demonstrates that younger individuals tend to display less sexual prejudice than older ones (Herek & McLemore, 2013; Herek & Gonzalez-Rivera, 2006; Hicks & Lee, 2006; Kite, 2011), and that sexual prejudice has a negative relationship with education, although this relationship appears to be weakening over time (Brewer, 2003; Herek & Gonzalez-Rivera, 2006; Hicks & Lee, 2006; Jelen, 2017; Kozloski, 2010). Interestingly, our SONA sample consisted of highly educated young students whose range of MHS scores was much narrower than participants from the general population recruited through MTurk, possibly suggesting an effect of education in the present study. However, scores were clustered in the middle of the distribution rather than negatively skewed as would be expected of such an effect. A participant's broader cultural background has also been found to correlate with their levels of trait sexual prejudice such that participants from highly conservative, traditionalist cultures are likely to display more sexual prejudice than individuals from more pluralistic cultures (Herek & Capitanio, 1995; Kite et al., 2019; Lee & Cunningham, 2014; Licciardello et al., 2011; Rosenthal et al., 2011).

Social desirability is often used as a covariate in studies of sexual prejudice and few studies suggest that trait social desirability is negatively, albeit weakly associated with sexual and anti-trans prejudice outside of highly prejudiced samples (Aosved et al., 2009; Gerhardstein & Anderson, 2010; Ratcliff et al., 2006; Tsang & Rowatt, 2007), but there are many more studies that find no such relationship (Ahuja et al., 2019; Brassel & Anderson, 2019; Callahan & Zukowski, 2019; Chonody, 2013; Glotfelter & Anderson, 2017; Kranz, 2022; McCullough et al., 2019; Rattazzi & Volpato, 2003; Rye et al., 2019; Smith et al., 2011; White & Jenkins, 2017). While social desirability may have affected the present results, we did not measure this trait in the present study and in any case the general consensus of existing literature does not support

this possibility. Although there are many other factors that contribute to sexual prejudice such as religiosity and conservative sociopolitical ideology (Andersen & Fetner, 2008; Brewer, 2003; Herek & McLemore, 2013; Herek, 2000; Hicks & Lee, 2006; Jelen, 2017), the present study did not measure such factors and therefore cannot speak to their prevalence within the present samples.

Regardless, this research would suggest that both MTurk and SONA samples, which were composed of 18-to-30-year-old individuals and (in the SONA sample) relatively highly educated students, were on average low in sexual prejudice (see Table 2), although participants in the MTurk sample displayed much more variability than those in the SONA sample. Some of the participants in the present study were also LGBTQ+ themselves, especially within the SONA sample (see Table 1), making them even less likely to display meaningful sexual prejudice. Although the present study included no direct measures of participants' cultural backgrounds, the demographic questionnaire revealed differences in the ethnic makeup of the two samples that may reflect cultural differences, even though ethnicity itself is not a reliable predictor of sexual prejudice (Miller et al., 2012; Miller, 2021; Wuestenenk et al., 2022). Such differences between samples may also help explain the differences in response variability, significance, and effect size observed across samples in the present results. However, definitive conclusions about the influence of culture on the present results are impossible without direct measures of cultural background. Since negative attitudes towards LGBTQ+ people have been on decline among much of general (Western) population for many years (Andersen & Fetner, 2008; Brewer, 2003; Hicks & Lee, 2006; Jelen, 2017; Kite, 2011; Scott, 1998), it is possible that meaningful effects of sexual prejudice at the state level may now only be observable within highly socially conservative samples, including cultural groups in which homosexuality is simply not accepted.

Therefore, the predictions of the current study might yet hold true if one were to run this study with a sample of older, less educated, and strongly conservative participants, especially those with highly intolerant cultural backgrounds.

If state empathy effects could be found within samples with greater levels of sexual prejudice, it would suggest that sexual prejudice must be high on average to modulate state empathy. Because existing research, including the present study, clearly demonstrates a relationship between sexual prejudice and trait empathy outside of strongly conservative samples (Burke et al., 2015; Marsden & Barnett, 2020; Poteat et al., 2013), the threshold for state effects of sexual prejudice appear significantly higher than the threshold for trait effects of sexual prejudice. This discrepancy would suggest that although an individual's "baseline" (trait) empathy can be modulated by situational (state) factors, a trait influence such as sexual prejudice would have to be significantly stronger than other state influences to produce a state-level effect on empathy. Unfortunately, there is no empathy research to date that explicitly explores this hypothesis, which future studies should investigate. Future research should also determine whether measures of trait empathy like the TEQ and IRI truly measure participants' "baseline" levels of empathy without being affected by state factors. Interestingly, results from the TEQ stability analysis included in the present study may provide some preliminary findings that speaks to this issue.

4.6 TEQ Stability Analysis

Stability is the defining characteristic of a personality trait, and previous research on empathy indicates that individuals' baseline levels of empathy are relatively stable across time (Banissy et al., 2012; Baron-Cohen & Wheelwright, 2004; Cuff et al., 2016; Davis, 1983; Jolliffe & Farrington, 2006; Spreng et al., 2009; Vachon & Lynam, 2016). However, such research is

often limited to specific measures of empathy like the IRI; until now the stability of the TEQ has not been assessed within published literature. Existing research suggests that the TEQ performs comparably to the IRI on most psychometric measures but has superior external validity with behavioural and state measures of empathy (De Lima & Osório, 2021; Kourmousi et al., 2017; Spreng et al., 2009; Totan et al., 2012; Xu et al., 2020). It is therefore important to assess TEQ stability to determine whether it is truly psychometrically superior to the IRI as a measure of trait empathy. Our initial stability analysis revealed that scores on the second iteration of the TEQ (TEQ2, administered at the end of the experiment) were not correlated with scores on the first iteration (TEQ1, administered during the pre-study and during mass testing) across samples. Within the MTurk sample TEQ2 scores were also significantly lower than TEQ1 scores, a difference most likely resulting from factors unique to the MTurk sample and not necessarily a function of TEQ instability *per se*. Unlike TEQ1 scores, TEQ2 scores were not significantly correlated with average empathy ratings in either sample. However, after this analysis it is still unclear as to whether this reflected a psychometric problem with the scale itself or the influence of external factors.

Because participants completed TEQ2 after completing the vignettes, their ability to empathize may have been exhausted by the study procedure as discussed above, resulting in TEQ scores that were not correlated across time. This scenario would suggest that state-level influences can cross the trait-state boundary and influence trait-level measures. Even if this is not the case and the TEQ is merely an unstable measure, both explanations imply that the TEQ is at least somewhat affected by individuals' state levels of empathy, regardless of the specific mechanism that results in changing scores across time. To determine whether the instability observed in the present analyses was the result of interference from the study design or a

fundamental lack of test-retest reliability for the TEQ, we are currently conducting a follow-up study that collects over a thousand participants' TEQ scores during Mass Testing over the course of more than a year. However, based on the present study alone, it seems safe to assume that the TEQ is more sensitive to state influences on empathy than has been previously assumed. Therefore, the IRI should still be considered a more stable measure suitable for assessing trait empathy across time, but researchers should be aware that its external validity is limited compared to other trait measures (De Lima & Osório, 2021; Murphy & Lilienfeld, 2019). The TEQ is still a better tool for researchers interested in comparing trait and behavioural measures of empathy due to its high external validity with behavioural tasks, but researchers should use the TEQ with caution and carefully plan their research designs to minimize the chance of state-level interference. An effective compromise may be to simply administer both the TEQ and IRI in the same study, keeping their respective limitations in mind when interpreting results.

