To Err is Human, to Forgive Divine:
Studying the Effect of Empathy For Race Groups on Forgiveness

Abstract

Studies show that people are more forgiving when primed with questions or social prompts that require them to put themselves into the other person's shoes. But, how do racial differences of the transgressor factor into an individual's ability to forgive and how do these differences affect policy decisions in judicial rulings, police officer training, and corporate applications for managers and employees? To answer these questions, our team developed a randomized control study, where individuals were primed with an act of transgression, then were shown an individual's face, either of the same race or of a different race as them, and then were measured on their forgiveness levels, either by first priming them with empathy or not. The observed null results suggested that we do not have enough evidence to conclude that there is a difference between people's level of forgiveness when primed with different race groups. These null effects gave us a deeper understanding of how individuals perceive in-group and out-group identities when presented with pictures of people. We conclude with possible explanations for the observed null effects for race-related priming and discuss interesting age- and gender-related effects that were observed in our analyses.

Keywords: Race, Discrimination, Gender, Forgiveness, Empathy, Group identity

By:

- Gregory Donworth, donworth@sas.upenn.edu
- Shiqing Lin, lshiqing@sas.upenn.edu
- Himani Mehta, himani22@sas.upenn.edu
- Nyaknno Owodiong-Idemeko, nidemeko@sas.upenn.edu
- Ankit Saxena, asaxena7@sas.upenn.edu
- Ladasa Tiraviriyapol, ladasa@sas.upenn.edu

Total word count: 6107

Contents

1. Introduction

Race relations in the United States have had a very long, complex history. Starting from the Native American genocide in the name of civilization and slavery in the 1700s, the inequalities in treatment amongst minorities is not unknown. Racial discrimination[1] is a prominent and important matter in American life, bearing significant harmful effects on the wellbeing of individuals in our society. Lessons from American history have shown us the damaging disparities between this "in-group love and out-group hate" (Allport, 1954). For instance, the 1882 Chinese Exclusion Act which banned the immigration of the Chinese to the U.S. and prevented Chinese residents living in the country were prevented from naturalizing as American citizens as well as the war waged on African Americans during the Jim Crow Era from the 1800s to 1950s. Research from Souring et al., posits that if a transgressor is from an in-group, the likelihood of forgiveness increases. For example, a report from the South Asian American Leaders for Tomorrow found that in the week after 9/11, the U.S. media ''reported 645 bias incidents directed towards Americans perceived to be of Middle Eastern descent.'' American civilians took it upon themselves to seek justice for the events that happened by punishing anyone wearing a turban[2], or with an Arab resemblance, even though many assaulted were not at all of Arab descent.

Racism has real effects, and minority races bear the brunt of these attacks. Research from Nellis et al. (2016), posits that the U.S. criminal justice has been cruel, as certain races have been shown to have a higher rate of incarceration than others, even when the same crimes have been committed. For instance, African Americans and Hispanics are treated more harshly than Whites, by judges and police officers, sometimes as a result of implicit biases and stereotype activation (Beckett, Nyrop, and Pfingst, 2006; Spohn, 2013). Additional studies have found that Blacks are more likely to be stopped by the police for legally invalid reasons (Tonry and Melwsky, 2008). For example, Beckett, Nyrop, and Pfingst (2006) found that, in Seattle, police target racially diverse drug markets (rather than White drug markets) because of the stereotype surrounding the racial makeup of drug traffickers. Further studies have also shown

[1] Historically, racial identity has been understood as relating to responses to racism and prejudice (Helms, 2007).
[2] Headgear also worn by Osama Bin Laden.

that Hispanics and African Americans are more likely to be denied bail than their white counterparts (e.g. Schlesinger, 2005)[3].

Additional studies from Mecleish et al. (2007), argue that people show more solidarity with others of the same identity, which suggests that people will be more likely to forgive a transgressor belonging to the same in-group. Since social identity development relies on recognizing and valuing their racial group (Paukner et al., 2015), it can be argued that as a result of identity threat strengthening the relationship between social identity and in-group favoritism, forgiveness will increase towards transgressors within an in-group. (Voci, 2006). This in-group favoritism may also lead to discrimination towards individuals outside the group, leading to unforgivenss towards members belonging to an out-group (Falk et al. 2008). We have observed this at play multiple times in American history. For example, the tragedy of Emmet Till, a 14 year-old boy who was lynched in 1955 by white adult men for allegedly whistling at a white woman. Over the past 10 years, we have observed the strength and effects of in-group favoritism with 17-year-old African-American Trayvon Martin, being shot and killed by a Hispanic man, George Zimmerman, on his way home from buying candy because he looked 'suspicious', and the many more that have followed since then. In the realm of sentencing by judges, various studies have shown that harsher sentencing results are a result of judges' stereotypes of minorities (Bishop and Frazier, 1996; Bridges et al., 1987) which can be attributed to the defendants' educational attainment and other biographical factors that judges use to help determine blameworthiness. Recent psychological studies have used hypothetical vignettes to determine these sentencing results to judges' implicit biases (Rachlinski et al., 2009).

Inspired by these events and studies, our research aims to further explore key questions on whether individuals are more forgiving towards members of their in-group and less towards members belonging to another group. Specifically, in this study, we define and measure in-group as members belonging to the same race and out-group, as members belonging to a different race. Furthermore, we explore how these racial differences of the transgressor may affect the likelihood of forgiveness of others amongst men and women. To test for empathy, we employ the same strategy used by Exline et al., 2008 by prompting these feelings through asking a series of questions.

[3] Several other studies have confirmed these findings across the U.S (e.g. Demuth, 2003; Kutateladze et al., 2014).

2. Theoretical Background

2.1 Association between empathy and forgiveness

When looking at transgression, theorists have argued that the responses victims adopt toward their offenders have ramifications not only for their cognition, but also for their emotion, physiology, and health. It is supported by several theoretical and empirical literature (e.g., Brown, 2003; Fincham et al., 2002; Exline et al.,2008) which revealed a positive association between empathy and forgiveness. One of the conditions that people are more likely to forgive is when offended parties cultivate empathic feelings toward their offenders (McCullough et al., 1997 & 1998). Affective empathy toward the offender appears to be a crucial social-cognitive determinant of forgiving, explaining considerable variability in people's self-reported forgiving of a transgressor. Many researches focused on forgiveness in close relationships found that victims are more likely to develop empathy for their transgressor when their relationship is close and satisfactory (McCullough et al., 1998). Because our study is targeting a wider population, in order to cultivate a sense of empathy, participants were asked several questions to put themselves in the transgressor's position. As auxiliaries, there are additional questions to foster the capability for similar wrongdoing and to perceived similarity to offenders.

For a bigger picture, Witvliet, Ludwig, and Vander Laan (2001) studied the immediate emotional and physiological effects that occurred when participants rehearsed hurtful memories and nursed grudges, comparing them with cases when they cultivated empathic perspective-taking and imagined granting forgiveness toward real-life offenders. Compared to research on close relationships, the research' target population becomes larger by adding more social influence. It provides an insight that, in order to reduce negative emotion and facilitate optimism, empathic perspective taking is important to study to manipulate forgiveness by reducing anger and increasing perceived control.

Empirically, scenario-based experiments are popularly used to test the effects of a victim's perspective-taking on interpersonal forgiveness, for example Takaku's study in 2001, which focused on personal capability made participants more likely to accept hypothetical apologies from offenders. One of the advantages of using scenario-based experiments, such as a vignette, is to address the problem of ethics, in which some situations cannot be experimentally constructed, manipulated or controlled, and subjects can't be randomly assigned

to experimental and control groups. Vignettes allow research participants to respond in various ways to 'short stories about hypothetical characters in specified circumstances' (Finch, 1987), and can thus be used by researchers to manipulate contextual variables experimentally. The main advantage of the vignette technique over other qualitative methods is that it effectively picks up subtleties and nuances of character that only the insider can be aware of and express (Lieberman, 1987).

We decided to use vignettes, replacing memory retrieval in our experimental design, in order to reduce recall bias, which is a systematic error caused by differences in the accuracy or completeness of the recollections retrieved by participants regarding experiences from the past. By using a vignette, we could control the severity of the transgression and eliminate outliers, for example, a person with a high forgiveness level just because the situation he/she recalled the incident as trivial.

2.2 Race in-group and out-group and forgiveness

A policy-related reason that prompted us to explore the relationship between races and level of forgiveness in depth was that certain races have been observed to have a higher rate of incarceration than others (Nellis, 2016). Understanding and studying the relationship between races and level of forgiveness would help inform policies related to the justice system. In Becker's book The Economics of Discrimination (2010), the competitive models of discrimination showed that individual-maximizing behavior may include discrimination because of their race, sex, or other noneconomic characteristics. This means that when faced with the same scenario, it is plausible that individuals may find people of certain races as more forgiving. With the theory of economics of discrimination, potential losses could be monitored and sources of errors could be traced in real situation.

However, our research is not interested in finding out which race is most forgiving. Rather, we are examining if people are more forgiving towards those belonging to their in-group (the same race as them) than that of an out-group (different race as them.) Evidence by Brown et al. (2008) suggests a positive relationship between secondhand forgiveness and the level of identification with a group, which is a form of group empathy. According to social identity theory (Tajfel and Turner, 2004), for example, people interpret events based on their group memberships and the structural relations that exist between in-groups and out-groups. When people think of themselves as group members, their desire to protect their collective identity can

lead them to appraise events in biased ways that benefit the in-group (Brown, Wohl and Exline, 2008).

Therefore, to check if there are differences in people's forgiveness levels when the transgressor in question is of the same or different race as them, we formulated our first hypothesis.

> ***Hypothesis 1:*** *People primed with same race are more likely to be forgiving compared to those primed with a different race*[4]

2.3 Gender differences between empathy and forgiveness

There are studies suggesting that the magnitude of empathy–forgiveness correlations might differ for men and women (Fincham et al., 2002; Toussaint & Webb, 2005). However, previous studies took different stands on this question. Commonly held stereotypes and some empirical researchers have found that women have higher levels of empathy than men, however, greater levels of empathy does not directly imply higher levels of forgiveness. Many studies (Berry, Worthington, Parrott, O'Connor, and Wade, 2001; Macaskill et al., 2002) failed to find any significant gender differences in the participant's forgiveness of oneself or others.

On the other hand, another recent study found that within married couples, emotional empathy was a stronger predictor of forgiveness for husbands than for wives (Fincham et al., 2002). Fincham et al. examined empathy and forgiveness in the context of marriage with 171 Italian husbands and wives, they examined the relationship separately in husbands and in wives. Though the study indicated a difference in the relationship between empathy and forgiveness across gender, there was a limitation of the study that the researchers did not run a formal statistical test to evaluate whether the difference between these correlations were statistically significant (Toussaint and Webb, 2005). Another limitation of Fincham's study was that it only examined marital offenses, whereas our study included a hypothetical offense that was relatable to most demographics, genders, races, and cultures.

Therefore, our second hypothesis is derived from a conceptual replication of the original study (Exline et al., 2008). Exline et al. (2008) found that males who are made to empathize with a transgressor tend to be more forgiving. Our study attempts to conceptually replicate that

[4] Pre-registered on AsPredicted.org

finding, while using an empathy priming technique by combining vignettes and images, which is discussed in detail in the subsequent section.

> ***Hypothesis 2:*** *Males primed with capability, similarity, and empathy questions are more likely to be forgiving than males who are not primed with capability, similarity, and empathy questions*[5]

Additionally, there were two studies showing contradictory results in the relationship between empathy and forgiveness across gender. Macaskill et al. (2002) showed a stronger relationship between empathy and forgiveness in women whereas Toussaint and Webb (2005) found an opposite pattern. Possible explanations for the observations were cultural or racial differences, since one experiment collected data from British participants and the other experiment recruited participants from the U.S. Which is why results from both our hypotheses, when coupled together, can throw light on the dynamics of gender and race on forgiveness.

[5] Pre-registered on AsPredicted.org

3. Experimental Design

3.1 The survey

To test our hypotheses we conducted a randomized control trial experiment with a between-subject design as the identification strategy, to best determine the relationship between transgressor race and forgiveness. As the key outcome of this study is the forgiveness measurement, we have used the Transgression Related Interpersonal Motivations Inventory-Revised (TRIM) to measure this outcome (McCullough, 2002). We have used capability, similarity, and empathy questions as variables that mediate forgiveness, as done in the original study that we tried to replicate. The TRIM consists of three subscales: avoidance motivations, revenge motivations, and benevolence motivations. The revenge motivations subscale (for example, I'll make him/her pay, I want to see him/her hurt and miserable, etc.) will be the dependent variable of interest. Appendix C provides a snapshot of the entire TRIM questionnaire.

