**Supplementary materials**

**Supplementary materials 1**

**Replication project Extraction and exclusion details**

Open Science Collaboration (2015)

Three original studies which did not report that their findings were indicative of a non-zero effect were excluded from those studies extracted from Open Science Collaboration (2015). Three studies for which z transformed correlation coefficients could not be extracted due to missing data in the downloaded data set were also excluded from analysis. Effect sizes for original and replication studies are included for 94 out of 97 studies replicated studies from Open Science Collaboration (2015) which reported having found a non-zero effect.

Camerer et al. (2018)

Original and replication effect sizes were extracted for all 21 studies included in Camerer et al. (2018). In some cases in the Nature Science reproducibility project (Camerer et al., 2018) multiple replication studies were performed for a single effect. In each of these cases we performed a fixed effects meta-analysis using the metafor package (Viechtbauer, 2010) to estimate a meta-analytic effect size estimate. The effect size, standard errors and sample sizes used in the current study reflect this pooled estimate. This method leads to one study more “replicating” according to the ‘statistical significance in the same direction of the original study’ criterion than was originally reported in Camerer et al. (2018), where they used the p value from the largest performed study instead of a pooled estimate.

Soto (in press)

Effect sizes were extracted for original and replication studies for 101 out of 121 included studies, and one original study’s sample size was not available. In Soto (in press) Effect sizes which were only reported in this dataset as beta coefficients were not converted to Fisher z scores as not enough information was available in the data set. A total of 100 of 121 effects were included in the current analysis. As some replication studies used shorter form versions of the original data collection instruments, all results presented have been disattenuated using the Spearman-Brown prediction formula and Spearman disattenuation formula to estimate the trait-outcome associations that would be expected if our outcome measure had used the same number of items as the original study (Lord & Novick, 1968). Following the other large scale replication studies, the signs of the original and replication study effects were switched if the original effect was negative.

Cova et al. (2018)

Cova et al. (2018) included three replications of original studies which were non-significant (and which did not claim to provide evidence for the effects under test), these were removed from analysis. Effect sizes were reported by Cova et al. (2018) and are included in the current study for 33, original and replication studies, out of an original 37 replicated studies with significant original results. The four studies for which no effect sizes were reported performed analyses for which Cova et al. (2018) could not develop reasonable effect size estimates (e.g., a Sobel test, GEE analysis).

Many labs 1 R. Klein et al. (2014)

Many labs 1 (R. Klein et al., 2014) examined whether effects from 13 original papers replicated, one of which did not report an effect size or test statistic so is not included in the current sample. No effect size was extractable for one original study, and this effect was excluded for the purposes of the current analysis. Four different operationalisations of anchoring effects were tested, all of which are included in the current analysis, leading to a total of 15 paired data-points being included from this study. The multilevel models reported below accounts for non-independence between effects by including a random effect for study.

Many labs 2 R. A. Klein et al. (2018)

A total of 22 of 28 paired original and replication effects sizes are included for this analysis. Four studies from (R. A. Klein et al., 2018) were removed because the original and replication studies examined a difference in effect sizes seen in different conditions, and the effects were not directly tested against each other making it difficult to derive an appropriate effect size. Two additional studies were excluded because their effect sizes were only available as Cohen’s q.

Ebersole et al. (2016)

Original and replication effect sizes were extracted for all 9 original and replication studies from (Ebersole et al., 2016), excluding a study they term a “conceptual replication”. Most effects (6/9) were converted to correlation coefficients from the Cohen’s d values reported in this replication project. The results of three additional studies reported as partial Eta squared were converted to correlation coefficients from F statistics using the formula:

Camerer et al. (2016)

The economics replication project (Camerer et al., 2016). Original and replication effect sizes for all 18 studies were reported in correlation coefficients and all are included in this analysis.

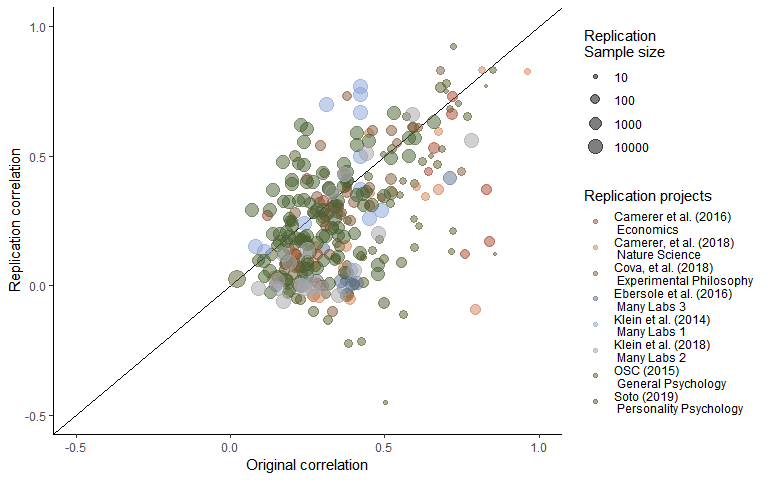
**Supplementary Material 2**

**Plots and multilevel model output of the relationship between original and replication correlation coefficients using varied exclusion criteria**

The following output shows scatter plots and model output for all of the multilevel meta-analyses performed using the varied exclusion criteria explained in the main text.