4.7 Conclusion

Although we did not find support for our novel predictions about the relationship between sexual prejudice, gender, and state empathy, the results of the present study still have practical and theoretical value for the study of empathy and sexual prejudice. We replicated a number of prior findings on the relationship between trait empathy and sexual prejudice, adding to a sparse literature and providing support for existing studies. Our non-significant results pose interesting questions about the trait and state facets of the broader empathy construct. Particularly interesting avenues for future research include investigating the factors that allow trait influences like prejudice to affect empathy at the state level, including emotional content and effect strength. Our supplementary TEQ stability analysis suggests that regardless of whether this measure is stable across time, it appears to be subject to state-level influences

despite supposedly providing measurement of trait empathy. This research highlights the need for additional studies on how social factors like prejudice influence the state and trait dimensions of empathy.

References

Ahrold, T. K., & Meston, C. M. (2008). Ethnic differences in sexual attitudes of U.S. college students: gender, acculturation, and religiosity factors. *Archives of Sexual Behavior, 39*(1), 190-202. https://doi.org/10.1007/s10508-008-9406-1

Ahuja, K. K., Dhillon, M., Juneja, A., Deepak, S., & Srivastava, G. (2019). Subverting heteronormativity: An intervention to foster positive attitudes toward homosexuality among Indian college students. *Journal of Homosexuality, 66*(6), 746-768. https://doi.org/10.1080/00918369.2018.1484230

Andersen, R., & Fetner, T. (2008). Cohort differences in tolerance of homosexuality: Attitudinal change in Canada and the United States, 1981-2000. *Public Opinion Quarterly, 72*(2), 311-330. https://doi.org/10.1093/poq/nfn017

Aosved, A. C., Long, P. J., & Voller, E. K. (2009). Measuring sexism, racism, sexual prejudice, ageism, classism, and religious intolerance: The Intolerant Schema Measure. *Journal of Applied Social Psychology, 39*(10), 2321-2354. https://doi.org/10.1111/j.1559-1816.2009.00528.x

Arditte Hall, K. A., Joormann, J., Siemer, M., & Timpano, K. R. (2018). The impact bias in self and others: Affective and empathic forecasting in individuals with social anxiety. *Behaviour Research and Therapy, 106*, 37-46. https://doi.org/10.1016/j.brat.2018.05.001

Bäckström, M., & Björklund, F. (2007). Structural modeling of generalized prejudice. *Journal of Individual Differences, 28*(1), 10-17. https://doi.org/10.1027/1614-0001.28.1.10

Banissy, M. J., Kanai, R., Walsh, V., & Rees, G. (2012). Inter-individual differences in empathy are reflected in human brain structure. *NeuroImage, 62*(3), 2034-2039. https://doi.org/10.1016/j.neuroimage.2012.05.081

Banks, A. J., & Hicks, H. M. (2016). Fear and implicit racism: Whites' support for voter ID laws. *Political Psychology, 37*(5), 641-658. https://doi.org/10.1111/pops.12292

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163-175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Batson, D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety, & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 3-16). Massachusetts Institute of Technology.

Brassel, S. T., & Anderson, V. N. (2019). Who thinks outside the gender Box? Feminism, gender self-esteem, and attitudes toward trans people. *Sex Roles, 82*(7-8), 447-462. https://doi.org/10.1007/s11199-019-01066-4

Brewer, P. R. (2003). The shifting foundations of public opinion about gay rights. *The Journal of Politics, 65*(4), 1208-1220. https://doi.org/10.1111/1468-2508.t01-1-00133

Burke, S. E., Dovidio, J. F., Przedworski, J. M., Hardeman, R. R., Perry, S. P., Phelan, S. M., Nelson, D. B., Burgess, D. J., Yeazel, M. W., & van Ryn, M. (2015). Do contact and empathy mitigate bias against gay and lesbian people among heterosexual first-year medical students? A report from the medical student CHANGE study. *Academic Medicine, 90*(5), 645-651. https://doi.org/10.1097/ACM.0000000000000661

Callahan, M. P., & Zukowski, K. T. (2019). Reactions to transgender women and men in public restrooms: Correlates and gender differences. *Journal of Homosexuality, 66*(1), 117-138. https://doi.org/10.1080/00918369.2017.1395661

Cavanagh, N., Cockett, G., Heinrich, C., Doig, L., Fiest, K., Guichon, J. R., Page, S., Mitchell, I., & Doig, C. J. (2020). Compassion fatigue in healthcare providers: A systematic review and meta-analysis. *Nursing Ethics, 27*(3), 639-665. https://doi.org/10.1177/0969733019889400

Chonody, J. M. (2013). Measuring sexual prejudice against gay men and lesbian women: Development of the Sexual Prejudice Scale (SPS). *Journal of Homosexuality, 60*(6), 895-926. https://doi.org/10.1080/00918369.2013.774863

Coetzee, S. K., & Laschinger, H. K. S. (2018). Toward a comprehensive, theoretical model of compassion fatigue: An integrative literature review. *Nursing & Health Sciences, 20*(1), 4-15. https://doi.org/10.1111/nhs.12387

Conger, A. J., Dygdon, J. A., & Rollock, D. (2012). Conditioned emotional responses in racial prejudice. *Ethnic and Racial Studies, 35*(2), 298-319. https://doi.org/10.1080/01419870.2011.581760

Contreras-Huerta, L. S., Hielscher, E., Sherwell, C. S., Rens, N., & Cunnington, R. (2014). Intergroup relationships do not reduce racial bias in empathic neural responses to pain. *Neuropsychologia, 64*, 263-270. https://doi.org/10.1016/j.neuropsychologia.2014.09.045

Craig, B. M., Chen, N. T. M., & Lipp, O. V. (2022). Featural vs. Holistic processing and visual sampling in the influence of social category cues on emotion recognition. *Cognition and Emotion*, https://doi.org/10.1080/02699931.2022.2057442

Craig, B. M., Mallan, K. M., & Lipp, O. V. (2012). The effect of poser race on the happy categorization advantage depends on stimulus type, set size, and presentation duration. *Emotion, 12*(6), 1303-1314. https://doi.org/10.1037/a0028622

Crawford, J. T., Inbar, Y., & Maloney, V. (2014). Disgust sensitivity selectively predicts attitudes toward groups that threaten (or uphold) traditional sexual morality. *Personality and Individual Differences, 70*, 218-223. https://doi.org/10.1016/j.paid.2014.07.001

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review, 8*(2), 144-153. https://doi.org/10.1177/1754073914558466

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. https://doi.org/10.1037/0022-3514.44.1.113

De Lima, F. F., & Osório, F. D. L. (2021). Empathy: Assessment instruments and psychometric quality – A systematic literature review with a meta-analysis of the past ten years. *Frontiers in Psychology, 12*, e781346. https://doi.org/10.3389/fpsyg.2021.781346

Decety, J. (2011). Dissecting the neural mechanisms mediating empathy. *Emotion Review, 3*(1), 92-108. https://doi.org/10.1177/1754073910374662

Decety, J., & Jackson, P. (2006). Empathy: A social neuroscience perspective. *Current Directions in Psychological Science, 15*(2), 541-548. https://doi.org/10.1111/j.0963-7214.2006.00406.x

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The Scientific World Journal, 6*, 1146-1163. https://doi.org/10.1100/tsw.2006.221

Decety, J., & Lamm, C. (2009). Empathy versus personal distress: Recent evidence from social neuroscience. In J. Decety, & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 199-214). Massachusetts Institute of Technology.

