Replication of Stern et al, 2015
"Conservatives negatively evaluate counterstereotypical people to maintain a sense of certainty"
PNAS 112 (50), 15337-15342.
https://www.pnas.org/content/112/50/15337

The original paper includes several studies. We randomly chose study 7. In this within-subject experiment, participants read an excerpt of a story describing two fictitious groups. Participants are then shown four photographs of men, two of which belong to each group. In both groups, one target confirms the stereotype from the story of having / not having facial moles, and the other target deviates from the stereotype. Afterwards, participants are asked to allocate money (as gifts) to the four targets. In addition, they are asked to rate their political ideology. Conservatives allocate less money to counter-stereotypical targets than stereotypical targets.

Hypothesis to replicate and bet on: Conservatives allocate less money to targets who deviate from stereotypes (i.e., counterstereotypical targets). To evaluate this hypothesis, the authors perform a generalized estimating equations regression analysis ($B = 0.52$, $SE = 0.14$, $z = 3.65$, $p < 0.001$); p. 15339.

Criteria for replication: The criteria for replication are an effect in the same direction as in the original study and a p-value < 0.05 in the Wald χ^2-test using a generalized estimating equations (GEE) approach as proposed by Liang and Zeger (1986). Note that the z-statistic reported in the original article is the square root of the Wald χ^2-statistic.

Power analysis: The original study had 273 participants after exclusion. The standardized effect size (Cohen's d) was $d = 0.221$. To have 90% power to detect 67% of the original effect size, a sample size of $n = 485$ is required.

Sample: In the original study, seventy-four participants were excluded from failing an attention check question. The replication experiment will implement the same exclusion criteria as in the original study, i.e., (1) exclude participants who fail a general attention check and (2) fail to correctly recall the information learned in the manipulation. We will make sure that participants can only participate once from the same account in this specific study, and we will only recruit participants with a HIT approval rate of 95% or above. We will also check all IP addresses via https://www.ipqualityscore.com/; and we will remove any participants where one or more of the following is true: fraud score >= 85; TOR = True; VPN = True; Bot = True; abuse velocity = high. The replication sample size is the sample size after any exclusions of participants.

Materials: We will use the same material as in the original study, kindly provided by the original authors. However, due to restrictions of the IRB associated with the original study, we cannot use the exact same photographic stimuli. Instead, the original authors recommended stimuli from the Chicago Face Database that capture the important characteristics of the original photos. We obtained these recommended stimuli. The rest of the materials are directly taken from the original study, including the exact wording and the visual presentation of the survey questions.

Procedure: We will closely follow the procedure of the original experiment. The following summary of the experimental procedure is therefore largely based on the description of the experiment in the article (p. 15341).

Participants will first be shown a Captcha, and will thereafter provide informed consent. After this we will include an attention check that participants will need to pass to continue to the study. This attention check is in addition to any other potential attention check(s) used in the original study. Participants will then read that researchers are interested in how people allocate money based on minimal amounts of information. Participants will learn that there are two groups of people, "Niffites" and "Luupites", with either the Niffites or the Luupites being more likely to possess moles, and the other group being more likely to have long fingers. The characteristics will be randomly varied between participants. Participants will be asked to view four photographs of white men in a 2×2 matrix pattern. Two of the men will have moles on their face, the other two will not have moles. Of these, one man with moles and one man without moles are randomly labeled as "Luupites" and the other pair as "Niffites", creating a stereotypical and counterstereotypical stimulus in each group. The exact presentation order of the faces within the matrix will be randomized across participants. Four sliding bars will be presented below the photographs that participants use to make their allocation. The bars are labeled as persons 1–4, and a number corresponding to one of the bars will appear next to each photograph. Participants can use the bars to decide how to make their allocation, and can change their answers until they submit their responses for all four targets. The only constraint is that the total allocation must amount to $30 and that the allocations must be made in whole dollars.

Political ideology will be measured on a 9-point scale ranging from "extremely liberal - 1" to "extremely conservative - 9" with "moderate - 5" at the middle of the scale.

To ensure that participants encode and remember the stereotype information presented to them at the beginning of the study, participants will separately indicate which group is more likely to have moles and which group is more likely to have long fingers. For each question, participants will indicate that the correct answer is either the Luupites, Niffites, neither group, or that they are unsure. Participants who fail to correctly answer both questions will be excluded from analysis, as in the original study. A general attention check question will also be included, prompting participants not to answer a multiple choice scale and enter "yes" in a text box. As in the original study, we will exclude those that answer this question incorrectly.

Following the original study, we will conduct two manipulation checks. The first manipulation check will be considered successful if we can reject the null hypothesis that the mean belief in the stereotype is below or equal to the scale midpoint of 4 using a one-sided one sample t-test (if normality assumption is met) or a Wilcoxon signed-rank test (if assumption of normality is violated). For that, participants will be asked to respond to the question "To what extent does whether a person has moles on their face help to accurately categorize them as belonging to the Luupite or Niffite group?" using a 1 (not at all) to 7 (very much so) scale. We consider the second manipulation check successful if we cannot reject the null hypothesis that the responses do differ as a function of political ideology and stereotype description (i.e., "Niffites" or the "Luupites" possess moles, dummy-coded), or their interaction. For that we will conduct a linear regression and assess the main effects and interaction of political ideology and stereotype description. The second manipulation check will be considered successful if both factors do not contribute significantly to the model. The conclusion about whether the study replicates or not will only be based on the main replication test (i.e., the results of the manipulation checks do not affect this conclusion). The tests of the manipulation checks are mainly relevant for understanding why the study failed to replicate if it should fail to replicate.

Analysis: The analysis will be performed as in the original paper, and the analysis code was kindly provided by the original authors. We will use a generalized estimating equations regression analysis (to account for the dependency in the data) to compare whether conservatives are more or less likely to allocate a higher percentage of money to stereotypical targets vs counterstereotypical targets).[1]

Subject payment: We are standardizing payments across all replications so that studies have a certain show-up fee depending on the expected length of the study, with an hourly wage from the show-up fee of $8 and a minimum payment of $1 (for studies with incentive payment we use the same incentive payment as in the original study; and this payment is paid in addition to the show-up fee). If we have problems recruiting, we will increase the show-up fee.

---

[1] We will only assess the effect of stereotypicality for conservatives, and omit the analysis for liberals (and the interaction effect).