The extension performed in our experiment surrounds a new independent variable, the race of the transgressor. In order to control for differences in experiences across participants, we made an adjustment from the original study, and introduced a vignette of a transgression committed to the participant, which intended to induce a feeling of revenge motivation, rather than a thinking exercise as used in the original study. Each of the independent and outcome variables are described in the following section.

3.2 Independent variables

Race

Race was manipulated by showing participants either a photo of the transgressor as an in-group (same race as the participants) or out-group (different race as the participant). The races we include are African American, Asian, Caucasian, and Hispanic. The images used to represent the transgressors were taken from a comprehensive database compiled by Ma et al. (2015). Appendix A discusses the process behind sampling of the pictures. Also, individuals who selected 'other' as their race were shown an out-group race.

Capability for similar offense

Participants were asked to rate the responses to four questions from 0 (no, definitely not) to 10 (yes, definitely). The questions included:

- Given the right circumstances, do you think that you could be capable of doing something just as bad (i.e., just as harmful or wrong) as what the other person did?
- Can you imagine a situation in which you could do something as bad as what the other person did?
- Do you think it's possible that you could ever do something as bad as what the other person did?
- Thinking back over your entire life, do you think that you have ever done anything as bad as what the other person did?

Capability for similar wrongdoing

Participants were asked to rate the responses to four questions from 0 (no, definitely not) to 10 (yes, definitely), with questions like:

- Given the right circumstances, do you think that you could be capable of doing something similar in type to what the other person did?
- Can you imagine a situation in which you could do something similar in type to what the other person did?
- Do you think it's possible that you could ever do something similar in type to what the other person did?
- Thinking back over your entire life, do you think that you have ever done anything similar in type to what the other person did?

Empathic understanding

Participants were asked to rate the responses to four questions from 0 (no, definitely not) to 10 (yes, definitely).

- To what extent can you understand why the other person acted as s/he did?
- To what extent can you see the situation from the other person's perspective?
- To what extent can you see his/her behavior as making sense?
- To what extent can you think of valid reasons why s/he acted as s/he did?

3.3 Dependent variable

Unforgiveness

We used the 18-item Transgression Related Interpersonal Motivations Inventory – Revised (TRIM) to assess level of forgiveness toward the transgressor. Participants responded to 18 items on a scale of 1 (strongly disagree) to 5 (strongly agree)[6]. Same as in the original study, we averaged the scores from the Revenge Motivation questions, which is a subset of the TRIM questions (below), to compute the forgiveness score[7]. The questions include:

- I'll make him/her pay.
- I wish that something bad would happen to him/her.
- I want him/her to get what he/she deserves.
- I'm going to get even.
- I want to see him/her hurt and miserable.

3.4 Structure of the experiment

The experiment was performed using Qualtrics software, and the participants of the survey were MTurk paid respondents. When beginning the experiment, participants first passed through a CAPTCHA security feature to ensure there was an actual participant taking the survey, and not a robot or an automated system[8].

[6] A technique for the measurement of attitudes (Likert, 1932)
[7] Pre-registered on AsPredicted.org
[8] An automated test that humans can but computers can't pass (Von Ahn, 2003)

Next, participants saw the Informed Consent form, and had the opportunity to exit the survey prior to beginning if they did not agree with the terms. If the participant continued to the next page, they were requested to confirm their agreement with the informed consent. All participants then answered a series of demographic questions, which included the participant's gender, ethnicity, age, and level of education. All participants then read a vignette of a transgression, in which they were the victim. The vignette page on Qualtrics had a timer of 12 seconds before the 'next page' button would appear. The vignette read as follows:

> "Two weeks had passed since you went out shopping for groceries due to social distancing. You had run out of many staples, including surface cleaning spray. You walked to the detergent aisle, and to your luck, you saw that there were only two bottles left. While you grabbed one, you noticed an elderly lady nearby with an empty basket strolling towards the last bottle. You left your cart to grab the last bottle of surface cleaning spray and gave it to the lady. You could see her smile through her mask as she thanked you. As you turned back to where your cart was, you saw a person, whom you recognized as your neighbor, grabbing the spray from your cart and putting it in her/his cart. When your neighbor noticed you staring in shock, she/he just smirked and walked away towards the cashier."

After reading the vignette, participants went through a comprehension and attendance check to ensure they read the vignette. If the answer was incorrect, the timer would restart and they would be able to re-read the vignette and select the correct answer (unlimited tries were offered).

All participants were then shown a photograph of the transgressor. The race of the transgressor was randomized into the in-group (same-race) or out-group (different race) treatment, based upon their race answer in the demographic question set. The gender of the transgressor was also randomized into in-group with the same gender, in-group with the different gender, out-group with the same gender, and out-group with the different gender.

All participants were then asked to complete the questions on a block of capability, similarity, and empathy counterbalanced by TRIM questions (18-item Transgression Related Interpersonal Motivations Inventory–Revised, which measures forgiveness) in random order. Participants were then required to submit their MTurk ID number. Lastly, participants had the

option of submitting feedback or comments in an open-ended question at the end of the survey. Appendix B provides the survey flow in depth.

3.5 Extension

The original study which we aimed to replicate found that forgiveness increased when men were primed by being asked to answer questions about their personal capability for committing similar wrongdoings and their ability to view the situation from the transgressor's perspective.

Our extension was interested in teasing apart how different attributes of a transgressor influences the victim's forgiveness. We specifically aimed to identify how racial similarity between victim and transgressor influence the level of forgiveness the victim has with regards to the transgression.

We designed a 2 x 2 double-blind between-subject experiment where subjects were asked to read and answer a series of questions. Figure 1 visually represents these four treatments. There are a total of four conditions described as follows:

- Condition 1: In-Group Non-Prime

  The participants will be shown an image of a perpetrator different from their own race and also will not be primed with empathy-inducing questions before we measure their forgiveness levels towards the transgressor.

- Condition 2: In-Group Prime

  The participants will be shown an image of a perpetrator similar to their own race and also will be primed with empathy-inducing questions before we measure their forgiveness levels towards the transgressor.

- Condition 3: Out-Group Non-Prime

  The participants will be shown an image of a perpetrator different from their own race and also will not be primed with empathy-inducing questions before we measure their forgiveness levels towards the transgressor.

- Condition 4: Out-Group Prime

  The participants will be shown an image of a perpetrator different from their own race and also will be primed with empathy-inducing questions before we measure their forgiveness levels towards the transgressor.

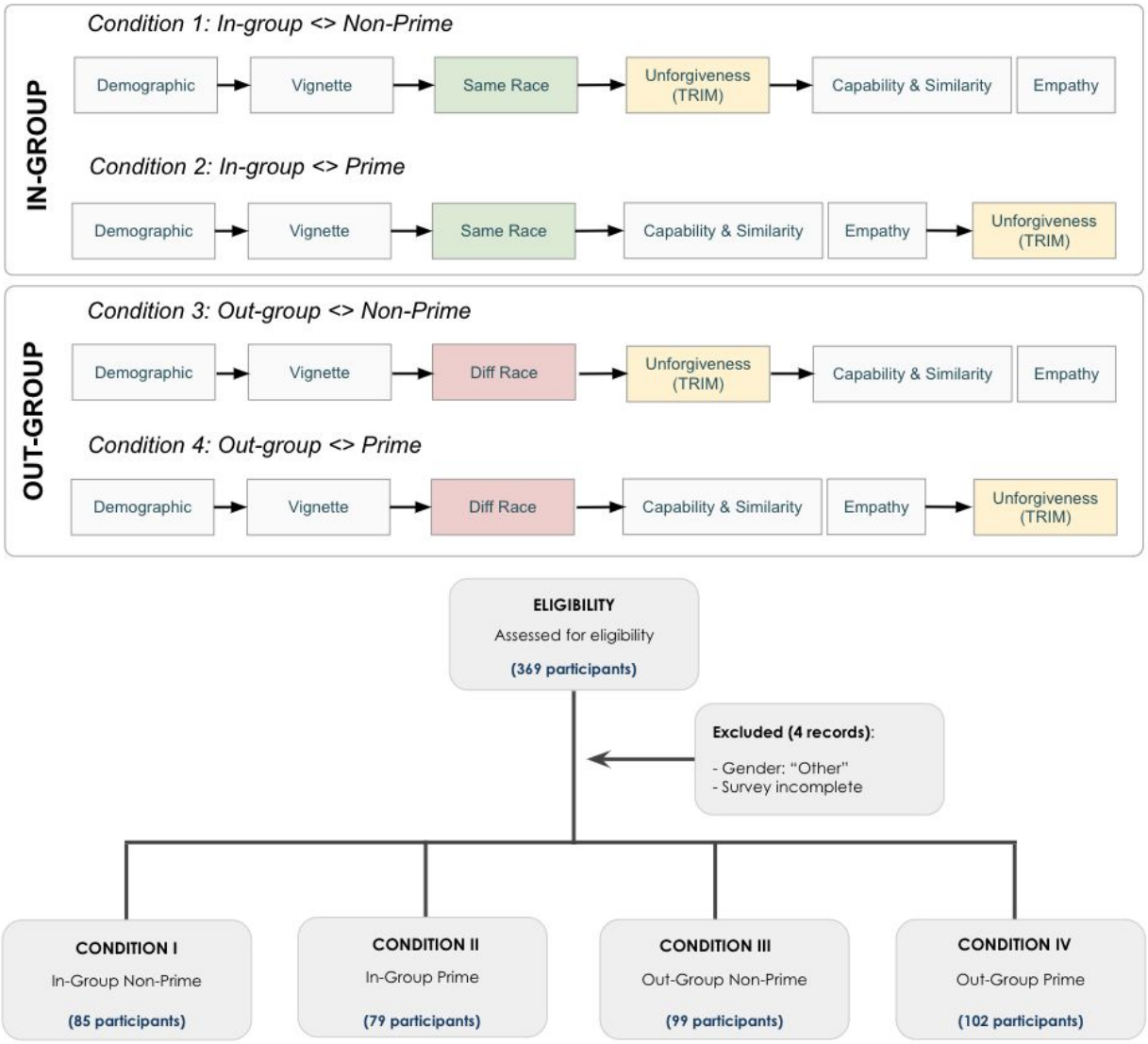


Figure 1: Different treatments in the study design and CONSORT diagram

4. Methodology

4.1 Data collection

The data were collected on Amazon Mechanical Turk (MTurk) between April 24, 2020 and April 25, 2020. A total of 369 records were collected.

4.2 Data exclusion

We excluded 4 participants who reported 'Other' as their gender, leaving us with 365 participants. Since all the participants completed their surveys, this was not used to exclude any participants[9]. Also, all the participants passed our attention check question, and hence were not excluded based on this criteria[10]. Total participants used for the analysis were 365. Outliers did not need to be filtered because our numeric values of the variables Age and Unforgiveness Scores were not many standard deviations away from the mean[11]. Table 1 shows the mean unforgiveness values across different groupings of our data and Figure 2 shows the means and confidence intervals for the four treatment groups used in our analysis.

