Replication of Study 1 by Lau, Kay, & Spencer (2008, *Psychological Science*)

Anna C. Merritt
Stanford University Department of Psychology
amerritt@stanford.edu

Introduction

Lau, Kay, and Spencer (2008) sought to demonstrate that, in line with system-justification theory, people are powerfully motivated to restore the legitimacy of their sociopolitical systems when those systems come under threat. One way people bolster their belief in the status quo is by justifying societal inequalities, often through the endorsement of negative stereotypes about socially disadvantaged groups. The authors propose that the system-justification motivation can influence everyday interpersonal attraction, such that men who are under system threat would find women who conform to benevolent female stereotypes (e.g., vulnerable and pure) more attractive than counter-stereotypic women. Expressing a preference for stereotypic women implicitly communicates that gender inequality in society is justified. Consistent with this hypothesis, the original study found that male participants who read a negative editorial about their home country preferred benevolent stereotypic women to non-stereotypic women, while participants who read a neutral / positive editorial showed no significant preference.

Methods

Power Analysis

The paper described the original effect size of the primary analysis (a 2 x 2 within-between interaction ANOVA) as $ш^2 = .02$. However, I am unable to calculate the power of the effect because the authors do not provide the means of each condition, the cell sizes, or the correlation between the within-subjects measures. I plan to run a pilot study with the same number of participants as the original study (n=36) and do a power analysis at that point.

Power analysis *is* possible on two other effects described in the paper. Participants under system threat rated benevolent-sexist-stereotypic women as more attractive, compared to participants not under threat (between subjects). The power of this effect at n=36 is .54, suggesting that the original study was underpowered. To achieve power levels of .85, .90, and .95, one would need to run 74, 86, and 106 participants, respectively. Also, participants who were under threat rated benevolent-sexist-stereotypic women as more attractive than non-stereotypic women (within subjects). The power of this effect at n=18 (assuming that half of the participants were assigned to the threat condition) is .84. To achieve power levels of .85, .90, and .95, one would need to run 19, 22, and 26 participants, respectively.

Planned Sample

Participants should be single males, preferably around the age of college undergraduates (i.e., late teens - 20s). They will be recruited through the Mechanical Turk subject pool. Because MTurk does not allow me to screen participants by gender, age, or marital status, I will survey all participants and then discard those who do not meet the criteria.

Materials

"Participants read an excerpt from a British newspaper. The excerpt used in the system-threat condition threatened their belief in the legitimacy of their federal system, offering the opinion of a foreign journalist who believed the social, economic, and political climate in Canada was worsening. In the no-system-threat condition, the excerpt described the social, economic, and political climate in Canada as relatively stable and positive (see Kay, Jost, & Young, 2005)." In the replication, the article will be edited to refer to the United States (participants' home country). Because the article used in the original study was actually adapted from a U.S. version used in Kay, Jost, & Young (2005) I will use the version from the 2005 paper.

"Participants then rated their romantic interest in each of eight profiled women. Each profile included a picture and self-description, ostensibly taken from a dating Web site. Half of the profiles portrayed the women as embodying the three dimensions of benevolent sexism identified by Glick and Fiske (1996); that is, these women were portrayed as vulnerable, pure, and ideal for making men feel complete. The portrayals in the other four profiles were inconsistent with the benevolent sexist stereotype in various ways; for example, the women were presented as career oriented, party seeking, active in social causes, or athletic."

Procedure

As described above, participants will begin by reading one of two excerpts to manipulate system threat. Then they will see the women's' profiles. "The profiles were presented randomly, with the only constraint being that two stereotypical or two nonstereotypical profiles could not be presented sequentially. To ensure that any differences in romantic interest across the profiles were not due to differences in the pictures, we randomly paired pictures with self-descriptions for each participant. After viewing each profile, participants used 7-point scales to answer eight questions gauging their romantic interest. They reported the extent to which they found the woman attractive, were interested in chatting with her on-line, would like to get to know her better, would like to meet her personally, would likely invite her to a party, would be pleased to have a date with her, were interested in starting a relationship with her, and viewed her as the ideal romantic partner. We then calculated an index of romantic interest."

Analysis Plan

As mentioned above, I will discard participants who are female, married, or above the age of 29. In keeping with the original study, I will create two composite scores for each participant by averaging their ratings of stereotypic and non-stereotypic women, respectively. These composites assume that there will be a high degree of internal reliability, both among the 8 items used to measure romantic interest and participants' ratings of the 4 women of the same type ($a = .97$ in the original paper). I will run a repeated-measures ANOVA to test for the interaction reported in the original paper, and follow up with simple effects tests to examine the between- and within-subjects effects the authors reported.

Differences from Original Study

The major difference will be in the participant population - they will come from MTurk and not from an undergraduate subject pool. Another major difference is that they will complete the study online. Although the details of the setting are not described in the original paper, I assume that participants completed the study on a computer in some kind of lab space with an experimenter present or nearby. Finally, the pictures and text in the online dating profiles for the women will not be the same. I was unable to obtain these materials from the

original experimenter, so I created my own versions of the profiles based on the descriptions in the paper (see materials section). I obtained pictures from the website hotornot.com to use as stimuli, because these pictures are matched for attractiveness by many raters. I selected pictures that had an average rating of 9.0, and only chose those that were neutral in content (i.e., just a head and shoulders shot, no piercings or tattoos, normal facial expression). Because these women were all rated as quite attractive, there is a possibility of a ceiling effect on participants' interest and a chance that the study will be less believable than the original. However, because I was not able to get information on the original womens' pictures, it is impossible to know how much the ones used in the current study differ from the originals.

(Post Data Collection) Methods Addendum

Actual Sample

Participants were 40 members of the MTurk online participant pool, paid $0.70 each. One participant was excluded for marking their gender as female.

Differences from pre-data collection methods plan

I had planned to collect data from American participants, but unfortunately I forgot to include location as a condition for participants to complete the survey. This makes the results less valid, especially considering that the threat manipulation referred to conditions in the US.

Results

Data preparation

Alpha scores for the two types of profiles were computed and were satisfactory (a = .95 for nonstereotypical, .97 for stereotypical). For each participant, I created a composite of the 4 profiles of each type.

Confirmatory analysis

Neither threat nor stereotypicality influenced ratings of the profiles, nor was there the predicted interaction between threat and stereotypicality, $F(1, 38) = 1.51$, $p > .20$.

Discussion

Summary of Replication Attempt

The confirmatory analysis failed to replicate the original findings.

Commentary

Unfortunately, the failure to replicate can be easily explained by the weakness of the threat manipulation in this replication vs. the original study. Although I did not collect data on the location / nationality of the participants in the replication (which should certainly be added in a future replication attempt) it is quite likely that many were non-American, based on the overall demographics of the MTurk pool. Thus, the system threat manipulation, which was designed to raise concerns about the United States' economy, would not have been very effective. Ideally, the passage participants read should threaten their own home country's socioeconomic stability.