***Test for Excess Significance.*** Another power-based test for publication bias is the Test for Excess Significance (Ioannidis & Trikalinos, 2007). This test estimates the number of expected studies with statistical significance given some anticipated effect size (usually the naïve meta-analytic estimate), then compares that expectation against the number of observed significant results. A significant test suggests censoring of nonsignificant results or the manipulation of results into statistical significance.

This test has a number of weaknesses that prevent its inclusion in the main text. It has poor statistical power (Ioannidis & Trikalinos, 2007), and the validity of its *p*-value rests on strong, perhaps unwarranted assumptions about researcher behavior (Morey, 2013). Like other tests for bias, its results may be spurious when there exists genuine between-study heterogeneity. The reader is urged to interpret these results with considerable caution.

The results are summarized in Table S1. TES indicates a significance of excess significance in the full sample of experiments of aggressive behavior and its best-practices subsample. This test is also significant in the full sample of experiments of aggressive affect and near the significance threshold in the best-practices subsample.

***Multiplicative-error meta-regression models.*** Random-effects meta-analysis models heterogeneity across studies through the use of an additive error term τ2. An alternative strategy for handling heterogeneity in meta-regression is instead to fit models with a multiplicative error term φ. (see, e.g., Moreno et al., 2009).

We fit additional multiplicative-error models. The results are summarized in Table S2. The most significant change is that the PET estimate for best-practices aggressive behavior reaches statistical significance in these models, and the PEESE estimate increases to *r* = .17.

**Checks for confounds with small-study effects.** One concern is that some other moderator may be confounded with sample size in such a way as to cause a small-study effect that is not due to publication bias. Correlation tables for study moderators in experiments of aggressive affect, behavior, and cognition are provided in Tables S3, S4, and S5, respectively.

**Sensitivity analyses.** Some data points appeared to be outliers, having unusually high effect sizes and thus considerable influence on the meta-analytic adjustments. One had sufficient influence that the results bear consideration after its exclusion.

In experiments of aggressive affect, one best-practices study (Ballard & Wiest, 1996) appeared to be an outlier, having a very large effect size (*r* = .87) and low precision (*N =* 30). Exclusion of this outlier brought the estimators into greater agreement, reducing the naïve estimates (*r* = .27, fixed- and random-effects, *I2* = 0.01, [0.00, 62.8]), the *p*-uniform estimate (*r* = .20), and the *p*-curve estimate (*r* = .19), while increasing the PET (*r* = -.01, *I2* = 0.00, [NA, NA]) and PEESE (*r* = .17, *I2* = 0.00, [NA, NA]) estimates. We think that the exclusion of this outlier is therefore a good idea, as it prevents PET from returning a negative effect size. Again, these null-set confidence intervals on *I2* indicate unusual homogeneity of residuals after adjusting for small-study effects.

Table S1. Results of the Test for Excess Significance.

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome | Setting | Best | TES.pval |
| Affect | Experiment | Best-only | .079 |
| Affect | Experiment | All | **.022** |
| Affect | Cross-Section | Best-only | - |
| Affect | Cross-Section | All | .211 |
| Behavior | Experiment | Best-only | **.015** |
| Behavior | Experiment | All | **.020** |
| Behavior | Cross-Section | Best-only | .785 |
| Behavior | Cross-Section | All | .309 |
| Cognition | Experiment | Best-only | .309 |
| Cognition | Experiment | All | .339 |
| Cognition | Cross-Section | Best-only | .994 |
| Cognition | Cross-Section | All | .581 |
| Arousal | Experiment | Best-only | .636 |
| Arousal | Experiment | All | .704 |

Table S2. Naïve, PET, and PEESE estimates, with Egger tests, in fixed-effect dispersion models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Setting | Group | Naïve | *p(Egger)* | PET | PEESE |
| Aggressive Affect | | | | | |
| Experiment | Best | .33 [.23, .42] | .001 | -.10 [-.32, .14] | .18 [.05, .30] |
| Experiment | Full | .21 [.13, .28] | .002 | -.10 [-.29, .09] | .08 [-.02, .18] |
| Cross-Section | Best | - | - | - | - |
| Cross-Section | Full | .16 [.08, .23] | .761 | .13 [-.05, .30] | .15 [.04, .26] |
| Aggressive Behavior | | | | | |
| Experiment | Best | .23 [.20, .27] | .001 | .10 [.02, .17] | .17 [.12, .22] |
| Experiment | Full | .18 [.13, .22] | .248 | .12 [.01, .23] | .15 [.08, .21] |
| Cross-Section | Best | .30 [.25, .34] | .723 | .28 [.16, .39] | .29 [.23, .36] |
| Cross-Section | Full | .23 [.20, .27] | .151 | .19 [.11, .26] | .22 [.18, .27] |
| Aggressive Cognition | | | | | |
| Experiment | Best | .23 [.18, .28] | .024 | .09 [-.03, .21] | .18 [.12, .24] |
| Experiment | Full | .18 [.13, .22] | .036 | .09 [-.02, .20] | .15 [.09, .21] |
| Cross-Section | Best | .21 [.15, .28] | .950 | .22 [.05, .38] | .22 [.13, .31] |
| Cross-Section | Full | .20 [.14, .26] | .480 | .16 [.01, .29] | .19 [.11, .27] |
| Physiological Arousal | | | | | |
| Experiment | Best | .21 [.11, .30] | .881 | .19 [-.09, .44] | .21 [.05, .35] |
| Experiment | Full | .16 [.09, .22] | .022 | -.03 [-.18, .13] | .09 [.01, .16] |