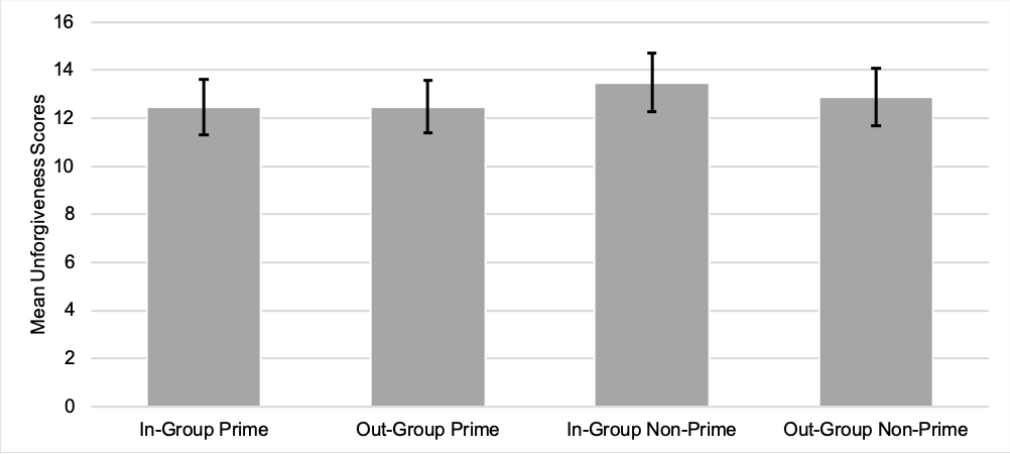


Figure 2: Means and confidence intervals for the four treatment groups

[9] Pre-registered on AsPredicted.org
[10] Pre-registered on AsPredicted.org
[11] Pre-registered on AsPredicted.org

Grouped Variable	Variable Value	Mean Unforgiveness
Gender	Female	12.0106
	Male	13.6932
Race	African	14.8846
	Asian	13.2791
	Caucasian	12.6288
	Hispanic	12.0500
	Other	12.2500
Education Level	Bachelor	13.3879
	Doctorate	12.0000
	High School	11.3809
	Master	13.1333
	Some High School	5.0000
	Trade School	14.6429
	Other	12.0000
Image Race	Asian	13.4324
	Black	12.0000
	Caucasian	13.0855
	Hispanic	12.4328
Image Gender	Female	12.5824
	Male	13.0601
Empathy Prime	Prime	12.4751
	Non-Prime	13.1630
Group	In-Group	13.0000
	Out-Group	12.6766

Table 1: Means of unforgiveness scores across different categorical groups

4.3 Data summary

Overall, we had enough records per treatment group to achieve our pre-registered power of 80 percent for pairwise comparisons[12]. Table 2 shows the summary of participants for each treatment. We also collected enough data to conduct an adequately powered hypothesis test for our first hypothesis. However, our second hypothesis test was underpowered. More details on our a priori power analyses can be found in Appendix D.

	Unforgiveness-Empathy	Empathy-Unforgiveness
In-group	85	79
Out-group	99	102

Table 2: Sample size for each treatment

4.4 Dependent variable

The dependent variable, i.e., the total unforgiveness score[13], was calculated by adding the absolute scores of the revenge-related TRIM questions (Appendix C).

4.5 Empirical strategy

We performed the Mann Whitney U test to determine differences between the samples of each group. The test was performed for each pairwise combination of the four treatments. Hence, the test was performed six times between the treatment groups[14].

Additionally, to test our first hypothesis, we performed the Mann Whitney U Test between all in-group and out-group participants[15]. Table 3 shows the sample sizes of the two groups. To test our second hypothesis, we performed the same test between males who were primed with empathy questions before answering the TRIM questions and males who were not primed with empathy questions before answering the TRIM questions. Table 4 shows the sample sizes of the two groups. Figure 3 shows the means and confidence intervals of the two

[12] Pre-registered on AsPredicted.org
[13] Pre-registered on AsPredicted.org
[14] Pre-registered on AsPredicted.org
[15] Pre-registered on AsPredicted.org

groups for the first hypothesis. Similarly, Figure 4 shows the same for the samples in the second hypothesis.

Finally, we performed a linear regression to cross-verify our results and added the control variables Age, Gender, Race, and Education, to check for other confounding effects[16].

In-group	Out-group
164	201

Table 3: Sample size for In-group and Out-group treatments

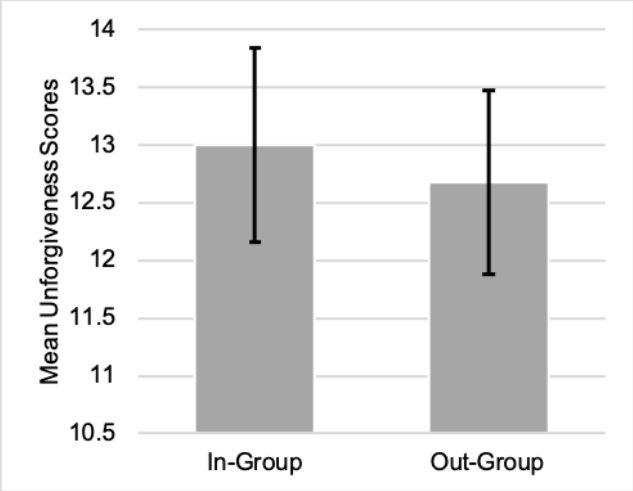


Figure 3: Means and confidence intervals for the four treatment groups[17]

Empathy Primed Males	Non-Empathy Primed Males
81	95

Table 4: Sample size for Empathy Primed Males and Non-Empathy Primed Males

[16] Pre-registered on AsPredicted.org
[17] Confidence intervals calculated at the 5 percent confidence level

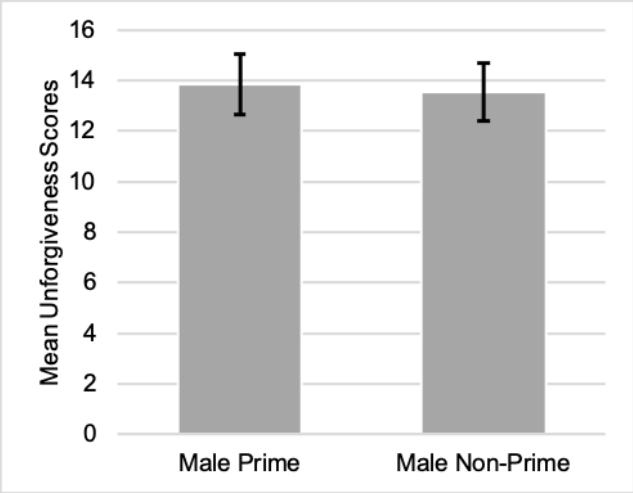


Figure 4: Means and confidence intervals for the four treatment groups[18]

4.6 Assumptions

- **Scale**: The measures are in ordinal or continuous format
- **Independence**: The independent variables were not highly correlated with each other
- **Distribution**: The observations weren't normally distributed

[18] Confidence intervals calculated at the 5 percent confidence level

5. Analysis Results

5.1 Pairwise test for differences

We obtained null effects from each of the pairwise Mann Whitney U tests performed between the treatments with unforgiveness scores of each treatment as the dependent variable with p-values greater than our pre-registered significance level of 5 percent[19] — 0.323[20] (In-Group Non-Prime and In-Group Prime), 0.495[21] (In-Group Non-Prime and Out-Group Prime), 0.176[22] (In-Group Non-Prime and Out-Group Prime), 0.649[23] (In-Group Prime and Out-Group Non-Prime), 0.750[24] (In-Group Prime and Oot-Group Prime), and 0.655[25] (Out-Group Non-Prime and In-Group Prime). Table 5 shows the U statistic[26] and p-value for the test for each pairwise combination.

	In-Group-Non-Prime	**In-Group-Prime**	**Out-Group-Non-Prime**	**Out-Group-Prime**
In-Group-Non-Prime	-	-	-	-
In-Group-Prime	U = 3057.5 p = 0.323	-	-	-
Out-Group-Non-Prime	U = 4453.0, p = 0.495	U = 3755.0 p = 0.649	-	-
Out-Group-Prime	U = 3837.5 p = 0.176	U = 4140.5 p = 0.750	U = 4864.5 p = 0.655	-

Table 5: Mann Whitney U test results for pairwise differences between treatments[27]

[19] Pre-registered on AsPredicted.org
[20] P-value associated with the Mann Whitney U test
[21] P-value associated with the Mann Whitney U test
[22] P-value associated with the Mann Whitney U test
[23] P-value associated with the Mann Whitney U test
[24] P-value associated with the Mann Whitney U test
[25] P-value associated with the Mann Whitney U test
[26] Obtained from the Mann Whitney U test
[27] All p-values not significant compared with the pre-registered significance level of 5 percent

5.2 Hypothesis test I

To perform the first hypothesis test[28], we divided the data into two groups, in-group and out-group members. We found null effects between these groups with a p-value of 0.485[29], hence preventing us from rejecting the null hypothesis. Table 7 shows the summary statistics.

Post-hoc power analysis[30] revealed an achieved power of 85.97[31] percent with a significance level of 5 percent. Table 6 summarizes the post-hoc power analyses. Based on our initial estimates, this hypothesis test is adequately powered.

5.3 Hypothesis test II

To perform the second hypothesis test[32], we first filtered the data to only include males. Then we divided the data into two groups, those primed with empathy questions and those not primed with empathy questions. We found null effects between these groups with a p-value of 0.736[33], hence preventing us from rejecting the null hypothesis. Table 7 shows the summary statistics.

Post-hoc power analysis[34] revealed an achieved power of 53.00[35] percent with a significance level of 5 percent. Table 6 summarizes the post-hoc power analyses with more details in Appendix E. Based on our pre-registration, this hypothesis test is underpowered.

	Achieved Power	Effect Size	Sample Sizes
Hypothesis I	84.97%	0.3234	In-group: 164 Out-group: 201
Hypothesis II	53.00%	0.3168	Primed Males: 81 Non-Primed Males: 95

Table 6: Post-hoc power analyses for a significance of 5 percent

[28] Pre-registered on AsPredicted.org
[29] P-value associated with the Mann Whitney U test
[30] Calculated using the software G*Power
[31] Post-hoc power analysis with an effect size of 0.3234 at a 5 percent significance
[32] Pre-registered on AsPredicted.org
[33] P-value associated with the Mann Whitney U test
[34] Calculated using the software G*Power
[35] Post-hoc power analysis with an effect size of 0.3168 at a 5 percent significance

	U-Statistic	P-Value
Hypothesis I	17181.0[36]	0.485[37]
Hypothesis II	3961.5[38]	0.736[39]

Table 7: Hypotheses test results

5.4 Linear regression

We performed an simple linear regression with Unforgiveness Scores as the dependent variable and the categorical independent variables Group (with two levels indicating if the participants were in-group or out-group), and Empathy Primed (with two levels indicating if the participants were primed with empathy questions or not before responding to the TRIM questions)[40]. Additionally, the following control variables were added to the equation:

- Age: Numeric value
- Gender: Categorical with two values, Male and Female (Male used as baseline for dummy variables)
- Race: Categorical with five values, Caucasian, Black, Asian, Hispanic, and Other (Other used as baseline for dummy variables)
- Education Qualification: Categorical with seven values, Bachelor, Doctorate, High School, Master, Some High School, Trade School, and Other (Other used as baseline for dummy variables)
- Empathy Prime: Categorical with two values, Prime and Non-Prime (Non-Prime used as baseline for dummy variables)

[36] Obtained from the Mann Whitney U test
[37] P-value not significant compared with the pre-registered significance level of 5 percent
[38] Obtained from the Mann Whitney U test
[39] P-value not significant compared with the pre-registered significance level of 5 percent
[40] Pre-registered on AsPredicted.org

- Group: Categorical with two values, In-Group and Out-Group (Out-Group used as baseline for dummy variables)

In the linear regression analysis, we were able to confirm that the variables Group[41] and Empathy Primed[42] did not have predictive power on the Forgiveness scores, with p-values of the coefficients 0.216[43] and 0.486[44] respectively. Interestingly, both the Age and Gender variables had statistically significant p-values — less than 0.05[45] for Age and less than 0.05[46] for Gender — for their coefficients implying there may be some effect of Gender and Age in the Forgiveness levels that is worth exploring. The low p-values for Table 8 summarizes the results. Appendix F and Appendix G discuss statistically-significant results for gender- and age-based relationships independently with forgiveness levels whereas Appendix H discusses their collective statistically-significant effects on forgiveness.[47] These three analyses were not pre-registered, hence we explain them in the Appendix section.

[41] Group In-Group was the dummy variable in the analysis and Group Out-Group was the baseline
[42] Empathy Prime was the dummy variable in the analysis and Empathy Non-Prime was the baseline
[43] Obtained from a simple linear regression analysis
[44] Obtained from a simple linear regression analysis
[45] Threshold significance level pre-registered on AsPredicted.org
[46] Threshold significance level pre-registered on AsPredicted.org
[47] This analysis was not pre-registered on AsPredicted.org

	Coefficient	Std Error	t	p-value	CI (0.025)	CI (0.975)
Constant	15.9892	2.2	7.268	0	11.663	20.316
Age	-0.0853	0.023	-3.688	0[48]	-0.131	-0.04
Gender Female	-1.3531	0.582	-2.324	0.021[49]	-2.499	-0.208
Race African	2.3184	1.933	1.199	0.231	-1.483	6.12
Race Asian	-0.1685	1.846	-0.091	0.927	-3.799	3.462
Race Caucasian	-0.1537	1.668	-0.092	0.927	-3.434	3.127
Race Hispanic	-1.281	2.069	-0.619	0.536	-5.351	2.789
Education Bachelor	1.0997	1.656	0.664	0.507	-2.158	4.357
Education Doctorate	0.1487	2.312	0.064	0.949	-4.398	4.695
Education High School	-0.4583	1.711	-0.268	0.789	-3.823	2.907
Education Master	1.0991	1.756	0.626	0.532	-2.354	4.552
Education Some High School	-6.1507	5.812	-1.058	0.291	-17.581	5.28
Education Trade School	3.0984	2.186	1.418	0.157	-1.2	7.397
Empathy Prime	-0.3992	0.572	-0.698	0.486	-1.524	0.726
Group In-group	0.7531	0.607	1.241	0.216	-0.441	1.947

Table 8: Simple linear regression analysis

[48] Statistically significant p-value according to the pre-registered significance level of 5 percent
[49] Statistically significant p-value according to the pre-registered significance level of 5 percent

6. Discussions and Conclusions

For the purpose of this study, we had two hypotheses — testing in-group versus out-group forgiveness and primed males versus non-primed males. The second hypothesis that males primed with empathy questions are more likely to be forgiving compared to males who are not primed with empathy, was a conceptual replication attempt of the Exline et al. 2008 paper. Our extension hypothesis was that people primed with the same race are more likely to forgive compared to those primed with a different race. We received a total of 365 responses from M-Turk. The sample sizes for the treatment groups are in Table 9.

