Documentation of Relevant Existing Data

Effect Sizes of Replication Studies

The aggregated effect size estimates of the replication studies were estimated as the correlation coefficient *r*, and were retrieved the Social Science Replication Project (SSRP; Camerer et al. 2018) and the Many Labs 2 project (ML2; Klein et al. 2018).

Social Sciences Replication Project (SSRP)
The effect size estimates of Camerer et al. (2018) were retrieved from the dataset called *D3 - ReplicationResults.csv* which is accessible via https://osf.io/abu7k/. Specifically, the effect size estimates are given in column *r_rp*, the corresponding study identifiers were retrieved from https://osf.io/fg4d3/.

Many Labs 2 Project (ML2)
The effect size estimates of Klein et al. (2018) were retrieved from the dataset *Data_Figure_NOweird_means.csv* which was published in https://osf.io/ag2pd/ and can be accessed in the subfolder *!!RawData*. Specifically, the effect size estimates are given in column *loc*, which specifies the location of the aggregated effect size estimates *r* that were used to create in Figure 2 in Klein et al. (2018, p. 470). The corresponding datafile is also accessible in our OSF repository under the name *ML2_ReplicationEffectSizes.csv* (https://osf.io/uagd2/).

Prediction Data of Experts

For both the SSRP and the ML2 project, binary accuracy data for each expert was not available. To create a vector that encodes correct and incorrect predictions, we therefore dichotomized the survey responses on the beliefs in replicability. If an expert indicated for a given study a belief in replicability that was above 50%, we coded this answer as “the study will replicate”. Values that were 50% or lower were coded as “the study will not replicate”. Based on these dichotomous answers, we then determined the rate of correct and incorrect predictions.

Social Sciences Replication Project (SSRP)
The survey beliefs of the experts in Camerer et al. (2018) were retrieved from the dataset *D5 - PreMarketSurvey.csv* which is accessible via https://osf.io/6cu54/. We chose to extract the survey beliefs from experts that were assigned to “Treatment 2” (more details on the different treatment conditions can be found in Camerer et al. (2018, Supplements, p. 27, section “Elicitation of peer beliefs about replicability”). In this treatment condition, experts were asked to assess for each replication study the likelihood that the hypothesis would be replicated in stage

1 or stage 2 of the replication project, or in neither. To reflect the belief in replicability, we reversed the rating for the “neither” response (we were not interested in the different stages). The corresponding column is denoted as *m3_b3*. The corresponding data is also accessible in our OSF repository under the name *SSRP_ExpertPredictions.csv* (https://osf.io/jqd2n/).

Many Labs 2 Project (ML2)
We retrieved the survey beliefs of the experts for the ML2 project (Forsell et al., 2018) from the dataset *ML2_Survey_a.csv* which was published at https://osf.io/uhmj4/. The relevant columns were *Q1A:Q28A.* We excluded 4 studies that were included in the ML2 project, but not in the prediction market and survey. The corresponding datafile is also accessible in our OSF repository under the name *ML2_ExpertPredictions.csv* (https://osf.io/uagd2/).

Replication Outcomes

Social Sciences Replication Project (SSRP)
The replication outcomes in Camerer et al. (2018) were retrieved from the dataset *D3 - ReplicationResults.csv* which is accessible via https://osf.io/abu7k/. The corresponding column is denoted as *rep_sr_rp*. The corresponding data is also accessible in our OSF repository under the name *SSRP_ReplicationOutcomes.csv* (https://osf.io/kxrm3/).

Many Labs 2 Project (ML2)
The replication outcomes for the ML2 project were retrieved fromTable 5 in Klein et al. (2018). The relevant column is *replication sample size, $p < .05$.* The corresponding datafile is also accessible in our OSF repository under the name *ML2_ReplicationOutcomes.csv* (https://osf.io/nqzjc/).

References

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating replicability of social science experiments in *Nature* and *Science*. *Nature Human Behaviour*, *2*, 637–644.

Forsell, E., Viganola, D., Pfeiffer, T., Almenberg, J., Wilson, B., Chen, Y., ... & Dreber, A. (2018). Predicting replication outcomes in the Many Labs 2 study. *Journal of Economic Psychology*.

Klein, R., Vianello, M., Hasselman, F., Adams, B., Adams, R., Alper, S., . . . Nosek, B. (2018). Many Labs 2: Investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science*, *1*, 443–490.