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2009). Multidimensional assessment of empathic abilities: Neural correlates and

gender differences. *Psychoneuroendocrinology, 35*(1), 67-82. https://doi.org/10.1016/j.psyneuen.2009.10.006

Di Tella, M., Miti, F., Ardito, R. B., & Adenzato, M. (2020). Social cognition and sex: Are men and women really different? *Personality and Individual Differences, 162,* e110045. https://doi.org/10.1016/j.paid.2020.110045

Fabi, S., & Leuthold, H. (2018). Racial bias in empathy: Do we process dark- and fair-colored hands in pain differently? An EEG study. *Neuropsychologia, 114*, 143-157. https://doi.org/10.1016/j.neuropsychologia.2018.04.024

Forgiarini, M., Gallucci, M., & Maravita, A. (2011). Racism and the empathy for pain on our skin. *Frontiers in Psychology, 2,* 108. https://doi.org/10.3389/fpsyg.2011.00108

Fourie, M. M., Stein, D. J., Solms, M., Gobodo-Madikizela, P., & Decety, J. (2017). Empathy and moral emotions in post-apartheid South-Africa: An fMRI investigation. *Social Cognitive and Affective Neuroscience, 12*(6), 881-892. https://doi.org/10.1093/scan/nsx019

Gehenne, L., Christophe, V., Eveno, C., Carnot, A., Turpin, A., Pannier, D., Piessen, G., & Lelorain, S. (2021). Creating scripted video-vignettes in an experimental study on two empathic processes in oncology: Reflections on our experience. *Patient Education and Counseling, 104*(3), 654-662. https://doi/org10.1016/j.pec.2020.08.041

Gerhardstein, K. R., & Anderson, V. N. (2010). There's more than meets the eye: Facial appearance and evaluations of transsexual people. *Sex Roles, 62*(5-6), 361-373. https://doi.org/10.1007/s11199-010-9746-x

Glotfelter, M. A., & Anderson, V. N. (2017). Relationships between gender self-esteem, sexual prejudice, and trans prejudice in cisgender heterosexual college students. *The International*

Journal of Transgenderism, *18*(2), 182-198. https://doi.org/10.1080/15532739.2016.1274932

Grey, J. A., Robinson, B. &. E., Coleman, E., & Bockting, W. O. (2013). A systematic review of instruments that measure attitudes toward homosexual men. *The Journal of Sex Research, 50*(3-4), 329-352. https://doi.org/10.1080/00224499.2012.746279

Groen, Y., Wijers, A. A., Tucha, O., & Althaus, M. (2013). Are there sex differences in ERPs related to processing empathy-evoking pictures? *Neuropsychologia, 51*(1), 142-155. https://doi.org/10.1016/j.neuropsychologia.2012.11.012

Halberstadt, A. G., Cooke, A. N., Garner, P. W., Hughes, S. A., Oertwig, D., & Neupert, S. D. (2022). Racialized emotion recognition accuracy and anger bias of children's faces. *Emotion (Washington, D.C.), 22*(3), 403-417. https://doi.org/10.1037/emo0000756

Han, S. (2018). Neurocognitive basis of racial ingroup bias in empathy. *Trends in Cognitive Sciences, 22*(5), 400-421. https://doi.org/10.1016/j.tics.2018.02.013

Hatfield, E., Rapson, R. L., & Yen-Chi, L. L. (2009). Emotional contagion and empathy. In J. Decety, & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 19-30). Massachusetts Institute of Technology.

Hein, S., Röder, M., & Fingerle, M. (2018). The role of emotion regulation in situational empathy-related responding and prosocial behaviour in the presence of negative affect. *International Journal of Psychology, 53*(6), 477-485. https://doi.org/10.1002/ijop.12405

Herek, G. M. (2002). Gender gaps in public opinion about lesbians and gay men. *Public Opinion Quarterly, 66*(1), 40-66. https://doi.org/10.1086/338409

Herek, G. M. (2000). The psychology of sexual prejudice. *Current Directions in Psychological Science, 9*(1), 19. https://doi.org/10.1111/1467-8721.00051

Herek, G. M., & Capitanio, J. P. (1995). Black heterosexuals' attitudes toward lesbians and gay men in the United States. *The Journal of Sex Research, 32*(2), 95-105. https://doi.org/10.1080/00224499509551780

Herek, G. M., & Gonzalez-Rivera, M. (2006). Attitudes toward homosexuality among U.S. residents of Mexican descent. *The Journal of Sex Research, 43*(2), 122-135. https://doi.org/10.1080/00224490609552307

Herek, G. M., & McLemore, K. A. (2013). Sexual prejudice. *Annual Review of Psychology, 64*(1), 309-333. https://doi.org/10.1146/annurev-psych-113011-143826

Hicks, G. R., & Lee, T. (2006). Public attitudes toward gays and lesbians. *Journal of Homosexuality, 51*(2), 57-77. https://doi.org/10.1300/J082v51n02_04

Hudson, S. T. J., Cikara, M., & Sidanius, J. (2019). Preference for hierarchy is associated with reduced empathy and increased counter-empathy towards others, especially out-group targets. *Journal of Experimental Social Psychology, 85*, e103871. https://doi.org/10.1016/j.jesp.2019.103871

Hunt, P. A., Denieffe, S., & Gooney, M. (2017). Burnout and its relationship to empathy in nursing: a review of the literature. *Journal of Research in Nursing, 22*(1-2), 7-22. https://doi.org/10.1177/1744987116678902

Hutchings, P. B., & Haddock, G. (2008). Look Black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *Journal of Experimental Social Psychology, 44*(5), 1418-1420. https://doi.org/10.1016/j.jesp.2008.05.002

Inbar, Y., Pizarro, D. A., Knobe, J., & Bloom, P. (2009). Disgust sensitivity predicts intuitive disapproval of gays. *Emotion, 9*(3), 435-439. https://doi.org/10.1037/a0015960

Jelen, T. G. (2017). Public attitudes toward abortion and LGBTQ issues: A dynamic analysis of region and partisanship. *SAGE Open, 7*(2), 1-6. https://doi.org/10.1177/2158244017697362

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence (London, England.), 29*(4), 589-611. https://doi.org/10.1016/j.adolescence.2005.08.010

Kaseweter, K. A., Drwecki, B. B., & Prkachin, K. M. (2012). Racial differences in pain treatment and empathy in a Canadian sample. *Pain Research and Management, 17*(6), 381-384. https://doi.org/10.1155/2012/803474

Kiebel, E. M., McFadden, S. L., & Herbstrith, J. C. (2017). Disgusted but not afraid: Feelings toward same-sex kissing reveal subtle homonegativity. *The Journal of Social Psychology, 157*(3), 263-278. https://doi.org/10.1080/00224545.2016.1184127

Kite, M. E., Togans, L. J., & Schultz, T. J. (2019). Stability or change? A cross-cultural look at attitudes toward sexual and gender identity minorities. In K. D. Keith (Ed.), *Cross-cultural psychology: Contemporary themes and perspectives* (pp. 427-448). Wiley-Blackwell.

