Replication of Guilbeault et al. 2018
"Social learning and partisan bias in the interpretation of climate trends"
PNAS 115(39), 9714-9719.
https://www.pnas.org/content/115/39/9714

The original paper contains one experiment. In this between-subject experiment, participants are shown a graph of the amount of Arctic sea ice 1980-2012 and in three rounds asked to predict the amount of ice in 2025. Participants are randomized into one of four treatments in which they receive differing amounts of feedback from their networks before they make their predictions in rounds 2 and 3. The four treatments are (i) a control group of participants with the same political ideology and no feedback from neighbors, and three treatments with networks with both conservatives and liberals where participants see their neighbors' average estimate together with (ii) No Partisan Cues, (iii) Democratic and Republican Logos, and (iv) information about the political identity of each of their neighbors. We focus on the comparison between treatment (ii) No Partisan Cues and treatment (iii) Democratic and Republican Logos. The fraction of participants that predict the correct trend in the climate data is higher in round 3 for the no partisan cues treatment than for the Democratic and Republican Logos treatment.

Hypothesis to replicate and bet on: Partisan priming reduces trend accuracy in networks when predicting climate trends. Trend accuracy is defined as the fraction of participants in a treatment that predicted the correct trend in the data, and the trends are compared in round 3. The authors test this hypothesis in a Mann-Whitney U test with each network of 40 individuals aggregated into one observation and 12 aggregated observations per treatment (Mann-Whitney U test, $n = 24$, $z = 3.207$, $p = 0.0013$; p. 9716). This test was randomly chosen.

Criteria for replication: The criteria for replication are an effect in the same direction as the original study and a p-value < 0.05 in a Mann-Whitney U test.

Power analysis: The original had 960 participants in two treatments, forming 24 aggregated network observations (12 per treatment, where each aggregated observation is based on 40 participants in a network). The standardized effect size (Cohen's d) was $d = 1.309$. To have 90% power to detect 67% of the original effect size, a sample size of $n = 56$ network observations (in total 2240 participants) is required.

Sample: As in the original study, we will make sure that participants can only participate once from the same account in this specific study. In the original study, only subjects who self-identified as conservative or liberal could participate – we will do the same. As the experiment is set in an American political context, we will restrict our HITs to US participants even though this restriction is not mentioned in the original study. We will only recruit participants with a HIT approval rate of 95% or higher. In the analysis of the original paper, all participants who had missing answers in rounds 1 and 3 were excluded from estimating the trend accuracy of each network observation - we will do the same. 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 code the experiment ourselves based on the information provided in the paper and the Supplementary Information, as we do not have access to the original software.

Procedure: We will closely follow the procedure of the original experiment. We will only replicate the No Partisan Cues and Democratic and Republican Logos treatments and not the other treatments included in the original study. The following summary of the experimental procedure is therefore largely based on the description of the experiment in the main text (p. 9715), in the "Materials and methods" section (pp. 9718–9719), and in the supplementary material (pp. 2–7).

Participants will be recruited to play an "Intelligence game" and asked about their self-identified political ideology. Participants who are neither conservative nor liberals will not be invited to the experiment. The remaining participants will then view instructions for the game, provide information about gender and age, and wait for enough subjects to arrive.

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. When enough participants have arrived, they will be randomized into one of two treatments. Participants will then become part of a decentralized social network with 40 participants. Each person will have four neighbors and each network consists of equally many conservatives and liberals. Participants will be presented with a graph showing the average monthly amount of Arctic sea ice 1979-2013 and asked to forecast the amount of Arctic sea ice in 2025. Participants will then provide estimates in three rounds, which will last one minute each. In round 1, participants provide their own estimate. In rounds 2 and 3, participants will be shown the average estimates of their network neighbors along with their previous answer. In the No Partisan Cues treatment, participants will be shown nothing else. In the Democratic and Republican Logos treatment, participants will also be shown the logos of the Democratic and Republican parties when providing estimates in rounds 2 and 3. At the end of the experiment, participants are informed of their accuracy and their payment. The experiment will last for five minutes.

Analysis: The analysis will be performed as in the original article. In particular, the trend accuracy in round 3 will be compared between the No Partisan Cues treatment and the Democratic and Republican Logos treatment with each network of 40 individuals aggregated into one observation using a Mann-Whitney U test (referred to as a Wilcoxon rank sum test in the original article).

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.