Treatment	In-Group-Prime	In-Group-Non-Prime	Out-Group-Prime	Out-Group-Non-Prime
Sample	79	85	102	99

Table 9: Sample sizes for each treatment

The dependent variable being calculated was the unforgiveness score. We did not find any significant effect of In-Group versus Out-Group treatments (p-value = 0.485[50]) as well as for the males in the Prime vs Non-Prime treatments (p-value = 0.736[51]) when it came to the unforgiveness scores. Post analyzing the responses, we saw that we had null results for the two hypotheses. Upon running regressions on the various variables in our data collected, we saw a significant effect of participant's age and gender on unforgiveness scores.[52] Our pretest conducted within our personal networks using the Qualtrics survey showed that on average, to get an 80 percent powered study and to have a significance level of 5 percent for pairwise comparisons of means, we needed a sample size of 228 participants. We received a final sample of 365 responses in our M-Turk study, however, it was still underpowered for a lot of conditions due to an uneven distribution of samples across treatments. This limitation has been discussed in our conclusion section.

The results of Exline et al. 2008 did not replicate in our study[53]. There were certain modifications made to the original design which may have led to the failure of replication.[54] For

[50] Obtained from the Mann Whitney U test
[51] Obtained from the Mann Whitney U test
[52] People tend to be more forgiving with increasing age and males are more likely to be unforgiving compared to females. However these were not a part of our initial hypothesis.
[53] The replication was a conceptual one, with slight changes in the experiment design discussed earlier
[54] What was interesting is that the original paper did not find an effect of empathy priming on the unforgiveness scores of females. In our study we do find a stronger correlation for females in the Prime

the 4 treatment groups, we also did a pairwise analysis using the Mann-Whitney-U (two way) test. We did not find significant results for any of the 6 pairs (Table 5). In order to identify if any of our hypotheses were true nulls or not, we did a power calculation using G-Power software. We saw that the study had a power of 84.97 percent for the first hypothesis (in-group vs out-group) and the second hypothesis had a power of 53 percent (males primed vs males not primed). While hypothesis 2 was underpowered, the first hypothesis that people are more forgiving towards in-group members versus out-group members was sufficiently powered. This leads us to believe that it is likely that the result for hypothesis 1 is a true null (meaning there is not a significant difference between forgiveness levels of those primed with the same race versus those primed with a different race transgressor) and there could be further verification of that since the effect size is around 0.3. A possible consequence of this is that people may not have been primed with race from the vignettes and pictures, which is further supported by our regression analysis. In fact, we've explored the effect of gender on unforgiveness after observing the regression results, which is statistically significant. In case of hypothesis 2 (males primed vs males not primed), we had an underpowered study despite meeting the total sample size requirements because of a potentially weak pre-test. Further investigation is needed to conclude with our pre-registered power and confidence if this indeed is a true null.

versus Non-Prime treatments (p-value = 0.11) compared to males (p-value = 0.736). While not significant at the 5 percent level, we do think this could be an insight worth pursuing in further extensions or replications of this study.

7. Self Evaluation

Given that this was the first time we ran our own experiment, we have achieved quite a lot in terms of experience of conducting an end-to-end experiment, from the literature review to the documentation of our results. Although overall, the experiment went smoothly, we definitely feel there are many areas for improvement.

7.1 What went well

Due to the fact that the backbone of our study design was based on a previous study, we saved a lot of time by selecting the right tests to test our research topic — forgiveness and unforgiveness on the race of transgressor — using the TRIM questionnaire to measure the revenge motivation. Moreover, because we did not have to run a field experiment and collect our own data and, instead, used M-Turk as the data collection tool, the data collection process was relatively quick and smooth. Our team complemented each other's strengths extremely well. Because our team members have a diverse set of skills, each member contributed to the work in their own areas of expertise, while also seeking out opportunities to learn new skills from one another. We were able to corroborate each other's quantitative analyses, which helped us identify and rectify each other's mistakes, which otherwise would have gone unnoticed if only one person was in-charge of the data analysis. We also tested the calculations using a combination of Excel, Python, and R for robustness.

Furthermore, our pre-registered regression analysis gave us interesting avenues to conduct research, for example, exploring the effect of gender and age on forgiveness. Our team was eager to explore them in further analyses, which we've expanded in the Appendix section.

7.2 Areas of improvement

In the previous section, we discussed how post-analyzing the final results showed that our hypothesis tests were underpowered. This means that although the total number of participants we needed for some of the treatments were more than sufficient, we underestimated the sample sizes for other treatments. This is a possible indicator that our pretest may not have been as comprehensive as it could have been and resulted in us unable to make solid conclusions on whether or not to reject the null hypothesis. It is possible that the

respondents from our pretest, whom we recruited from our close circles may not have been a good representation of the population — a case of sampling error. In retrospect, we could have run a pretest with a small sample from M-Turk, in order to mitigate this issue.

Moreover, although finding null effects is acceptable and may indicate the true lack of effects, however, we feel that the design of our study may not have been as rigorous and it could have been. Upon reflection and after receiving the professor's and our classmates' feedback, there are a couple of areas in our design that could have been improved.

Firstly, the association between the priming of out-groups and in-groups and the revenge motivation may have been too weak. In our study, we decided to use an image of the person, while varying their race, as a manipulation of our independent variable. The image was shown after the participants read the vignette on a separate screen. While allocating the prime its own screen may have increased its salience, it may also have disassociated that prime from the revenge motivation induced by the vignette. A possible change for future studies could be to place the image of the transgressor on the same screen as the vignette to create a strong association between the priming cues and the induced revenge motivation. Another interesting design-choice could be to place the image of the transgressor on each screen of the questionnaire in order to deepen the associate and maintain the strength of in-group and in-group priming.

It is also possible to add more varieties of races in future designs. In our current design, participants from Indian backgrounds were lumped with other categories of Asian ethnicity. Given that a large population of Asian M-Turk participants are of Indian ethnicity[55], the priming of in-group may have been too weak when showing the participant of an Eastern Asian looking transgressor. Of all the methods we have mentioned, we think it would be prudent to run a small pilot test to find supporting evidence on the methods that work best to use in our main study.

Moreover, because the TRIM questionnaire was designed to measure the revenge motivation toward a transgressor whom the participants had some sort of relationship before, we could have provided a brief context on the profile of the transgressor to set a prior-relationship baseline with the participant. This can serve as a reference point in which the

[55] Based on a study from 2018, the top countries where the M-Turk workers are from are USA (75%), India (16%), Canada (1.1%), Great Britain (0.7%), Philippines (0.35%), and Germany (0.27%) (Difallah, 2018).

participants, after reading the transgression vignette, can use to measure how much they have shifted in closeness with the transgressor. The perceived closeness of relationship with the transgressor can be measured using oneness scale. For instance, we could have used the "Inclusion of the Other in the Self" (IOS) Scale, which is a pictorial measure of closeness, developed by Aron and Smollan (Aron et. al, 1992), to measure how close the participants feel toward the transgressor before and after reading the vignette.

Furthermore, the vignette presented may not have been as effective in inducing the revenge motivation from the participants. In the original study, participants were asked to think of a time somebody they knew had hurt them. This may have been a more effective method in terms of inducing the revenge motivation due to the fact that people have different levels of perceptions on what is considered as acceptable or unacceptable. By asking the participants about their own experience, the manipulation is personalized to each participant. However, in our study, because we were also interested in manipulating the race of the transgressor, it would have been difficult to ask our participants to think of an experience they were hurt by a person of a particular race due to individual differences in life experience. What we could have done was to prepare multiple vignettes with varying themes of transgression, ask a question prior to the vignette on which types of transgressions they feel most strongly about, and display the vignette that corresponds to the participants' answer.

Lastly, studies conducted with online participants, especially on M-Turk, are not always representative of the population. It is possible that our M-Turk participants may have been overexposed to many social science experiments, and thus, reducing the efficacy of the primes on them. Also, participants may have been more motivated to complete the experiment as quickly as possible or were participating in multiple experiments at the same time while working on our study. Such possibilities could be mitigated by comparing the results from M-Turk with another offline study in future studies.

8. Appendix

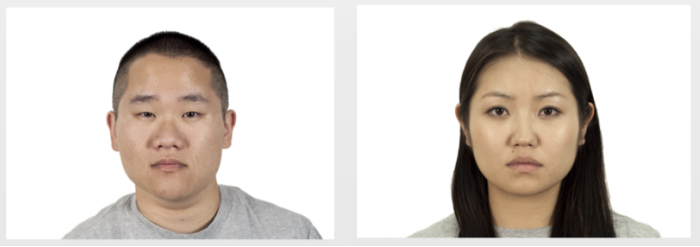
Appendix A: Chicago Face Images Database

To make the racial identities of participants more salient, we used photographs from the Chicago Face Database (Ma et al., 2015). This allowed us to show participants a set of players' faces that was homogeneous in facial expression, lightning, angle, and image resolution.

In the database, each image had associated numeric values for subjective emotional responses of people to the images, like angry, attractive, disgusted, threatening, etc. They also had numeric values for objective facial features like nose length, lip thickness, etc. Finally, there was a value for lumination.

We calculated weighted scores for each image based on the above factors. We weighed the emotional responses as 1 and assigned each of the objective facial features and lumination values with a weight of 0.1. After calculating the weighted sum for each image, we found the median value within each race-gender combination — Asian-Male, Asian-Female, Black-Male, Black-Female, Caucasian-Male, Caucasian-Female, Hispanic-Male, and Hispanic-Female — and selected the image for each group with the median weighted sum. The final images used in the survey as shown below:

