Reviewer 1:

We ask Reviewer 1 to reconsider. The PEESE estimate Reviewer 1 considers is the single most optimistic estimator of the effect and is known to be upwardly biased when the null is true. Even so, Reviewer 1 may find it important that a decrease from *r* = .21 to *r* = .15 implies a twofold loss of explanatory power (R­2 falls from 4.4% to 2.3%) and the need for a twofold increase in sample size (80% one-tailed power at 136 vs. 270).

We recognize that the current findings would be inconsistent with contemporary theories of human aggression. However, to discard the evidence because it does not support the current theory would seem to be an unfortunate inversion of the scientific method.

It was not our intention to argue that the effects are necessarily minimal; rather, we meant to say that the adjusted effect size estimates are close to zero. We do not mean to argue that this is strong evidence that there is no effect. Instead, we interpret this as a sign of uncertainty and the need for a strong preregistered replication effort. We thank Reviewer 1 for pointing out the inconsistencies in our language and have made appropriate edits.

Reviewer 2:

Reviewer 2 made a number of valuable suggestions where our rhetoric grew sloppy, e.g., in our unqualified claim that PET, PEESE, and *p*-curve “provide better adjustments” for publication bias. We have softened such claims and used more precise rhetoric as advised.

We have elaborated a bit more on the problem of outcome-switching, as requested. Indeed, the flexible quantification of the Competitive Reaction Time Task (see Elson et al., 2014) may constitute a form of outcome switching.

TODO: STUFF ABOUT TRIM-AND-FILL

We have added a section cautioning the reader as to the potential influence of heterogeneity in meta-analysis in general and the degradation of trim-and-fill results in particular. We caution the reader that such heterogeneity can also influence conclusions drawn from meta-regression and p-curve techniques.

TODO: VEVEA & HEDGES 1995

TODO: USE RANDOM EFFECTS FOR EVERYTHING, EXPLICATE RE MODEL (E.G. “REML”), REPORT I2

Reviewer 3:

Reviewer 3 similarly indicated a number of places where our rhetoric was too strident. We thank Reviewer 3 for these constructive critiques and have amended our language as appropriate.

In our analysis, we ignored the partial effect sizes as collected by Anderson et al. We were most interested in the effects in experiments, for which the “best partials” estimates are the same as the “best raw” estimates in Anderson et al. (2010). The “best partials” in cross-sectional studies are difficult to interpret and the source of much controversy--see, e.g., Ferguson (2015), Rothstein & Bushman, (2015). We thought it best not to touch these. We now explicate this decision in the text so that it is clear we only analyze the raw effect sizes.

We wish to point out some reasons why we think it unlikely that the funnel plot asymmetry is caused by a combination of heterogeneity and *a priori* power analysis. First, we understand the use of power analysis to be quite uncommon in 2010 and earlier. Reforms that placed an emphasis on power calculation did not happen until some years later, e.g. Simmons, Nelson, & Simonsohn’s (2012) 21-word solution. Second, we doubt researchers could have known the causes of heterogeneity in effect size. The Anderson et al. (2010) meta-analysis looked for many suspected moderators of the effect (e.g., sex, age, game perspective) and found none. Additionally, if Anderson knew what would permit acceptable statistical power at both N = 515 (Anderson, Gentile, & Buckley, 2007) and N = 39 (Bartholow & Anderson, 2002), then we expect he would have mentioned such moderators in the 2010 meta-analysis. Finally, whereas this sort of power analysis will create an asymmetrical funnel plot, it will still lead to a right-skewed *p*-curve, as *p*-curve is a function of statistical power alone. We have uploaded a simulation demonstrating this to the OSF repo at <https://osf.io/y2jc6/>.

Reviewer 4:

We were previously using the weighted regression, unaware of the distinction between the Sterne & Egger (2005) method and the Egger 1997 method. Thus, the results have not changed, but the citation has.

Per the reviewer’s suggestions, we have added p-uniform and the Test for Excess Significance. We have also switched to zero-centered, contour-enhanced funnel plots to better illustrate the p = .05 threshold.

We are deeply sympathetic to the reviewer’s request for data. We feel that it is very important that the data be publicly available so that readers and reviewers can better scrutinize our analyses and results. However, Dr. Anderson made it clear when he sent us his data that we did not have his permission to share it with anybody else. Prior to our original submission, we emailed him to request permission to add the data to the online repository, but he has yet to reply. We are reluctant to hector Dr. Anderson any further ourselves, but suggest Reviewer 4 might contact him directly.