Kite, M. E. (2011). (Some) things are different now. *Psychology of Women Quarterly, 35*(3), 517-522. https://doi.org/10.1177/0361684311414831

Kite, M. E. (1996). Sex differences in attitudes toward homosexual persons, behaviors, and civil rights: A meta-analysis. *Personality and Social Psychology Bulletin, 22*(4), 336-353. https://doi.org/10.1177/0146167296224002

Kline, R. B. (1998). *Principles and practice of structural equation modeling* (2nd ed.). Guilford.

Kobach, M. J., & Weaver, A. J. (2012). Gender and empathy differences in negative reactions to fictionalized and real violent images. *Communication Reports (Pullman, Wash.), 25*(2), 51-61. https://doi.org/10.1080/08934215.2012.721087

Kourmousi, N., Amanaki, E., Tzavara, C., Merakou, K., Barbouni, A., & Koutras, V. (2017). The Toronto Empathy Questionnaire: Reliability and validity in a nationwide sample of Greek teachers. *Social Sciences, 6*(2)https://doi.org/10.3390/socsci6020062

Kozloski, M. J. (2010). Homosexual moral acceptance and social tolerance: Are the effects of education changing? *Journal of Homosexuality, 57*(10), 1370-1383. https://doi.org/10.1080/00918369.2010.517083

Kranz, D. (2022). On the attribution of parental competence: Parents' behavior matters, not their sexual orientation. *Journal of Child and Family Studies*, https://doi.org/10.1007/s10826-022-02335-9

Lee, M. D., & Wagenmakers, E. (2013). *Bayesian cognitive modeling: A practical course* (1st ed.). Cambridge University Press.

Lee, W., & Cunningham, G. B. (2014). Imagine that: examining the influence of sport-related imagined contact on intergroup anxiety and sexual prejudice across cultures. *Journal of Applied Social Psychology, 44*(8), 557-566. https://doi.org/10.1111/jasp.12247

Licciardello, O., Castiglione, C., & Rampullo, A. (2011). Intergroup contact, value system and the representation of homosexuality. *Procedia - Social and Behavioral Sciences, 30*, 1467-1471. https://doi.org/10.1016/j.sbspro.2011.10.284

Longobardi, E., Spataro, P., & Rossi-Arnaud, C. (2019). Direct and indirect associations of empathy, theory of mind, and language with prosocial behavior: Gender differences in primary school children. *The Journal of Genetic Psychology, 180*(6), 266-279. https://doi.org/10.1080/00221325.2019.1653817

Mahaffey, A. L., Bryan, A., & Hutchison, K. E. (2005). Sex differences in affective responses to homoerotic stimuli: Evidence for an unconscious bias among heterosexual men, but not

heterosexual women. *Archives of Sexual Behavior, 34*(5), 537-545. https://doi.org/10.1007/s10508-005-6279-4

Marsden, A. D., & Barnett, M. D. (2020). The role of empathy in the relationship between social political ideology and sexual prejudice in heterosexual college students in the U.S. *Archives of Sexual Behavior, 49*(5), 1853-1861. https://doi.org/10.1007/s10508-019-01545-5

McCrackin, S. D., & Itier, R. J. (2021). Feeling through another's eyes: Perceived gaze direction impacts ERP and behavioural measures of positive and negative affective empathy. *NeuroImage, 226*, 117605. https://doi.org/10.1016/j.neuroimage.2020.117605

McCullough, R., Dispenza, F., Chang, C. Y., & Zeligman, M. R. (2019). Correlates and predictors of anti-transgender prejudice. *Psychology of Sexual Orientation and Gender Diversity, 6*(3), 359-368. 10.1037/sgd0000334

McGrath, M. J., & Haslam, N. (2020). Development and validation of the Harm Concept Breadth Scale: Assessing individual differences in harm inflation. *PloS One, 15*(8), e0237732. https://doi.org/10.1371/journal.pone.0237732

Meiring, L., Subramoney, S., Thomas, K. G. F., Decety, J., & Fourie, M. M. (2014). Empathy and helping: Effects of racial group membership and cognitive load. *South African Journal of Psychology, 44*(4), 426-438. https://doi.org/10.1177/0081246314530280

Miller, A. K. (2021). A dilemma of dogma: Specifying the personality root of sexual prejudice. *Journal of Homosexuality, 68*(1), 3-22. https://doi.org/10.1080/00918369.2019.1624454

Miller, A. K., Wagner, M. M., & Hunt, A. N. (2012). Parsimony in personality: Predicting sexual prejudice. *Journal of Homosexuality, 59*(2), 201-214. https://doi.org/10.1080/00918369.2012.638550

Molenberghs, P., Gapp, J., Wang, B., Louis, W. R., & Decety, J. (2016). Increased moral sensitivity for outgroup perpetrators harming ingroup members. *Cerebral Cortex, 26*(1), 225-233. https://doi.org/10.1093/cercor/bhu195

Monto, M. A., & Supinski, J. (2014). Discomfort with homosexuality: A new measure captures differences in attitudes toward gay men and lesbians. *Journal of Homosexuality, 61*(6), 899-916. https://doi.org/10.1080/00918369.2014.870816

Morrison, M. A., & Morrison, T. G. (2003). Development and validation of a scale measuring modern prejudice toward gay men and lesbian women. *Journal of Homosexuality, 43*(2), 15-37. https://doi.org/10.1300/J082v43n02_02

Morrison, M. A., Morrison, T. G., & Franklin, R. (2009). Modern and old-fashioned homonegativity among samples of Canadian and American university students. *Journal of Cross-Cultural Psychology, 40*(4), 523-542. https://doi.org/10.1177/0022022109335053

Morrison, T. G., Kenny, P., & Harrington, A. (2005). Modern prejudice toward gay men and lesbian women: Assessing the viability of a measure of modern homonegative attitudes within an Irish context. *Genetic, Social, and General Psychology Monographs, 131*(3), 219-250. https://doi.org/10.3200/MONO.131.3.219-250

Murphy, B. A., & Lilienfeld, S. O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. *Psychological Assessment, 31*(8), 1062-1072. https://doi.org/10.1037/pas0000732

O'Handley, B. M., Blair, K. L., & Hoskin, R. A. (2017). What do two men kissing and a bucket of maggots have in common? Heterosexual men's indistinguishable salivary α-amylase

responses to photos of two men kissing and disgusting images. *Psychology and Sexuality, 8*(3), 173-188. https://doi.org/10.1080/19419899.2017.1328459