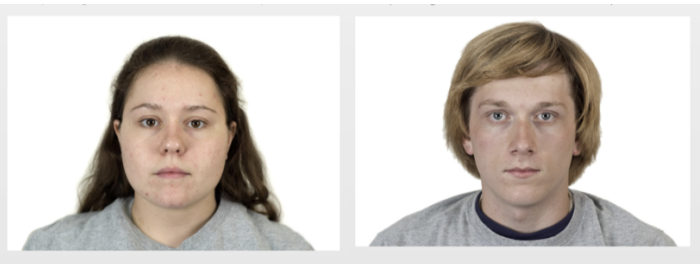
Asian



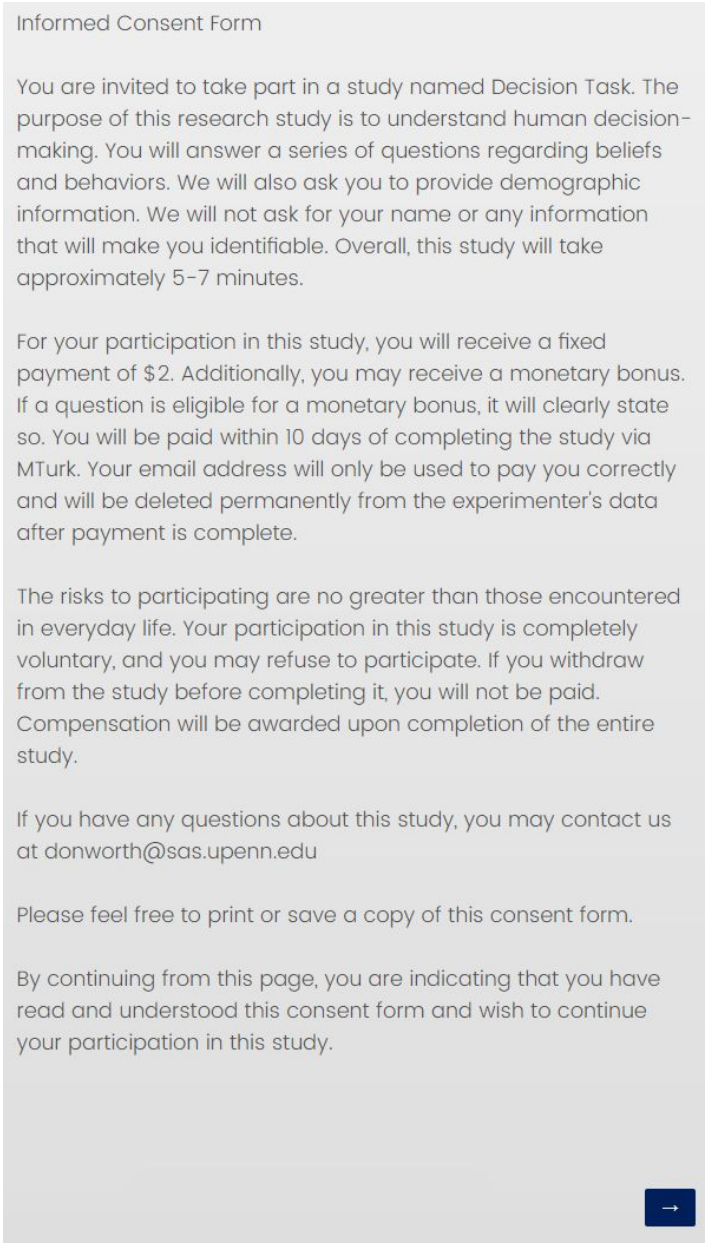
Black

Hispanic

Caucasian



Appendix B: Survey Flow



Please indicate your gender:

- Male
- Female
- Other

Please tell us your age:

Which race best describes you?

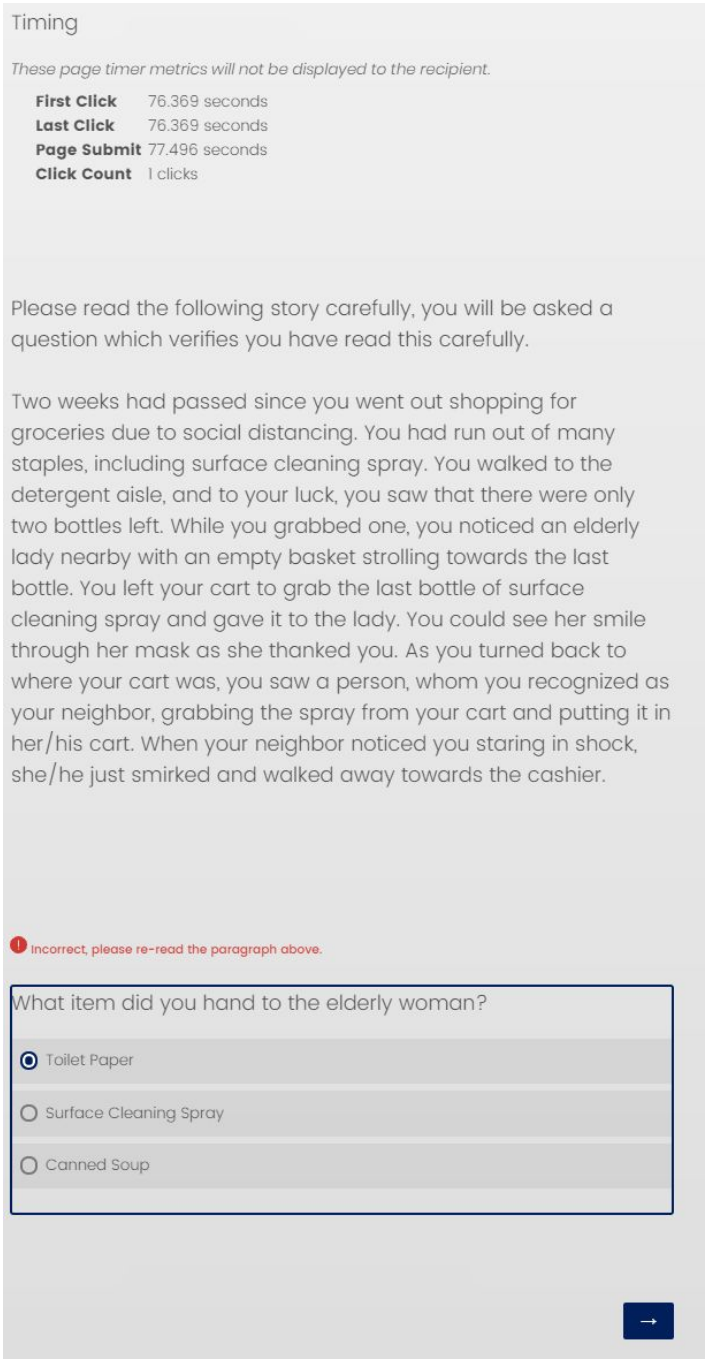
- African or African American
- Asian, Indian or Middle Eastern
- Caucasian or European
- Hispanic
- Other

What is the highest degree or level of education you have completed?

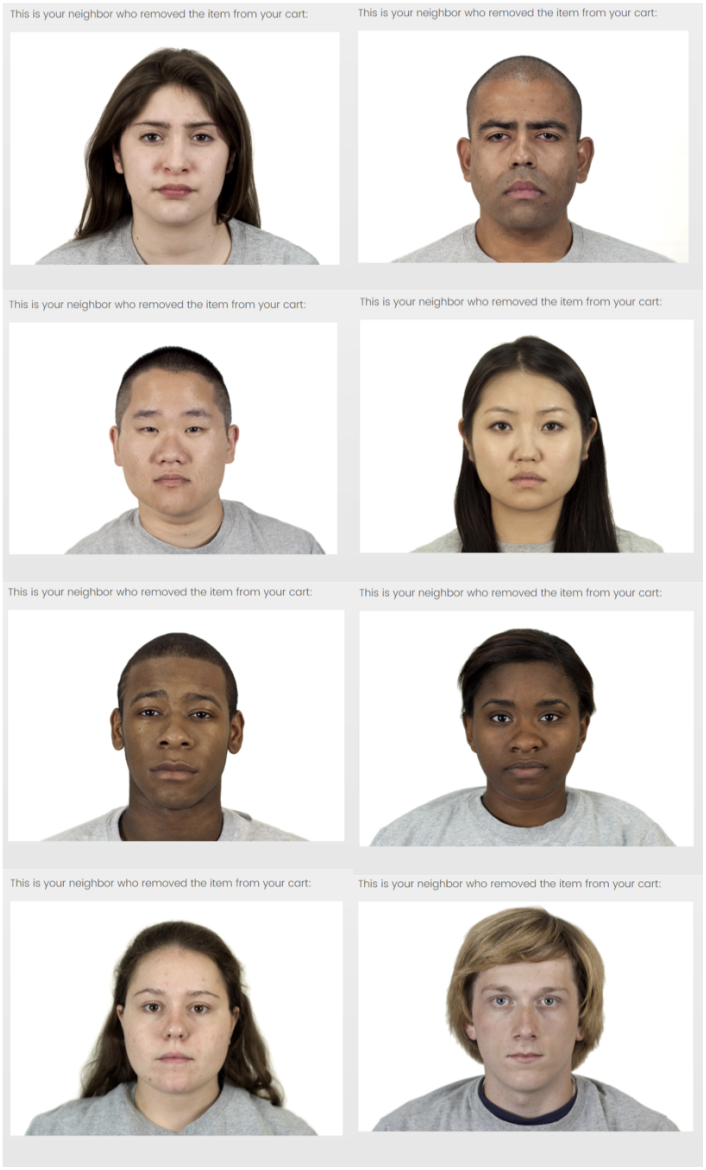
- Some High School
- High School
- Bachelor's Degree (e.g. BA, BS)
- Master's Degree (e.g. MA, MS, MEd)
- Doctorate or higher (e.g. PhD, EdD)
- Trade School
- Other

→

Vignette has an 11 second timer. The question below the vignette must be answered correctly (Surface Cleaning Spray) to move on. Incorrect answers will be shown the message below, and timer resets.



Randomized images of transgressors, one image is shown per survey. No timer used on this page.



TRIM Questions (immediately below) and CSE questions (shown after TRIM questions) will be randomly shown one before the other. TRIM Questions will be shown in the same order with no randomization within the 18 questions. TRIM Questions have a 15 second timer before the ‘next’ button appears.

Timing

These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

For the following questions, please indicate your current thoughts and feelings about the person who hurt you; that is, we want to know how you feel about that person right now. Indicate which answer best describes your current thoughts and feelings.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
I'll make him/her pay. (1)	○	○	○	○	○
I am trying to keep as much distance between us as possible. (2)	○	○	○	○	○
Even though his/her actions hurt me, I have goodwill for him/her. (3)	○	○	○	○	○
I wish that something bad would happen to him/her. (4)	○	○	○	○	○
I am living as if he/she doesn't exist, isn't around. (5)	○	○	○	○	○
I want us to bury the hatchet and move forward with our relationship. (6)	○	○	○	○	○
I don't trust him/her. (7)	○	○	○	○	○
Despite what he/she did, I want us to have a positive relationship again. (8)	○	○	○	○	○

The CSE questions below all have a timer of 5 seconds individually. One question shown per page. The first 8 questions shown below are randomized, with the final 4 questions not-randomized and stuck at the bottom of this question set. This question set and the TRIM questions above are randomly shown one before the other.

Timing

These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

Do you think it's possible that you could ever do something as bad as what the other person did?

- ○ (0) No, definitely not
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Yes, definitely

→

Timing

These page timer metrics will not be displayed to the recipient.

First Click 0 seconds
Last Click 0 seconds
Page Submit 0 seconds
Click Count 0 clicks

Do you think it's possible that you could ever do something similar in type to what the other person did?

- ○ (0) No, definitely not
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Yes, definitely

→

Timing

These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

Given the right circumstances, do you think that you could be capable of doing something just as bad (i.e., just as harmful or wrong) as what the other person did?

- ○ (0) No, definitely not
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Yes, definitely

Timing

These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

Can you imagine a situation in which you could do something similar in type to what the other person did?

- ○ (0) No, definitely not
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Yes, definitely

→

Timing

These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

Thinking back over your entire life, do you think that you have ever done anything similar in type to what the other person did?

- ○ (0) No, definitely not
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Yes, definitely

→

Timing

These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

Thinking back over your entire life, do you think that you have ever done anything as bad as what the other person did?

- ○ (0) No, definitely not
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Yes, definitely

→

Timing

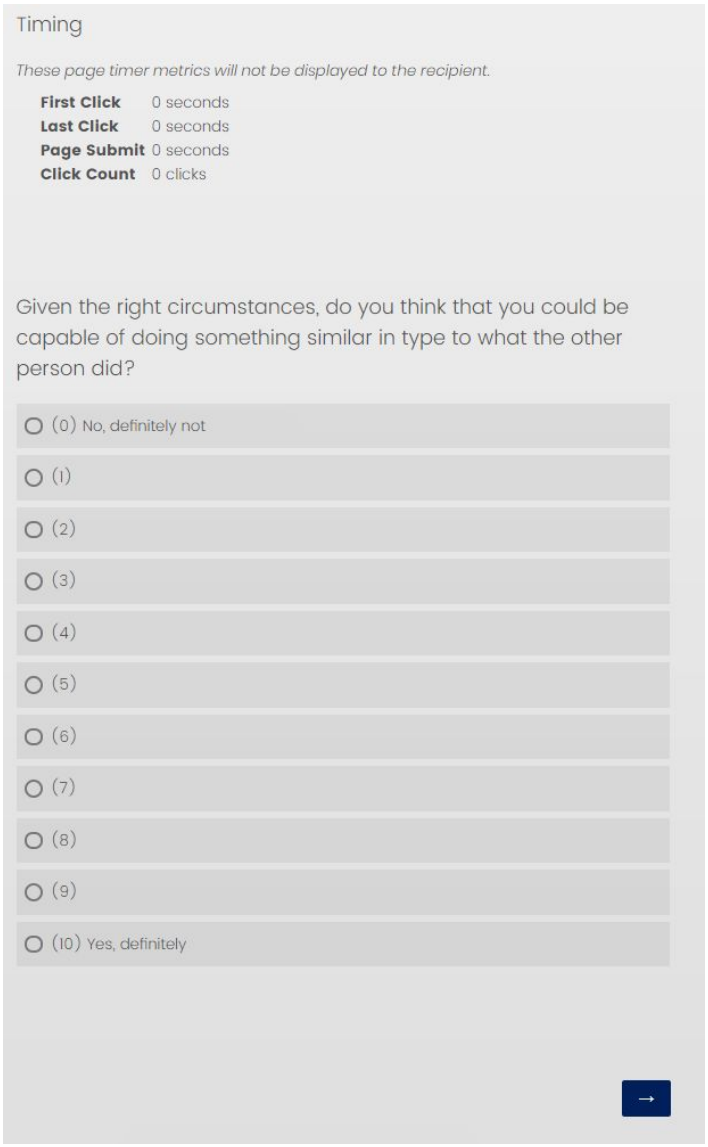
These page timer metrics will not be displayed to the recipient.