Table S3. Correlations between features of experimental effect sizes on aggressive affect.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sample.size | Fisher.s.Z | best | age | person | east | proceeding | diss |
| Sample.size | 1 | -0.437 | -0.227 | 0.127 | 0.449 | -0.197 | -0.096 | -0.281 |
| Fisher.s.Z | -0.437 | 1 | 0.568 | 0.268 | -0.092 | 0.159 | 0.113 | -0.373 |
| best | -0.227 | 0.568 | 1 | 0.249 | -0.104 | 0.236 | 0.164 | -0.503 |
| age | 0.127 | 0.268 | 0.249 | 1 | 0.197 | 0.127 | 0.089 | -0.354 |
| person | 0.449 | -0.092 | -0.104 | 0.197 | 1 | -0.173 | -0.12 | -0.415 |
| east | -0.197 | 0.159 | 0.236 | 0.127 | -0.173 | 1 | 0.696 | -0.15 |
| proceeding | -0.096 | 0.113 | 0.164 | 0.089 | -0.12 | 0.696 | 1 | -0.104 |
| diss | -0.281 | -0.373 | -0.503 | -0.354 | -0.415 | -0.15 | -0.104 | 1 |

*Note:* “Best” is coded 1 for best-practices, 0 for not-best-practices. “Age” is coded 1 for adults, 0 for children, NA for combined adult/child samples. “Person” is coded 1 for first-person games, 0 otherwise. “East” is coded 1 for Eastern samples, 0 for Western samples. “Proceeding” is coded 1 for conference proceedings, 0 otherwise. “Diss” is coded 1 for unpublished dissertations, 0 otherwise.

Table S4. Correlations between features of experimental effect sizes on aggressive behavior.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sample.size | Fisher.s.Z | best | age | person | east | proceeding | diss |
| Sample.size | 1 | -0.119 | 0.271 | 0.265 | 0.021 | -0.301 | -0.193 | -0.089 |
| Fisher.s.Z | -0.119 | 1 | 0.444 | 0.229 | 0.202 | 0.145 | 0.265 | -0.388 |
| best | 0.271 | 0.444 | 1 | 0.396 | 0.175 | 0.036 | 0.11 | -0.511 |
| age | 0.265 | 0.229 | 0.396 | 1 | 0.279 | 0.187 | 0.219 | -0.201 |
| person | 0.021 | 0.202 | 0.175 | 0.279 | 1 | -0.036 | 0.164 | -0.267 |
| east | -0.301 | 0.145 | 0.036 | 0.187 | -0.036 | 1 | 0.665 | -0.217 |
| proceeding | -0.193 | 0.265 | 0.11 | 0.219 | 0.164 | 0.665 | 1 | -0.144 |
| diss | -0.089 | -0.388 | -0.511 | -0.201 | -0.267 | -0.217 | -0.144 | 1 |

*Note:* “Best” is coded 1 for best-practices, 0 for not-best-practices. “Age” is coded 1 for adults, 0 for children, NA for combined adult/child samples. “Person” is coded 1 for first-person games, 0 otherwise. “East” is coded 1 for Eastern samples, 0 for Western samples. “Proceeding” is coded 1 for conference proceedings, 0 otherwise. “Diss” is coded 1 for unpublished dissertations, 0 otherwise.

Table S5. Correlations between features of experimental effect sizes on aggressive cognition.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sample.size | Fisher.s.Z | best | age | person | east | proceeding | diss |
| Sample.size | 1 | -0.253 | 0.351 | 0.313 | 0.151 | -0.33 | -0.27 | -0.384 |
| Fisher.s.Z | -0.253 | 1 | 0.199 | 0.283 | 0.2 | 0.288 | 0.039 | -0.296 |
| best | 0.351 | 0.199 | 1 | 0.372 | 0.312 | -0.068 | 0.039 | -0.485 |
| age | 0.313 | 0.283 | 0.372 | 1 | 0.275 | 0.14 | 0.12 | -0.315 |
| person | 0.151 | 0.2 | 0.312 | 0.275 | 1 | -0.036 | 0.021 | -0.327 |
| east | -0.33 | 0.288 | -0.068 | 0.14 | -0.036 | 1 | 0.854 | -0.167 |
| proceeding | -0.27 | 0.039 | 0.039 | 0.12 | 0.021 | 0.854 | 1 | -0.142 |
| diss | -0.384 | -0.296 | -0.485 | -0.315 | -0.327 | -0.167 | -0.142 | 1 |

*Note:* “Best” is coded 1 for best-practices, 0 for not-best-practices. “Age” is coded 1 for adults, 0 for children, NA for combined adult/child samples. “Person” is coded 1 for first-person games, 0 otherwise. “East” is coded 1 for Eastern samples, 0 for Western samples. “Proceeding” is coded 1 for conference proceedings, 0 otherwise. “Diss” is coded 1 for unpublished dissertations, 0 otherwise.

References

Ballard, M. E., & Wiest, J. R. (1996). Mortal Kombat: The effects of violent video game play on males’ hostility and cardiovascular responding. Journal of Applied Social Psychology, 26, 717-730. DOI:10.1111/j.1559-1816.1996.tb02740.x

Ioannidis, J. P. A., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*, 245-253. DOI:10.1177/1740774507079441

Morey, R. D. (2013). The consistency test does not - and cannot - deliver what is advertised: A comment on Francis (2013). *Journal of Mathematical Psychology*, *57* (5), 180 - 183. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022249613000291