Olatunji, B. O. (2008). Disgust, scrupulosity and conservative attitudes about sex: Evidence for a mediational model of homophobia. *Journal of Research in Personality, 42*(5), 1364-1369. https://doi.org/10.1016/j.jrp.2008.04.001

Outten, H. R., Schmitt, M. T., Miller, D. A., & Garcia, A. L. (2012). Feeling threatened about the future: Whites' emotional reactions to anticipated ethnic demographic changes. *Personality and Social Psychology Bulletin, 38*(1), 14-25. https://doi.org/10.1177/0146167211418531

Parrott, D. J., & Gallagher, K. E. (2008). What accounts for heterosexual women's negative emotional responses to lesbians? Examination of traditional gender role beliefs and sexual prejudice. *Sex Roles, 59*(3-4), 229-239. https://doi.org/10.1007/s11199-008-9436-0

Peterson, C. H., Dalley, L. M., Dombrowski, S. C., & Maier, C. (2017). A review of instruments that measure LGBTQ affirmation and discrimination constructs in adults. *Journal of LGBT Issues in Counseling, 11*(4), 230-246. https://doi.org/10.1080/15538605.2017.1380555

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*(6), 922-934. https://doi.org/10.1002/ejsp.504

Poteat, P. V., DiGiovanni, C. D., & Scheer, J. R. (2013). Predicting homophobic behavior among heterosexual youth: Domain general and sexual orientation-specific factors at the individual and contextual level. *Journal of Youth and Adolescence, 42*(3), 351-362. https://doi.org/10.1007/s10964-012-9813-4

Ratcliff, J. J., Lassiter, G. D., Markman, K. D., & Snyder, C. J. (2006). Gender differences in attitudes toward gay men and lesbians: The role of motivation to respond without prejudice.

Personality & Social Psychology Bulletin, 32(10), 1325-1338. https://doi.org/10.1177/0146167206290213

Rattazzi, A. M. M., & Volpato, C. (2003). Social desirability of subtle and blatant prejudice scales. *Psychological Reports, 92*(1), 241-250. https://doi.org/10.2466/pr0.2003.92.1.241

Romero, D. H., Morera, O. F., & Wiebe, J. S. (2015). Assessing the gender invariance of the Modern Homonegativity Scale. *Journal of Homosexuality, 62*(11), 1539-1559. https://doi.org/10.1080/00918369.2015.1073034

Rosenthal, L., Levy, S. R., & Moss, I. (2011). Polyculturalism and openness about criticizing one's culture: Implications for sexual prejudice. *Group Processes & Intergroup Relations, 15*(2), 149-165. https://doir.org/10.1177/1368430211412801

Rye, B. J., & Meaney, G. J. (2010). Measuring homonegativity. *Canadian Journal of Behavioural Science, 42*(3), 158-167. https://doi.org/10.1037/a0018237

Rye, B. J., Merritt, O. A., & Straatsma, D. (2019). Individual difference predictors of transgender beliefs: Expanding our conceptualization of conservatism. *Personality and Individual Differences, 149*, 179-185. https://doi.org/10.1016/j.paid.2019.05.033

Scott, J. (1998). Changing attitudes to sexual morality: A cross-national comparison. *Sociology, 32*(4), 815-845. https://doi.org/10.1177/0038038598032004010

Sessa, P., Meconi, F., Castelli, L., & Dell'Acqua, R. (2014). Taking one's time in feeling other-race pain: An event-related potential investigation on the time-course of cross-racial empathy. *Social Cognitive and Affective Neuroscience, 9*(4), 454-463. https://doi.org/10.1093/scan/nst003

Shields, S. A., & Harriman, R. E. (1984). Fear of male homosexuality. *Journal of Homosexuality, 10*(1-2), 53-67. https://doi.org/10.1300/J082v10n01_04

Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Ho, A. K., Sibley, C., & Duriez, B. (2013). You're inferior and not worth our concern: The interface between empathy and social dominance orientation. *Journal of Personality, 81*(3), 313-323. https://doi.org/10.1111/jopy.12008

Smith, S. J., Zanotti, D. C., Axelton, A. M., & Saucier, D. A. (2011). Individuals' beliefs about the etiology of same-sex sexual orientation. *Journal of Homosexuality, 58*(8), 1110-1131. https://doi.org/10.1080/00918369.2011.598417

Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2008). Gender differences in facial imitation and verbally reported emotional contagion from spontaneous to emotionally regulated processing levels. *Scandinavian Journal of Psychology, 49*(2), 111-122. https://doi.org/10.1111/j.1467-9450.2008.00626.x

Spreng, R. N., Mckinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment, 91*(1), 62-71. https://doi.org/10.1080/00223890802484381

Stathi, S., Humayun, S., Stoddart Isaac, R., & McCarron, D. M. (2021). Psychopathy and prejudice: The mediating role of empathy, social dominance orientation and right-wing authoritarianism. *Journal of Theoretical Social Psychology, 5*(4), 530-541. https://doi.org/10.1002/jts5.116

Stępień-Nycz, M., Bosacki, S., & Białecka-Pikul, M. (2021). Theory of mind in adolescence: Developmental change and gender differences. *The Journal of Early Adolescence, 41*(3), 424-436. https://doi.org/10.1177/0272431620919164

Suleiman, R., Yahya, R., Decety, J., & Shamay-Tsoory, S. (2018). The impact of implicitly and explicitly primed ingroup–outgroup categorization on the evaluation of others pain: The case of the Jewish–Arab conflict. *Motivation and Emotion, 42*(3), 438-445. https://doi.org/10.1007/s11031-018-9677-3

Terbeck, S., Savulescu, J., Chesterman, L. P., & Cowen, P. J. (2016). Noradrenaline effects on social behaviour, intergroup relations, and moral decisions. *Neuroscience and Biobehavioral Reviews, 66*, 54-60. https://doi.org/10.1016/j.neubiorev.2016.03.031

Terrizzi, J. A., Shook, N. J., & Ventis, W. L. (2010). Disgust: A predictor of social conservatism and prejudicial attitudes toward homosexuals. *Personality and Individual Differences, 49*(6), 587-592. https://doi.org/10.1016/j.paid.2010.05.024

Totan, T., Dogan, T., & Sapmaz, F. (2012). The Toronto Empathy Questionnaire: Evaluation of psychometric properties among Turkish university students. *Eurasian Journal of Educational Research, 12*(46), 179-199. https://files.eric.ed.gov/fulltext/EJ1057315.pdf

Toussaint, L., & Webb, J. R. (2005). Gender differences in the relationship between empathy and forgiveness. *The Journal of Social Psychology, 145*(6), 673-685. https://doi.org/10.3200/SOCP.145.6.673-686

Trawalter, S., & Hoffman, K. M. (2015). Got pain? Racial bias in perceptions of pain. *Social and Personality Psychology Compass, 9*(3), 146-157. https://doi.org/10.1111/spc3.12161

Tsang, J., & Rowatt, W. C. (2007). The relationship between religious orientation, right-wing authoritarianism, and implicit sexual prejudice. *The International Journal for the Psychology of Religion, 17*(2), 99-120. https://doi.org/10.1080/10508610701244122

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment, 23*(2), 135-149. https://doi.org/10.1177/1073191114567941

van Baaren, R. B., Decety, J., Dijksterhuis, A., van der Leij, A., & van Leeuwen, M. L. (2009). Being imitated: Consequences of nonconsciously showing empathy. In J. Decety, & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 31-42). Massachusetts Institute of Technology.

Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology, 11*, 59-63. https://doi.org/10.1016/j.copsyc.2016.06.007

Vera Cruz, G., & Mullet, E. (2019). Empathy and Forgiveness Among Mozambican Wives. *International Perspectives in Psychology : Research, Practice, Consultation, 8*(3), 124-143. https://doi.org/10.1037/ipp0000110

Watt, D. (2007). Toward a neuroscience of empathy: Integrating affective and cognitive perspectives. *Neuro-Psychoanalysis, 9*(2), 119-140. https://doi.org/10.1080/15294145.2007.10773550

White, C. R., & Jenkins, D. D. (2017). College students' acceptance of trans women and trans men in gendered spaces: The role of physical appearance. *Journal of Gay & Lesbian Social Services, 29*(1), 41-67. https://doi.org/10.1080/10538720.2016.1261749

Wilkinson, H., Whittington, R., Perry, L., & Eames, C. (2017). Examining the relationship between burnout and empathy in healthcare professionals: A systematic review. *Burnout Research, 6*, 18-29. https://doi.org/10.1016/j.burn.2017.06.003

Wuestenenk, N., van Tubergen, F., & Stark, T. H. (2022). Attitudes towards homosexuality among ethnic majority and minority adolescents in Western Europe: The role of ethnic

classroom composition. *International Journal of Intercultural Relations*, *88*, 133-147. https://doi.org/10.1016/j.ijintrel.2022.04.001

Xu, R. H., Wong, E. L., Lu, S. Y., Zhou, L., Chang, J., & Wang, D. (2020). Validation of the Toronto Empathy Questionnaire (TEQ) among medical students in China: Analyses using three psychometric methods. *Frontiers in Psychology*, *11*, 810. https://doi.org/10.3389/fpsyg.2020.00810

Appendices

Appendix A

The general form of the 25 vignettes used in the present study are listed below. Contextual information, denoted with curved brackets, was randomly chosen from the appropriate lists for each vignette presentation. Lists of potential names, ages, locations, and hobbies are displayed in Table 8 below. Gendered pronouns, denoted with square brackets, were chosen based on the condition a given trial is in. The list of possible conditions is displayed below. Emotional valence, denoted with curly brackets, was also chosen based on condition. For clarity, examples of vignettes in each condition are provided. An example positive vignette, which was presented but not analyzed in the present study, is also provided.

1. (CHARACTER NAME) is a (AGE) year old living in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). Outside of work, (CHARACTER NAME) enjoys (HOBBY) and taking care of [HIS/HER] pets. (CHARACTER NAME)'s pet dog was {FED/KILLED} yesterday afternoon.
2. (CHARACTER NAME) is (AGE) years old and lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). [HIS/HER] hobbies include (HOBBY) and taking care of [HIS/HER] pets. (CHARACTER NAME)'s pet cat was {FED/LOST} yesterday afternoon.
3. (CHARACTER NAME) is (AGE) years old and lives in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). Although (CHARACTER NAME) works full-time, [HE/SHE] always tries to make time for [HIS/HER] passion for (HOBBY). (CHARACTER NAME) {DOES/HATES} [HIS/HER] job and {WORKS WITH/HATES} [HIS/HER] boss.
4. In the city of (LOCATION), a (AGE) year old named (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) likes to split [HIS/HER] free

time between (HOBBY) and doing amateur research online. (CHARACTER NAME) recently {LEARNED ALL ABOUT/LEARNED [HE/SHE] HAS CONTRACTED} a deadly disease.

5. At (AGE) years old, (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). Although (PARTNER NAME) spends a lot of time shopping, (CHARACTER NAME) prefers (HOBBY). (PARTNER NAME) {WENT SHOPPING/PASSED AWAY} yesterday morning.

6. (CHARACTER NAME) is a (AGE) year old living in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) divides [HIS/HER] time between [HIS/HER] passion for (HOBBY) and taking care of [HIS/HER] son. (CHARACTER NAME)'s son {WAS DELAYED BEHIND/ 'S LIFE WAS LOST AFTER} a bad car crash on the highway.

7. (CHARACTER NAME) is (AGE) years old and lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). (PARTNER NAME) and (CHARACTER NAME) spend a lot of time together, but when they are apart (CHARACTER NAME) enjoys (HOBBY). (CHARACTER NAME) {WAS DOING HOUSEWORK WITH/BROKE UP WITH} (PARTNER NAME) yesterday.

8. (CHARACTER NAME) is (AGE) years old and lives in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) divides [HIS/HER] time between their passion for (HOBBY) and taking care of [HIS/HER] child. (CHARACTER NAME)'s child was {AT [HIS/HER] WOKPLACE WITH/TAKEN AWAY FROM) [HIM/HER] today.

9. In the city of (LOCATION), a (AGE) year old named (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) loves (HOBBY) and spending time with [HIS/HER] dog. (CHARACTER NAME)'s dog {ATE ALL HER FOOD/WAS TAKEN AWAY FROM [HIM/HER]} today.

10. At (AGE) years old, (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). Recently, (CHARACTER NAME) has not been able to spend as much time (HOBBY) as [HE/SHE] usually does, as [HE/SHE] is {WORKING AT THE DRIVING TEST CENTER/PREPARING FOR [HIS/HER] DRIVING TEST}. (CHARACTER NAME) {MARKED THE/FAILED [HIS/HER]} very important driving test today.

11. (CHARACTER NAME) is a (AGE) year old living in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). Over the past week, (CHARACTER NAME) has had to prioritize {WORKING AS A TA/STUDYING FOR A PHYSICS TEST} over [HIS/HER] passion for (HOBBY). (CHARACTER NAME) {MARKED THE/FAILED [HIS/HER]} very important physics test today.

12. (CHARACTER NAME) is (AGE) years old and lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). Recently, (CHARACTER NAME) has had to divide [HIS/HER] time between (HOBBY) and {WORKING AS A TA/STUDYING FOR A PSYCH TEST}. (CHARACTER NAME) {MARKED THE/FAILED [HIS/HER]} very difficult psych exam today.

13. (CHARACTER NAME) is (AGE) years old and lives in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). Recently, (CHARACTER NAME) really enjoys (HOBBY) but spent all day yesterday [WORKING AS A TA/STUDYING FOR A MATH TEST}. (CHARACTER NAME) [MARKED THE/FAILED [HIS/HER]} very difficult math exam today.

14. In the city of (LOCATION), a (AGE) year old named (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) and (PARTNER NAME) spend a lot of time shopping together, but they also enjoy (HOBBY). (CHARACTER NAME)

knows that (PARTNER NAME) {DOES NOT LIKE SHOPPING/IS NOT REALLY IN LOVE WITH [HIM/HER]}.