First Click 0 seconds
Last Click 0 seconds
Page Submit 0 seconds
Click Count 0 clicks

Can you imagine a situation in which you could do something as bad as what the other person did?

- ○ (0) No, definitely not
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Yes, definitely

→



The next 4 CSE questions are stuck at the bottom of this question set in a non-random order. Still shown as one question per page with a 5 second timer.

Timing

These page timer metrics will not be displayed to the recipient.

First Click 0 seconds
Last Click 0 seconds
Page Submit 0 seconds
Click Count 0 clicks

To what extent can you understand why the other person acted as s/he did?

- ○ (0) Not at all
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Totally

Timing

These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

To what extent can you see the situation from the other person's perspective?

- ○ (0) Not at all
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Totally

Timing

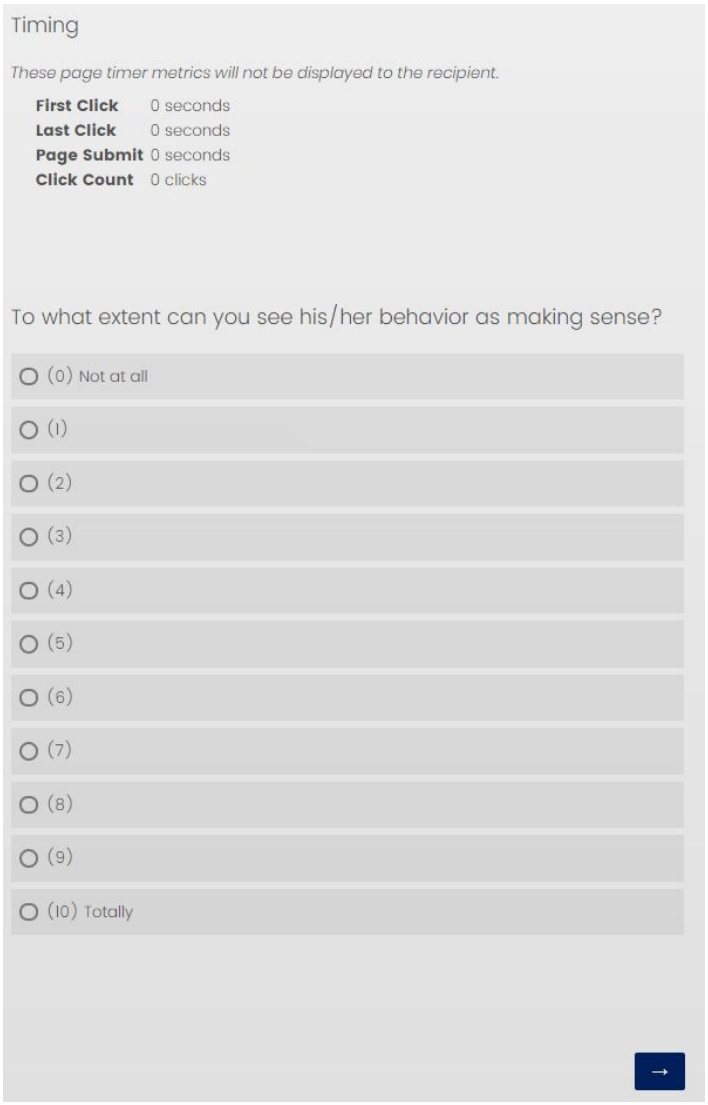
These page timer metrics will not be displayed to the recipient.

First Click	0 seconds
Last Click	0 seconds
Page Submit	0 seconds
Click Count	0 clicks

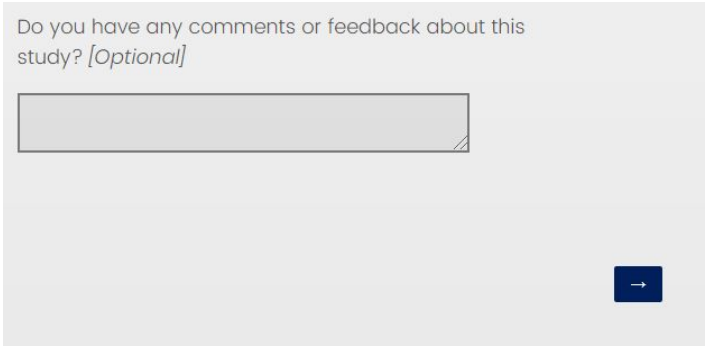
To what extent can you think of valid reasons why s/he acted as s/he did?

- ○ (0) Not at all
- ○ (1)
- ○ (2)
- ○ (3)
- ○ (4)
- ○ (5)
- ○ (6)
- ○ (7)
- ○ (8)
- ○ (9)
- ○ (10) Totally

→



Final survey question, and the only optional question of the survey:



We thank you for your time spent taking this survey.
Your response has been recorded.

Appendix C: Transgression Related Interpersonal Motivations Inventory–Revised

Trim-18
(McCullough, Root, & Cohen, 2006)

For the following questions, please indicate your current thoughts and feelings about the person who hurt you; that is, we want to know how you feel about that person **right now**. Next to each item, circle the number that best describes your current thoughts and feelings.

	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5
1. I'll make him/her pay.	1	2	3	4	5
2. I am trying to keep as much distance between us as possible.	1	2	3	4	5
3. Even though his/her actions hurt me, I have goodwill for him/her.	1	2	3	4	5
4. I wish that something bad would happen to him/her.	1	2	3	4	5
5. I am living as if he/she doesn't exist, isn't around.	1	2	3	4	5
6. I want us to bury the hatchet and move forward with our relationship.	1	2	3	4	5
7. I don't trust him/her.	1	2	3	4	5
8. Despite what he/she did, I want us to have a positive relationship again.	1	2	3	4	5
9. I want him/her to get what he/she deserves.	1	2	3	4	5
10. I am finding it difficult to act warmly toward him/her.	1	2	3	4	5
11. I am avoiding him/her.	1	2	3	4	5
12. Although he/she hurt me, I am putting the hurts aside so we can resume our relationship.	1	2	3	4	5
13. I'm going to get even.	1	2	3	4	5
14. I have given up my hurt and resentment.	1	2	3	4	5
15. I cut off the relationship with him/her.	1	2	3	4	5
16. I have released my anger so I can work on restoring our relationship to health.	1	2	3	4	5
17. I want to see him/her hurt and miserable.	1	2	3	4	5
18. I withdraw from him/her.	1	2	3	4	5

Scoring Instructions

Avoidance Motivations:
Add up the scores for items 2, 5, 7, 10, 11, 15, and 18

Revenge Motivations:
Add up the scores for items 1, 4, 9, 13, and 17

Benevolence Motivations
Add up the scores for items 3, 6, 8, 12, 14, and 16

Citation:

McCullough, M. E., Root, L. M., & Cohen, A. D. (2006). Writing about the personal benefits of a transgression facilitates forgiveness. *Journal of Consulting and Clinical Psychology, 74*, 887-897.

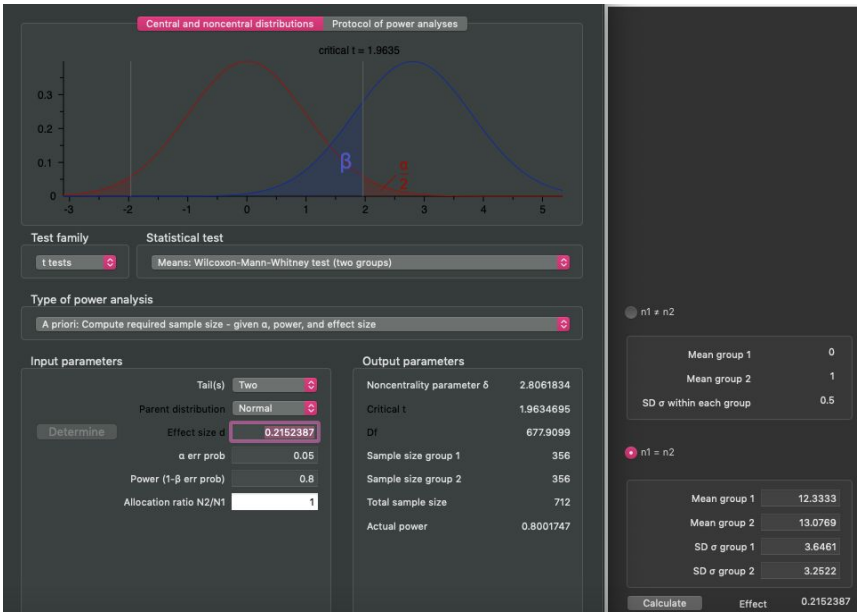
Appendix D: A priori power analysis for sample size determination

We conducted a pretest to calculate initial estimates of effect sizes to determine our sample size calculations for the main study. We were able to collect 59 records in total. We used the results from this pretest sample to conduct our a priori power analysis.

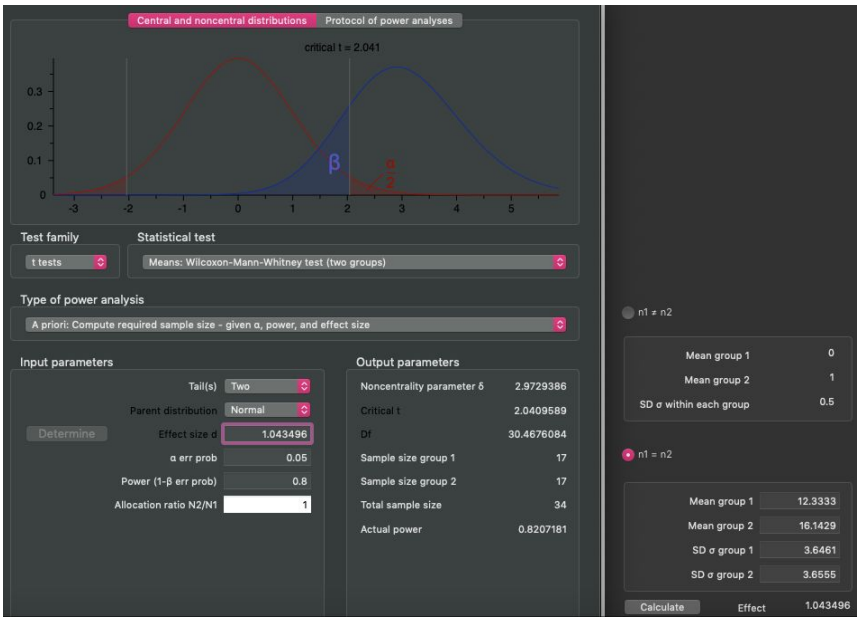
Per-Treatment Pre-Test Results

Treatment	Count	Mean	Standard Deviation
In-group <> TRIM-E	18	12.3333	3.6461
In-group <> E-TRIM	13	13.0769	3.2522
Out-group <> TRIM-E	14	16.1429	3.6555
Out-group <> E-TRIM	14	12.4286	4.3450