Table SM 1. Multilevel meta-analysis model estimates and random effects for all data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Estimate | 95% CI LB | 95% CI UB | SE | p | Random effects |
| -0.14 | -0.21 | -0.07 | 0.04 | < .001 |  |
|  |  |  |  |  | Project variance = 0.008, n = 8 |
|  |  |  |  |  | Article variance = 0.016, n = 229 |
|  |  |  |  |  | Effect variance = 0.012, n = 306 |
|  |  |  |  |  | QE(305) = 3531.9, p < .001 |

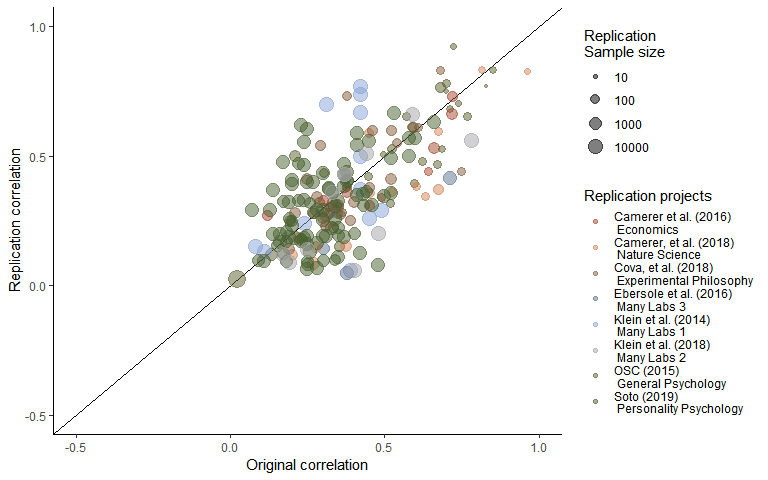


*Figure SM1.* Scatter plot of replication effect sizes (in correlation coefficients) plotted against original effects including all data.

Table SM2.

*Multilevel meta-analysis model estimates and random effects including only statistically significant replications.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Estimate | 95% CI LB | 95% CI UB | SE | p | Random effects |
| -0.05 | -0.11 | 0.01 | 0.03 | 0.1 |  |
|  |  |  |  |  | Project variance = 0.005, n = 8 |
|  |  |  |  |  | Article variance = 0.014, n = 132 |
|  |  |  |  |  | Effect variance = 0.009, n = 198 |
|  |  |  |  |  | QE(197) = 2715.24, p < .001 |

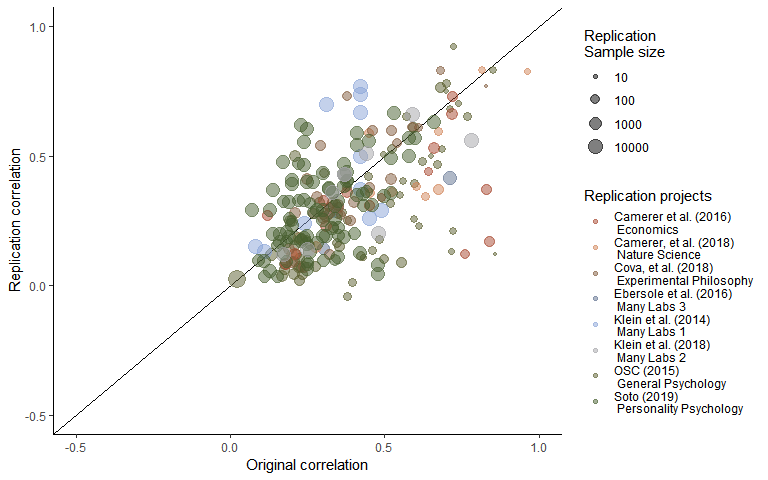


*Figure SM2.* Scatter plot of replication effect sizes (in correlation coefficients) plotted against original effects including only statistically significant replications.

Table SM3.

*Multilevel meta-analysis model estimates and random effects including studies which are not statistically equivalent to the null, using equivalence bounds set as the minimum effect size that would have been statistically significant in the original study.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Estimate | 95% CI LB | 95% CI UB | SE | p | Random effects |
| -0.08 | -0.15 | -0.01 | 0.04 | 0.03 |  |
|  |  |  |  |  | Project variance = 0.008, n = 8 |
|  |  |  |  |  | Article variance = 0.018, n = 169 |
|  |  |  |  |  | Effect variance = 0.009, n = 237 |
|  |  |  |  |  | QE(236) = 3031.58, p < .001 |



*Figure SM3.* Scatter plot of replication effect sizes (in correlation coefficients) plotted against original effects including studies which are not statistically equivalent to the null, using equivalence bounds set as the minimum effect size that would have been statistically significant in the original study.