15. At (AGE) years old, (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). In recent months, (CHARACTER NAME) has spent most days either looking for work or enjoying [HIS/HER] passion for (HOBBY). (CHARACTER NAME) {WENT TO/WAS REJECTED AT} a job interview today.

16. (CHARACTER NAME) is a (AGE) year old living in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) is interested in going to graduate school, but prefers to spend [HIS/HER] free time (HOBBY). (CHARACTER NAME) {READ ABOUT/WAS REJECTED BY} the school [HE/SHE] wanted to go to today.

17. (CHARACTER NAME) is (AGE) years old and lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). (CHARACTER NAME) took the weekend off to spend some time (HOBBY). Today, (PARTNER NAME) told (CHARACTER NAME) that she/he {REALLY LOVES CATS/NO LONGER LOVES [HIM/HER]}.

18. (CHARACTER NAME) is (AGE) years old and lives in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). While she/he spends a lot of time shopping, (CHARACTER NAME) prefers to spend [HIS/HER] free time (HOBBY). (CHARACTER NAME) knows that right now, (PARTNER NAME) is {SHOPPING/CHEATING}.

19. In the city of (LOCATION), a (AGE) year old named (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME). Since (CHARACTER NAME) has been on sick leave recently, [HE/SHE] has taken the opportunity to spend more time exploring [HIS/HER] passion for (HOBBY). (CHARACTER NAME) just learned that [HIS/HER] insurance {NEEDS A CODE FOR/WILL NOT PAY FOR} the medical treatment [HE/SHE] needs.

20. At (AGE) years old, (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). While (PARTNER NAME) has spent the past month working towards getting [HER/HIS] driver's license, (CHARACTER NAME) has been spending [HIS/HER] free time (HOBBY). (PARTNER NAME) has decided to {DRIVE WITH/BREAK UP WITH} (CHARACTER NAME).

21. (CHARACTER NAME) is a (AGE) year old living in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) tries [HIS/HER] best to divide [HIS/HER] time between work, taking care of [HIS/HER] cat, and (HOBBY). (CHARACTER NAME)'s cat's [FOOD WAS PURCHASED/LIFE WAS LOST} yesterday afternoon.

22. (CHARACTER NAME) is (AGE) years old and lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). (CHARACTER NAME) enjoys playing with [HIS/HER] dog and (HOBBY) every weekend. (CHARACTER NAME)'s pet dog was {FED/LOST} yesterday afternoon.

23. (CHARACTER NAME) is (AGE) years old and lives in (LOCATION) with [HIS/HER] partner, (PARTNER NAME). The couple moved to (LOCATION) recently, and (CHARACTER NAME) has taken up (HOBBY) for the first time. (CHARACTER NAME)'s close childhood friend just {PASSED THROUGH TOWN/PASSED AWAY} today.

24. In the city of (LOCATION), a (AGE) year old named (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME). (CHARACTER NAME) is currently taking care of [HIS/HER] sister, but tries to take breaks to enjoy [HIS/HER] hobby of (HOBBY) when possible. (CHARACTER NAME) {STUDIED ORGAN DONATION WITH/FOUND NO ORGAN MATCH TO SAVE} [HIS/HER] sister.

25. At (AGE) years old, (CHARACTER NAME) lives with [HIS/HER] partner, (PARTNER NAME), in (LOCATION). (CHARACTER NAME) enjoys (HOBBY), but most of [HIS/HER] day is now spent taking care of [HIS/HER] newborn baby. (CHARACTER NAME)'s newborn child was {FED/KILLED} yesterday afternoon.

Name (Male)	Name (Female)	Age	Location	Hobby
Alexander	Sarah	18	Toronto	Photography
Nathan	Emily	19	Montreal	Gardening
Simon	Emma	20	Vancouver	Playing video games
William	Jessica	21	Calgary	Reading
David	Hannah	22	Edmonton	Listening to music
Nicholas	Samantha	23	Ottawa	Travelling
Samuel	Catherine	24	Winnipeg	Playing board games
Gabriel	Rachel	25	Quebec City	Fishing
Kevin	Jade	26	Hamilton	Hunting
Charles	Amelia	27	Kitchener	Hiking
Antoine	Julia	28	London	Canoeing
Jordan	Audrey	29	Victoria	Journaling
Eric	Megan	30	Halifax	Woodworking
Vincent	Madison		Oshawa	Doing puzzles
Adam	Mary		Windsor	Cooking
Brandon	Ashley		Saskatoon	Baking
Ryan	Vanessa		St. Catharines	Swimming
Michael	Laura		Regina	Knitting
Benjamin	Stephanie		St. John's	Painting
John	Chloe		Kelowna	Birdwatching
Ethan	Rebecca		Barrie	Cycling
Matthew	Olivia		Sherbrooke	Horseback riding
Justin	Lauren		Guelph	Skateboarding
Kyle	Amanda		Kingston	Martial arts
Andrew	Sophie		Moncton	Jogging
Owen	Elizabeth		Brantford	Skiing
James	Anna		Thunder Bay	Playing sports
Joshua	Andrea		Nanaimo	Creative writing
Chris	Michelle		Sudbury	Kayaking
Luke	Alice		Lethbridge	Weightlifting

Table 5. Lists of randomized contextual information used in vignettes.

The following vignettes are examples, using randomly chosen contextual information and the appropriate gendered names, pronouns, and emotional valence option for the stated condition.

Condition 1 - Negative Male Gay: In the city of St. Catharines, a 25 year old named Justin lives with his partner, Matthew. Justin likes to split his free time between fishing and doing amateur research online. Justin recently learned that he has contracted a deadly disease.

Condition 2 - Negative Male Straight: David is 26 years old and lives with his partner, Megan, in Saskatoon. His hobbies include jogging and taking care of his pets. David's pet cat was lost yesterday afternoon.

Condition 3 - Negative Female Lesbian: Olivia is 21 years old and lives in Kingston with her partner, Lauren. Olivia divides her time between her passion for photography and taking care of her child. Olivia's child was separated from her today.

Condition 4 - Negative Female Straight: Laura is a 30 year old living in Kelowna with her partner, Brandon. Laura divides her time between her passion for knitting and taking care of her son. Laura's son's life was lost after a bad car crash on the highway.

Condition 5 - Neutral Male Gay: Ryan is 22 years old and lives in Barrie with his partner, Eric. Although Ryan works full-time, he always tries to make time for his passion for woodworking. Ryan does his job and works with his boss.

Condition 6 - Neutral Male Straight: Alexander is a 19 year old living in Montreal with his partner, Rachel. Outside of work, Alexander enjoys baking and taking care of his pets. Alexander's pet dog was fed yesterday afternoon.

Condition 7 - Neutral Female Lesbian: Chloe is 18 years old and lives with her partner, Vanessa, in Guelph. Vanessa and Chloe spend a lot of time together, but when they are apart, Chloe enjoys hiking. Chloe was doing housework with Vanessa yesterday.

Condition 8 - Neutral Female Straight: At 24 years old, Amanda lives with her partner, Adam, in Halifax. Although Adam spends a lot of time shopping, Amanda prefers gardening. Adam went shopping yesterday morning.