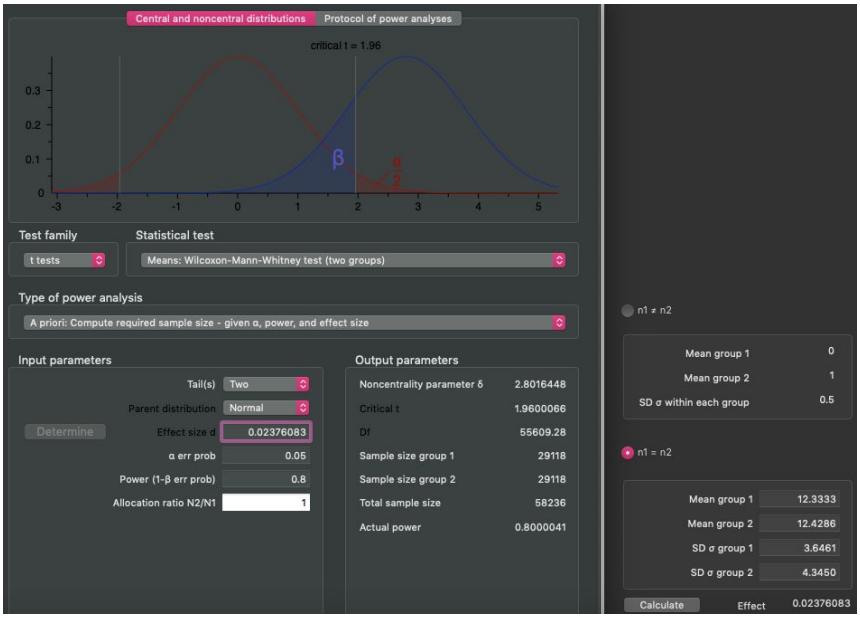
In-group <> TRIM-E and In-group <> E-TRIM: 712 (356 each)



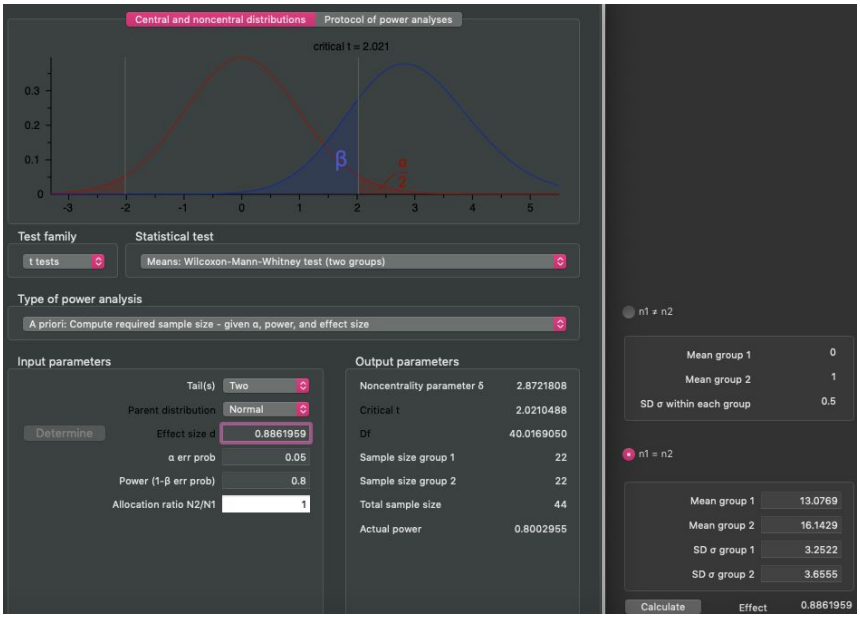
In-group <> TRIM-E and Out-group <> TRIM-E: 34 (17 each)



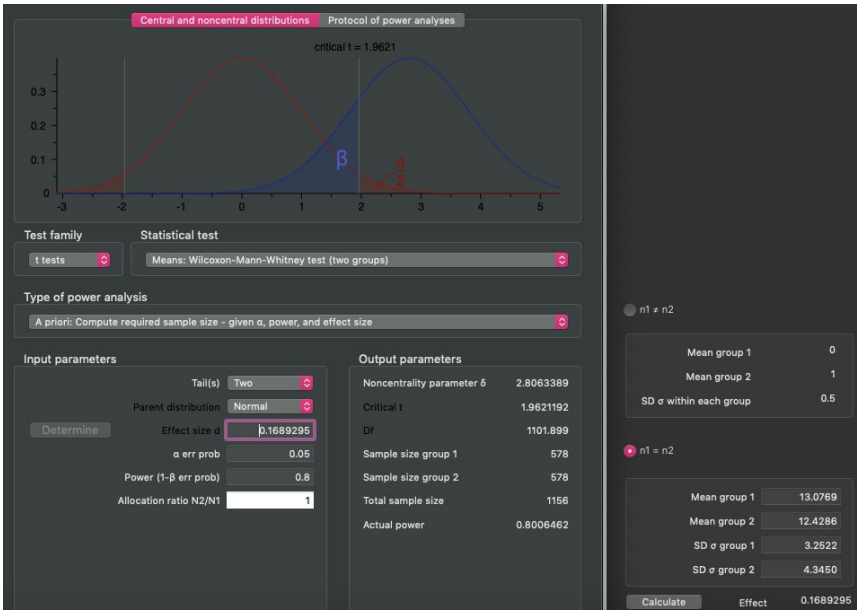
In-group <> TRIM-E and Out-group <> E-TRIM: 58236 (29118 each)



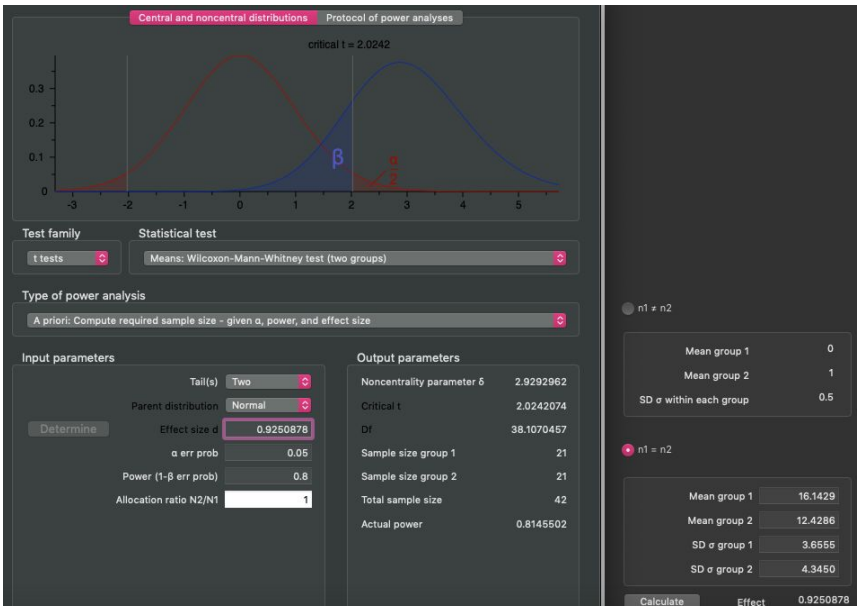
In-group <> E-TRIM and Out-group <> TRIM-E: 44 (22 each)



In-group <> E-TRIM and Out-group <> E-TRIM: 1156 (578 each)

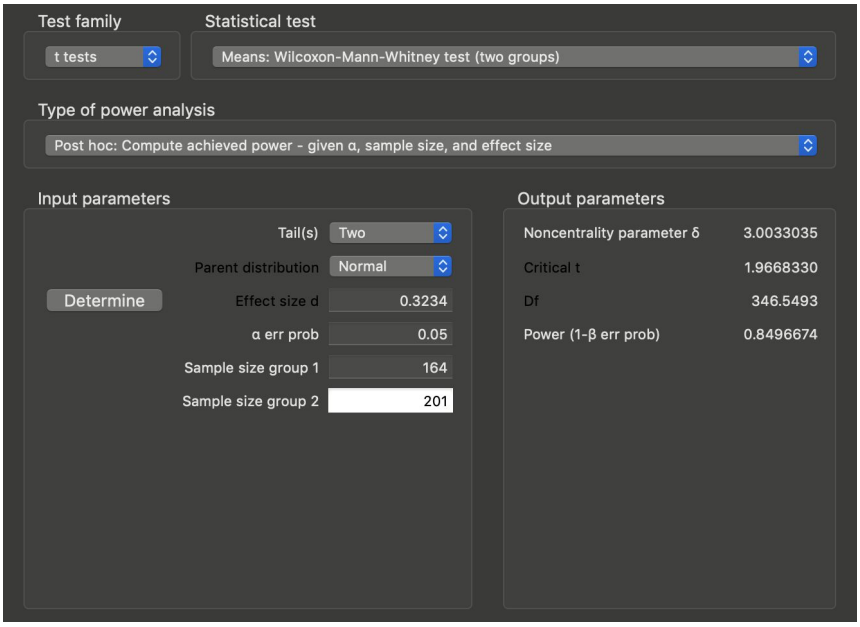


Out-group <> TRIM-E and Out-group <> E-TRIM: 42 (21 each)

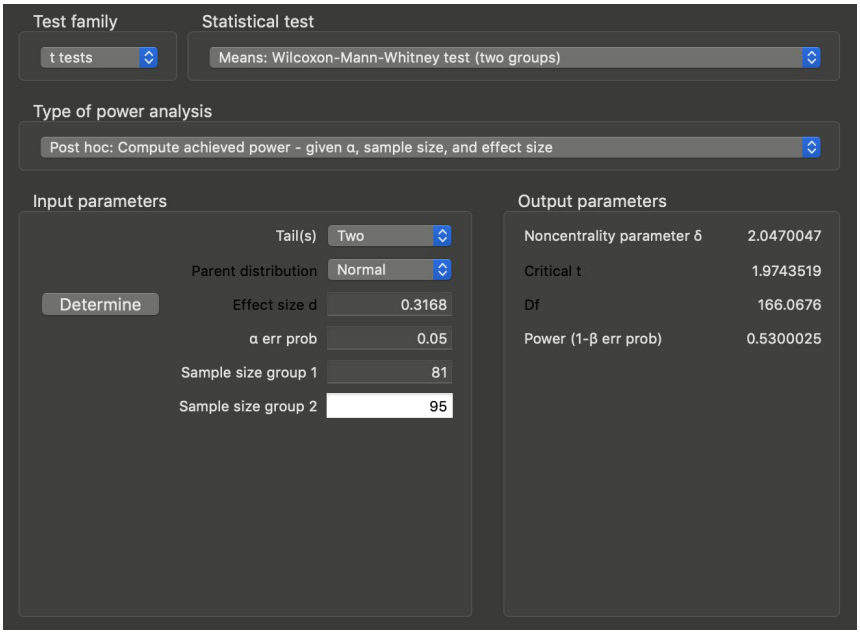

Central and noncentral distributions
Protocol of power analyses
critical t = 2.0242
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
4
5
β
α/2
Test family
t tests
Statistical test
Means: Wilcoxon-Mann-Whitney test (two groups)
Type of power analysis
A priori: Compute required sample size - given α, power, and effect size
Input parameters
Tail(s)
Two
Parent distribution
Normal
Determine
Effect size d
0.9250878
α err prob
0.05
Power (1-β err prob)
0.8
Allocation ratio N2/N1
1
Output parameters
Noncentrality parameter δ
2.9292962
Critical t
2.0242074
Df
38.1070457
Sample size group 1
21
Sample size group 2
21
Total sample size
42
Actual power
0.8145502
n1 ≠ n2
Mean group 1
0
Mean group 2
1
SD σ within each group
0.5
n1 = n2
Mean group 1
16.1429
Mean group 2
12.4286
SD σ group 1
3.6555
SD σ group 2
4.3450
Calculate
Effect
0.9250878

Appendix E: A post-hoc power analysis for achieved power

Hypothesis I power calculation



Hypothesis II power calculation



Appendix F: Gender-based effects on forgiveness levels

When comparing forgiveness levels of males with those of females, we found interesting effects that are worth exploring in future studies. Particularly, we observed that males, either primed or not primed, had statistically-significant different forgiveness levels from females who were primed with empathy questions. We used the non-parametric Mann Whitney U test for these analyses.

These results, paired with the results from the regression analysis in the table below, showed that females primed with empathy questions tended to be more forgiving than males, primed or not primed with empathy questions.

