**Replication of Gheorghiu et al. 2017**
**"Facial appearance affects science communication"**
**PNAS 114(23), 5970-5975.**
http://www.pnas.org/content/114/23/5970

*The original paper includes several studies. We randomly chose study 4. In this within-subject experiment, the authors investigate to what extent the competence and attractiveness of scientists influence interest in the scientists' work. The authors vary the attractiveness and competence of the scientists (2x2 design) and participants are asked to imagine that they were browsing a website where scientists describe their research in videos. For each of the four conditions, participants rate how likely they would be to watch the video. Interest is positively related to the facial competence of the scientist.*

**Hypothesis to replicate and bet on:** Participants are more likely to select articles that are paired with photographs of researchers that are associated with high competence. To evaluate this hypothesis, the authors perform a mixed-effects regression as reported in the Table S16 in the SI appendix (the test statistic and *p*-value for this result are not reported in the main text). The *p*-value for this result is $p = 0.032$. The original study does not report a test statistic corresponding to the treatment effect. Based on the original data and analysis scripts that were kindly shared by the original authors, the corresponding test statistic is $t(404) = 2.151$ (using Satterthwaite's correction for degrees of freedom). This particular result was chosen since it was the key result in study 4.

**Criteria for replication:** The criteria for replication are an effect in the same direction as the original study and a *p*-value < 0.05 from a two-sided *t*-test of the coefficient of the competence ratings in a mixed-effects regression.

**Power analysis:** The original study had 408 participants after exclusions who each rated a scientist in each of the four conditions. The replication focuses on the treatment comparison between high and low competence. The standardized effect size (Cohen's *d*) was $d = 0.106$. To have 90% power to detect 67% of the original effect size, a sample size of $n = 2085$ is required.

**Sample:** Only participants from the US over 18 years of age were allowed to participate in the original study. Moreover, workers were required to have a HIT approval rate of 98% or higher, and more than 1,000 HITs approved. Two participants were excluded for reporting technical issues. The replication experiment will implement the same criteria. Moreover, we will make sure that participants can only participate once from the same account in this specific study. 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. In particular, the experiment will be conducted using the original *Qualtrics* survey, including the same titles, photographs, and videos as in the original study.

**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. 5972), the section "Materials and Methods" (p. 5974), and the descriptions provided in section "Stimuli" in the Supplementary Information (pp. 3–4).

Study 4 of the paper builds on the finding that competence and attractiveness are two key predictors of interest judgments in studies 1 and 2, by varying the attractiveness and competence of the scientists in a 2 (attractiveness) × 2 (competence) within-subject design.

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 are asked to imagine that they were browsing a website hosting videos of scientists describing their research. Each trial will present one putative video, comprising one of the four biology article titles with the least-extreme interest preratings from Study 3, paired with a male scientist's photo taken from those scoring in the bottom or top octile on competence and attractiveness in Study 2. Participants are asked to rate how likely they would be to watch the video, completing one trial per cell of the treatment design (i.e., four trials in total). Ratings are on a seven-point scale.

The four article titles are paired with the four cells of the design (low/high attractiveness and low/high competence) using a 4x4 Latin Square; participants are randomly allocated to one of the four versions. One of the two photos with the appropriate attractiveness-competence combination was randomly selected on each trial.

At the end of all studies, participants provide demographic information and complete a questionnaire to measure their engagement with science (e.g., "I am knowledgeable about science," "I find scientific ideas fascinating").

**Analysis:** The analysis will be performed as in the original paper. In particular, interest ratings are regressed on competence, attractiveness, and their interaction, along with participant age, gender, science engagement, and their interactions with the facial traits using a mixed-effects model (using Satterthwaite's correction for degrees of freedom). The replication focuses on the *t*-test of the coefficient of the competence variable in the mixed-effects regression.

**Subject payments:** 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.