Positive Male Gay: At 20 years old, Owen lives with his partner, Andrew, in Calgary. Owen enjoys birdwatching, but most of his day is now spent taking care of his newborn baby. Owen's newborn child was saved yesterday afternoon.

Appendix B

To investigate potential effects of fatigue and habituation on empathy ratings further, we replicated the empathy ratings ANOVA and correlational analyses involving empathy ratings using the first quarter of trials within every condition. All details of these analyses were identical to those performed in the Empathy Ratings Analysis and Correlational Analyses subsections above, albeit on restricted datasets. Note that all variables used in these post-hoc analyses were subjected to the same data screening procedure described for the original results above. All average empathy ratings for the first quarter of trials were deemed normally distributed according to the criteria described above (Kline, 1998). The assumption of homogeneity of variance was not violated according to a series of Levene's tests. The results of these analyses are presented in Tables B.1 and B.2 below alongside the original results using the whole dataset for comparison. Note that only results that were significant or predicted are displayed below; all other results were non-significant. These results indicate that although participants' empathy ratings may have changed over time to some degree, these temporal changes did not meaningfully impact the significance of the effects that are most relevant to the present study's novel predictions (Predictions 2, 3, and 4). Therefore, it is unlikely that the lack of support for these predictions was due to fatigue and/or habituation among the participants.

Sample	Analysis	Effect	*F*	*MSE*	*p*	η^2
SONA	Original	Valence	356.237	856.702	<.001	.795
		Valence × Participant Gender	9.084	21.744	.003	.090
		Valence × Target Orientation	0.006	0.001	.940	.000
		Four-Way Interaction	0.178	.029	.674	.002
	First Quarter	Valence	310.133	807.698	<.001	.771
		Valence × Target Orientation	0.203	0.126	.653	.002
		Four-Way Interaction	0.315	0.228	.576	.003
MTurk	Original	Valence	264.571	948.161	<.001	.704
		Participant Gender	4.092	31.687	.045	.036
		Valence × Participant Gender	4.661	16.702	.033	.040
		Valence × Target Orientation	2.010	0.299	.159	.018
		Four-Way Interaction	0.061	0.012	.805	.001
	First Quarter	Valence	231.337	863.483	<.001	.676
		Target Gender × Target Orientation	6.713	4.238	.011	.057
		Valence × Target Orientation	0.902	0.468	.344	.008
		Four-Way Interaction	3.642	2.965	.059	.032

Table 6. Significant and predicted effects for original and first quarter versions of the empathy rating analysis. All other possible main effects and interactions were included in the model although they are not presented along with the focal effects and interactions here.

Sample	Analysis	Variable 1	Variable 2	*r*	*p*
SONA	Original	MHS	Average Gay/Lesbian	-.076	.468
		TEQ1	Average All	.207	.046
	First Quarter	MHS	Average Gay/Lesbian	-.010	.925
		TEQ1	Average All	.035	.737
MTurk	Original	MHS	Average Gay/Lesbian	-.140	.140
		TEQ1	Average All	.364	<.001
	First Quarter	MHS	Average Gay/Lesbian	-.072	.447
		TEQ1	Average All	.114	.229

Table 7. Significant and predicted correlations for original and first quarter versions of each correlational analysis. All other possible main effects and interactions were included in the model although they are not presented along with the focal effects and interactions here.

Appendix C

Variable	Min	Max	*M*	*SD*	Skew	Kurtosis
Negative Male Gay/Lesbian	3.10	9.00	7.06	1.28	-0.84	0.26
Negative Male Straight	3.24	8.83	7.06	1.26	-0.83	0.53
Negative Female Gay/Lesbian	3.53	9.00	7.14	1.20	-0.85	0.59
Negative Female Straight	3.93	9.00	7.16	1.08	-0.66	-0.02
Neutral Male Gay/Lesbian	1.47	8.00	4.90	1.15	-1.11	2.26
Neutral Male Straight	1.25	8.09	4.83	1.23	-0.93	1.76
Neutral Female Gay/Lesbian	1.16	8.00	4.83	1.25	-1.11	1.79
Neutral Female Straight	1.20	8.17	4.93	1.09	-1.20	3.59
Difference Male Gay/Lesbian	-1.13	5.67	2.17	1.27	0.24	0.67
Difference Male Straight	-0.29	5.25	2.22	1.25	0.14	-0.56
Difference Female Gay/Lesbian	-0.50	6.42	2.31	1.31	0.37	0.46
Difference Female Straight	-0.05	5.50	2.23	1.16	0.21	-0.10
TEQ1 Score	1.20	2.87	2.30	0.35	-0.71	0.60
MHS Score	2.17	3.33	2.70	0.27	0.14	-0.60
TEQ2 Score	1.27	3.27	2.30	0.45	-0.27	-0.43
Aggregate Gay/Lesbian	-0.29	5.67	2.24	1.23	0.32	0.39
Aggregate All	-0.11	5.09	2.23	1.15	0.18	-0.05

Table 8. Descriptive statistics for all dependent variables subjected to analysis within the SONA sample. Note that the first eight variables represent empathy ratings for the listed conditions; the subsequent four conditions represent the difference between empathy ratings in the negative and neutral versions of each condition.

Variable	Min	Max	*M*	*SD*	Skew	Kurtosis
Negative Male Gay/Lesbian	3.83	9.00	7.02	1.22	-0.54	-0.40
Negative Male Straight	3.49	9.00	7.13	1.18	-0.74	0.38
Negative Female Gay/Lesbian	3.42	9.00	7.08	1.20	-0.80	0.64
Negative Female Straight	3.46	8.92	7.09	1.16	-0.74	0.84
Neutral Male Gay/Lesbian	1.14	7.45	5.02	1.29	-1.09	1.51
Neutral Male Straight	1.17	7.67	5.00	1.37	-1.01	0.97
Neutral Female Gay/Lesbian	1.08	7.70	5.06	1.33	-1.08	1.26
Neutral Female Straight	1.00	7.58	5.06	1.39	-1.20	1.52
Difference Male Gay/Lesbian	-1.58	5.83	1.99	1.41	-0.12	0.06
Difference Male Straight	-2.55	5.92	2.12	1.47	-0.02	0.48
Difference Female Gay/Lesbian	-2.08	6.35	2.01	1.45	-0.04	0.50
Difference Female Straight	-2.83	6.27	2.01	1.54	-0.06	0.47
TEQ1 Score	1.13	3.93	2.81	0.65	-0.26	-0.72
MHS Score	0.92	5.00	2.52	1.10	0.41	-0.81
TEQ2 Score	1.20	3.47	2.31	0.55	-0.03	-0.70
Aggregate Gay/Lesbian	-1.31	5.76	2.00	1.37	-0.09	0.20
Aggregate All	-1.99	5.61	2.04	1.38	-0.10	0.21

Table 9. Descriptive statistics for all dependent variables subjected to analysis within the MTurk sample. Note that the first eight variables represent empathy ratings for the listed conditions; the

subsequent four conditions represent the difference between empathy ratings in the negative and neutral versions of each condition.