# Errata

## On Coppock and Green (2015), *Political Science Research and Methods* 3(1)

This note lists six misstated claims in "Assessing the Correspondence between Experimental Results Obtained in the Lab and Field: A Review of Recent Social Science Research," published in *Political Science Research and Methods* 3(1), 113–131, doi 10.1017/psrm.2014.10. **None of them changes a conclusion of the article.** The rank-order correlation of 0.73 that the abstract leads with is unaffected, and the disattenuated correlation still exceeds the raw one, so the finding that lab-field correspondence is strong stands. Each item is recorded because a reader who quoted the sentence would be quoting a claim the article's own data do not support.

The first three concern one calculation and share one cause. The last three are descriptions of the robustness exercise behind Figure 3, whose numbers are themselves correct.

Every corrected value below is computed from the deposited data at the time this document is rendered; bold marks the tokens that change. The code and output supporting each correction are at github.com/active-maintenance-aec/coppock_green_2015.

### 1. The disattenuated correlation

> The disattenuated correlation is found to be $0.64/\sqrt{0.78 * 0.82} = 0.80$, further attesting to the strength of lab-field correspondence in published work.

Corrected:

> The disattenuated correlation is found to be $\mathbf{0.65}/\sqrt{\mathbf{0.88 * 0.90}} = \mathbf{0.73}$, further attesting to the strength of lab-field correspondence in published work.

Three separate errors move this number. The two reliabilities are squared, which item 3 sets out. The raw Pearson correlation of 0.64 is 0.65 once the Gneezy and Rustichini (2004) female treatment effect is standardized by its own control group rather than by a male treatment group, as Equation 1 of the article requires. And the List (2006) field standard error, which enters the reliability simulation, is read from the lab vector in the deposited script.

## 2. The two reliabilities

> The reliabilities for the lab and field treatment effects are 0.78 and 0.82, respectively.

Corrected:

> The reliabilities for the lab and field treatment effects are **0.88** and **0.90**, respectively.

A correlation near 0.9 is much larger than its own square, which is why correcting the definition in item 3 moves these two values so far. Keeping the squared definition and fixing only the indexing gives 0.78 and 0.82, which is why the published pair is so close to what the corrected code produces under the article's own definition.

## 3. The definition of the reliability

> The estimated reliability is the square of the correlation between two treatment effect draws.

Corrected:

> The estimated reliability is **the correlation** between two treatment effect draws.

The reliability of a measure is the correlation between two parallel measurements of it, and Spearman's correction for attenuation divides the observed correlation by the geometric mean of the two reliabilities. Squaring the correlations first makes the denominator too small and the corrected estimate too large. The deposited script implements the footnote exactly as written, so the article and its code agree with each other and both depart from the correction they invoke. The error was identified by David Broockman (UC Berkeley).

## 4. The median correlation across permutations

> Unsurprisingly, the median correlation stays constant across all permutations.

Corrected:

> Unsurprisingly, the median correlation **stays between 0.72 and 0.74** across all permutations.

Figure 3 itself is right and is not reprinted here. Its median series takes 3 distinct values at two decimal places across the 21 combination sizes, running from 0.719 to 0.739. The deposited code gives the same answer at its own random draw, from 0.720 to 0.729, so this is a description that neither run supports rather than a number the pipeline has moved.

## 5. The note to Figure 3

> *Note*: we calculate the rank-order correlations of the 21 treatment effect pairs included in the meta-analysis, under all possible combinations duplicating or removing any 1, 2, 3… or 10 of them.

Corrected:

> *Note*: we calculate the rank-order correlations of the 21 treatment effect pairs included in the meta-analysis, under **all possible combinations, sampled down to 50,000 wherever more exist,** duplicating or removing any 1, 2, 3… or 10 of them.

The note and footnote 6 contradict each other, and footnote 6 is the one the code follows: 555,790 of the 2,097,150 available combinations were evaluated. Again the figure is right; only its note overstates what was computed.

## 6. The subset size at which sampling begins

> When dropping more than any six pairs, we randomly sampled 50,000 of the possible combinations.

Corrected:

> When dropping more than any **five** pairs, we randomly sampled 50,000 of the possible combinations.

Dropping six of the 21 pairs leaves 54,264 combinations, which is already above the cap, so the sampling applies there too; dropping five leaves 20,349, which is below it. The deposited code applies the same cap by the same test, so this is a description of the code rather than a change to it. The footnote is also silent on the duplication side of the exercise, where the cap binds at exactly the same point.