	Male-Non-Prime	**Male-Prime**
Female-Non-Prime	U = 4561.0 p = 0.355	U = 4011.0 p = 0.204
Female-Prime	U = 5810.5 p = 0.007	U = 5143.0 p = 0.002

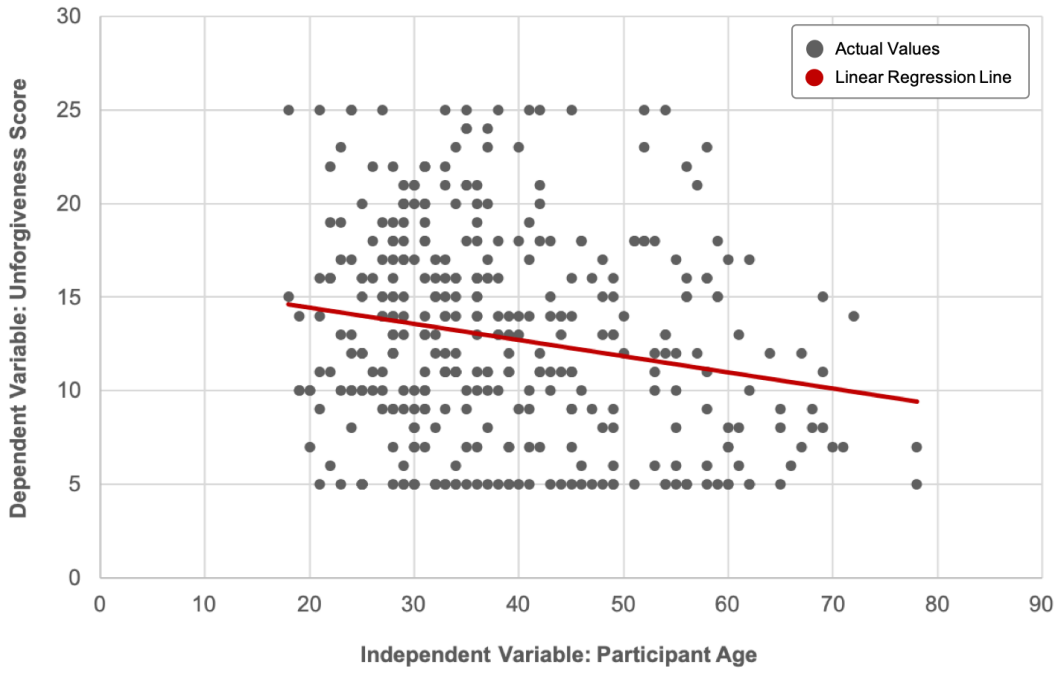
Sample size for Empathy Primed Males and Non-Empathy Primed Males

Appendix G: Age-based effects on forgiveness levels

When performing a linear regression for Unforgiveness scores with participant's age as the dependent variable we find a positive relationship between their age and their tendency to forgive.

	Coefficient	Std Error	t	p-value	CI (0.025)	CI (0.975)
Constant	16.177	0.919	17.610	0	14.370	17.983
Age	-0.087	0.023	-3.845	0[56]	-0.042	-0.131

Simple linear regression analysis for Age versus Unforgiveness Score



Simple linear regression plot for Age versus Unforgiveness Score

[56] Statistically significant p-value according to the pre-registered significance level of 5 percent

Appendix H: Age- and gender-based effects on forgiveness levels

When conducting a follow-up linear regression analysis for testing any relationship between Age and gender, we obtain the following results. Age has the statistically-strongest relationship with a coefficient p-value, 0.000 up to three decimal places, less than our pre-registered confidence level, whereas Gender also had a strong predictive power over forgiveness levels, with a p-value of 0.008. The table below shows the summary statistics. The negative coefficients for both Age and Gender (Female)[57] suggest that:

1. Older people tend to be more forgiving than younger people
2. Females tend to be more forgiving than males

	Coefficient	Std Error	t	p-value	CI (0.025)	CI (0.975)
Constant	16.787	0.940	17.868	0	14.939	18.635
Age	-0.082	0.022	-3.661	0[58]	-0.126	-0.038
Gender Female	-1.521	0.572	-2.659	0.008[59]	-2.647	-0.396

Simple linear regression analysis for Age and Gender

[57] Dummy variables used for the categorical factor Gender, with Gender value Male as baseline
[58] Statistically significant p-value according to the pre-registered significance level of 5 percent
[59] Statistically significant p-value according to the pre-registered significance level of 5 percent

9. Bibliography

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. Journal of personality and social psychology, 63(4), 596.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.

Beckett, Katherine, Kris Nyrop, and Lori Pfingst. 2006. Race, drugs, and policing: Understanding disparities in drug delivery arrests. Criminology 44: 105– 37.

Berry, J. W., Worthington Jr, E. L., Parrott III, L., O'Connor, L. E., & Wade, N. G. (2001). Dispositional forgivingness: Development and construct validity of the Transgression Narrative Test of Forgivingness (TNTF). *Personality and Social Psychology Bulletin*, *27*(10), 1277-1290.

Bishop, Donna M., and Charles E. Frazier. 1996. Race effects in juvenile justice decision-making: Findings of a statewide analysis. The Journal of Criminal Law and Criminology 86: 392– 414.

Bridges, George S., Robert D. Crutchfield, and Edith E. Simpson. 1987. Crime, social structure and criminal punishment: White and nonwhite rates of imprisonment. Social Problems 34: 345–61.

Brown, R. P. (2003). Measuring individual differences in the tendency to forgive: Construct validity and links with depression. *Personality and Social Psychology Bulletin*, *29*(6), 759-771

Brown, R. P., Wohl, M. J., & Exline, J. J. (2008). Taking up offenses: Secondhand forgiveness and group identification. *Personality and Social Psychology Bulletin*, *34*(10), 1406-1419.

Demuth, S. (2003). Racial and ethnic differences in pretrial release decisions and outcomes: A comparison of Hispanic, Black, and White felony arrestees. Criminology, 41(3), 873-908.

Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February). Demographics and dynamics of mechanical Turk workers. In Proceedings of the eleventh ACM international conference on web search and data mining (pp. 135-143).

Exline, J. J., Baumeister, R. F., Zell, A. L., Kraft, A. J., & Witvliet, C. V. (2008). Not so innocent: Does seeing one's own capability for wrongdoing predict forgiveness?. *Journal of personality and social psychology*, *94*(3), 495.

Finch, J. (1987). The vignette technique in survey research. *Sociology*, *21*(1), 105-114.

Fincham, F. D., Paleari, F. G., & Regalia, C. (2002). Forgiveness in marriage: The role of relationship quality, attributions, and empathy. *Personal Relationships*, *9*(1), 27-37.

Kutateladze, B. L., Andiloro, N. R., Johnson, B. D., & Spohn, C. C. (2014). Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. Criminology, 52(3), 514-551.

Lieberman, A. (1987). Documenting Professional Practice: The Vignette as a Qualitative Tool.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, 47(4), 1122-1135.

Macaskill, A., Maltby, J., & Day, L. (2002). Forgiveness of self and others and emotional empathy. *The Journal of social psychology*, *142*(5), 663-665.

McCullough, M. E., & Hoyt, W. T. (2002). Transgression-related motivational dispositions: Personality substrates of forgiveness and their links to the Big Five. *Personality and Social Psychology Bulletin*, 28(11), 1556-1573.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington Jr, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of personality and social psychology*, *75*(6), 1586.

McCullough, M. E., Worthington Jr, E. L., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of personality and social psychology*, *73*(2), 321.

Nellis, A. (2016). The color of justice: Racial and ethnic disparity in state prisons. Sentencing Project.

Rachlinski, Jeffrey J., Sheri Lynn Johnson, Andrew J. Wistrich, and Chris Guthrie. 2009. Does unconscious racial bias affect trial judges? Notre Dame Law Review 84: 1195– 246.

SAALT. 2001. American Backlash. South Asian American Leaders of Tomorrow. saalt.org/wp-content/uploads/2012/09/American-Backlash-report.pdf.

Schlesinger, Traci. 2005. Racial and ethnic disparity in pretrial criminal processing. Justice Quarterly 22: 170– 92.

Spohn, Cassia C. 2013. Racial disparities in prosecution, sentencing, and punishment. In The Oxford Handbook of Ethnicity, Crime, and Immigration, eds. Sandra M. Bucerius and Michael Tonry. New York: Oxford University Press.

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior. In J. T. Jost & J.

Takaku, S. (2001). The effects of apology and perspective taking on interpersonal forgiveness: A dissonance-attribution model of interpersonal forgiveness. *The Journal of social psychology*, *141*(4), 494-508.

Tonry, Michael, and Matthew Melewski. 2008. The malign effects of drug and crime control policies on Black Americans. Crime and Justice 37: 1– 44.

Toussaint, L., & Webb, J. R. (2005). Gender differences in the relationship between empathy and forgiveness. *The Journal of social psychology*, *145*(6), 673-685.

Von Ahn, L., Blum, M., Hopper, N. J., & Langford, J. (2003, May). CAPTCHA: Using hard AI problems for security. In *International Conference on the Theory and Applications of Cryptographic Techniques* (pp. 294-311). Springer, Berlin, Heidelberg.

Witvliet, C. V. O., Ludwig, T. E., & Laan, K. L. V. (2001). Granting forgiveness or harboring grudges: Implications for emotion, physiology, and health. *Psychological science*, *12*(2), 117-123.

Group Level Assessment

Who made the best contributions to the group project (receives +2% on the group's grade)?

Student Name (can remain blank if does not apply): Ankit Saxena

Explain in detail how exactly (if applies):

Ankit went above and beyond to make sure we had ticked all the boxes on our project, he always ensured our slides and reports were expertly formatted

Who contributed the least to the group project (receives -2% on the group's grade)?

Student Name (can remain blank if does not apply): _____

Explain in detail how exactly (if applies):

_____

_____

_____

Things that worked particularly well in the group:

We cross-verified our work (especially the analyses) and identified avoidable errors. Also, we complemented our skills well, while also learning from each other's strengths. Overall, it was a very motivated team that loved learning and balancing the workload evenly.

Things that could have worked better in the group:

As mentioned in the Self-reflection section, we could have done a better job at the pre-analysis section. We should have per-registered some of the interesting analyses, which we eventually added in our Appendix.

Individual Level Assessment

(0 = fully disagree, 5 = fully agree)

Important: per assessment item, the highest grade '5' can only be assigned once (i.e., no two students can both have a 5 for "contributions were of high quality", but a single student can have multiple 5s if he/she excelled in multiple domains)

Student #1

- Name: Gregory Donworth
- Was reliable and timely (0 – 5): 5
- Contributed actively (0 – 5): 4
- Contributions were of high quality (0 – 5): 4
- Contributed significantly to the final presentation (0 – 5): 4
- Contributed significantly to the final paper (0 – 5): 4
- Contributed the following written parts to the paper: Wrote "Experimental Design" and "Appendix B" Revised: "Theoretical Background" "Bibliography" "Appendices"

Student #2

- Name: Shiqing Lin
- Was reliable and timely (0 – 5): 4
- Contributed actively (0 – 5): 4
- Contributions were of high quality (0 – 5): 4
- Contributed significantly to the final presentation (0 – 5): 4
- Contributed significantly to the final paper (0 – 5): 5
- Contributed the following written parts to the paper: Wrote "Introduction and Theoretical Background" Revised: "Bibliography"

Student #3

Name: Himani Mehta

- Was reliable and timely (0 – 5): 4
- Contributed actively (0 – 5): 4

- Contributions were of high quality (0 – 5): 5
- Contributed significantly to the final presentation (0 – 5): 4
- Contributed significantly to the final paper (0 – 5): 4
- Contributed the following written parts to the paper: Wrote “Conclusion & Discussion”; Revised/reviewed “Experimental Design”

Student #4

Name: Nyaknno Owodiong-Idemeko

- Was reliable and timely (0 – 5): 4
- Contributed actively (0 – 5): 4
- Contributions were of high quality (0 – 5): 4
- Contributed significantly to the final presentation (0 – 5): 4
- Contributed significantly to the final paper (0 – 5): 4
- Contributed the following written parts to the paper: Wrote “Introduction and Theoretical Background” Revised: “Conclusion & Discussion” “Bibliography”

Student #5

- Name: Ankit Saxena
- Was reliable and timely (0 – 5): 4
- Contributed actively (0 – 5): 4
- Contributions were of high quality (0 – 5): 4
- Contributed significantly to the final presentation (0 – 5): 5
- Contributed significantly to the final paper (0 – 5): 4
- Contributed the following written parts to the paper: Wrote “Methodology”, “Analysis Results” and “Appendix A, F, G, and H.”

Student #6

- Name: Ladasa Tiraviriyapol

- Was reliable and timely (0 – 5): 4
- Contributed actively (0 – 5): 5
- Contributions were of high quality (0 – 5): 4
- Contributed significantly to the final presentation (0 – 5): 4
- Contributed significantly to the final paper (0 – 5): 4
- Contributed the following written parts to the paper: Wrote “Self Evaluation”; Reviewed/revised “Discussions and Conclusion” and “Analysis Results”

I hereby confirm that the above statements are correct (signed by all group members)

Nyaknno Owodiong-Idemeko Gregory Donworth Ankit Saxena

Himani Mehta Ladasa Tiraviriyapol Shiqing Lin