**Supplementary materials 3**

**Leave one out cross validation output**

Table [SM4](file:///C:\Users\fsingletonthorn\Documents\PhD\Effect%20size%20adjustment%20testing%20paper\SimplifiedEffectSizeAdjustment%2029%20Jan%202018.docx#loo-cross-validation-output).

*0th, 25th, 50th, 75th and 100th percentiles from leave one out cross validation for each multilevel model, for each exclusion method an, including only the sample indicated in “LOO exclusions”.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| LOO exclusions | Subsample | Proportion significant | Minimum estimate | 25th percentile | Median | 75th percentile | Maximum estimate |
| Replication project | Only non-equivalent replications | 0.38 | -0.10 | -0.09 | -0.08 | -0.07 | -0.06 |
| Replication project | Only significant replications | 0.12 | -0.07 | -0.06 | -0.06 | -0.04 | -0.03 |
| Replication project | P value as Moderator | 1.00 | -0.10 | -0.09 | -0.07 | -0.07 | -0.06 |
| Replication project | All data | 1.00 | -0.16 | -0.15 | -0.13 | -0.13 | -0.12 |
| Study | All data | 1.00 | -0.14 | -0.14 | -0.14 | -0.14 | -0.13 |
| Study | Only significant replications | 0.01 | -0.06 | -0.05 | -0.05 | -0.05 | -0.04 |
| Study | Only non-equivalent replications | 1.00 | -0.09 | -0.08 | -0.08 | -0.08 | -0.07 |
| Study | P value as Moderator | 1.00 | -0.09 | -0.08 | -0.08 | -0.08 | -0.07 |
| Effect | Only significant replications | 0.00 | -0.06 | -0.05 | -0.05 | -0.05 | -0.04 |
| Effect | Only non-equivalent replications | 1.00 | -0.09 | -0.08 | -0.08 | -0.08 | -0.07 |
| Effect | P value as Moderator | 1.00 | -0.09 | -0.08 | -0.08 | -0.08 | -0.07 |
| Effect | All data | 1.00 | -0.14 | -0.14 | -0.14 | -0.14 | -0.13 |

**Supplementary materials 4**

**Bayesian Mixture Model**

The mixture model results presented in text presents the model developed by Camerer et al,. (2018; see <https://osf.io/xhj4d/> for their detailed description of this model). All priors were chosen to be uninformative or vague. The mixture model assumes that the observed replication effect sizes either come from the null hypothesis, a true effect sampled from a normal distribution with a mean of zero and a estimated precision (tau). This model uses an errors-in-variables approach to account for possible attenuation of effect sizes due to measurement error and estimation uncertainty following (Matzke et al., 2017), which means the effect size attenuation factor is the factor change between the estimated true effect of the original and replication study effect size. Although this may be reasonable in that the true effect size of the effect may not be the true effect size of a particular study and analysis set up, this poses an interpretative problem in that alpha now represents the difference between the estimated original effect and the replication effect.

Box SM1. The original model reported in (Camerer et al., 2018) and reported on in the main text of the current article.

model{# Mixture Model Priors:alpha ~ dunif(0,1) # flat prior on slope for predicted effect size under H1tau ~ dgamma(0.001,0.001) # vague prior on study precisionphi ~ dbeta(1, 1) # flat prior on the true effect rate# prior on true effect size of original studies:for (i in 1:n){trueOrgEffect[i] ~ dnorm(0, 1)}# Mixture Model Likelihood:for(i in 1:n){clust[i] ~ dbern(phi)# extract errors in variables (FT stands for Fisher-transformed):orgEffect\_FT[i] ~ dnorm(trueOrgEffect[i], orgTau[i])repEffect\_FT[i] ~ dnorm(trueRepEffect[i], repTau[i])trueRepEffect[i] ~ dnorm(mu[i], tau)# if clust[i] = 0 then H0 is true; if clust[i] = 1 then H1 is true and# the replication effect is a function of the original effect:mu[i] <- alpha \* trueOrgEffect[i] \* equals(clust[i], 1)# when clust[i] = 0, then mu[i] = 0;# when clust[i] = 1, then mu[i] = alpha \* trueOrgEffect[i] }}

**Supplementary materials 5**

**Conversions**

All statistical tests extracted were transformed into correlation coefficients as follows, using the methods reported in (Open Science Collaboration, 2015).

t statistics:

Where is the observed t statistic and is the degrees of freedom of the t test.

F statistics:

Where is the observed F statistic and is the degrees of freedom of the numerator and is degrees of freedom of the denominator.

Chi square statistics:

Where is the observed statistic and is the associated degrees of freedom.

All values were then transformed into fisher Z transformed correlation coefficients using: