

A 21 Word Solution

By Joe Simmons, Leif Nelson and Uri Simonsohn

About a year ago we published "[False-Positive Psychology](#)," in which we argued that scientific journals should require authors to disclose how data were collected and analyzed. Here we propose a simple 21-word statement that achieves this, and then reflect on reservations people have expressed regarding disclosure.

"False-Positive" has received much more attention than we ever expected, but this attention has not yet materialized into concrete action designed to reduce the share of false-positives in the literature. We write this article hoping to move the discussion towards action.

1. If you are not [p-hacking](#) and you know it, clap your hands.

Many support our call for transparency, and agree that researchers should fully disclose details of data collection and analysis. Many do not agree. What follows is a message for the former; we begin by preaching to the choir.

Choir: There is no need to wait for everyone to catch-up with your desire for a more transparent science. If *you* did not *p-hack* a finding, *say it*, and your results will be evaluated with the greater confidence they deserve.

If you determined sample size in advance, *say it.*
If you did not drop any variables, *say it.*
If you did not drop any conditions, *say it.*

These 21 words in a Methods section can *say it* succinctly:

"We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study."

When needed, supplemental materials can be used to ensure the 21 words are accurate.

When sample size is not determined in advance, one could write:

"We added 50 observations after analyzing the first 100".

A small but energetic choir may get the entire congregation to sing along.

2. Big-picture skeptics

Three big-picture concerns have been raised:

- a) Trust: Requirements assume or create lack of trust among peers
- b) Policing: Enforcement turns journals into the police (not Sting's)
- c) Effectiveness: Disclosure won't do anything

Our reactions:

a) Trust.

For trust to exist, people must agree on what it is they are trusting. We cannot “trust” our colleagues to run and report their studies “properly” if there is no shared understanding of what “properly” is.

There is no shared understanding of what “properly” is.

Some think dropping conditions is fine, others do not. Some think collecting 10 subjects at a time is fine, others do not. Some think dropping measures is fine, others do not. And so on.

Note how disclosure depends on trust.

If Faith says she collected only one dependent variable we will believe her.

But if she does not say anything, what should we think?

Because many forms of *p*-hacking are common, currently acceptable, and often encouraged by journals, we should wonder, in the absence of disclosure, whether Faith actually collected more measures.

Note how lack of disclosure fosters distrust.

Asking authors to disclose does not take trust out of our scholarly exchanges, it creates a framework for trust to meaningfully exist.

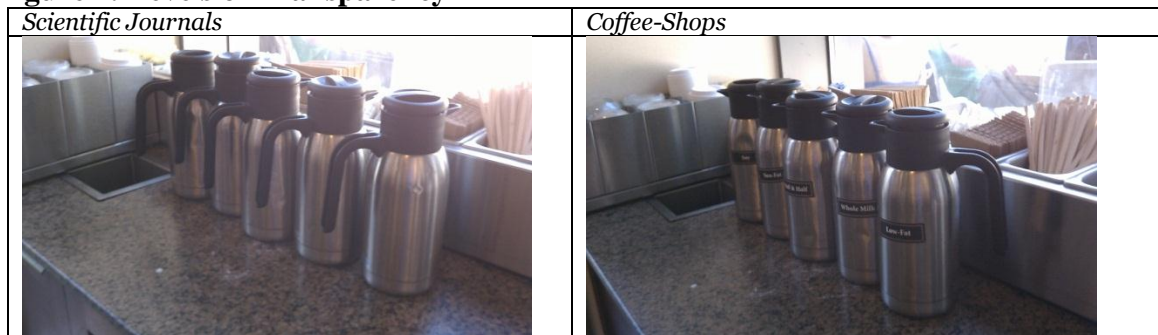
b) Policing.

At coffee-shops we can freely ask, “hey, is this 1% or 2% milk?”

Our aspiration is that our journals achieve this coffee-shop grade level of transparency, so that we can freely ask when reading papers “hey, is this a 1 or 2 dependent variable study?”

The Starbucks barista does not pull out a milk-fat-assessment instrument when asked, he just reads the carton. He trusts the carton to be truthful. Journals, similarly, would merely ask authors to indicate the *p*-hack content in their research and trust them to be truthful. No policing, just asking.

Figure 1. Levels of Transparency



Note: pictures taken at a Peet’s coffee, not Starbucks.

c) Effectiveness.

If you think about it, the only way in which disclosure does not reduce the negative impact of *p*-hacking, does not reduce the likelihood that false-positives will be published in our journals, is if authors *lie* when they disclose. They explicitly state their study had three conditions when in fact it had five. We don't think that is very likely to happen, so we think disclosure is very likely to work.

Those who disagree must not trust psychologists as much as we do.

3. Small-picture skeptics.

Two small-picture concerns that have been raised are:

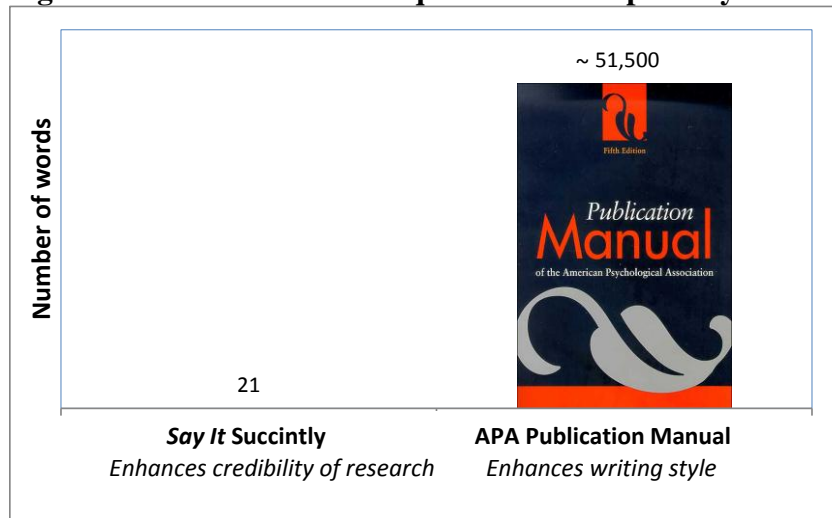
- a) Red-tape: Disclosure requirements add excessive regulation to the publication process
- b) Exceptions: What if once in a while disclosure is impractical?

Our reactions:

a) Red-Tape.

The disclosure requirements we proposed in “False-Positive” fit in half a column of *Psychological Science* (see Table 2 in p.1362). The *say it* succinctly version we proposed above has but 21 words. Both can be implemented in supplemental materials, keeping the length of published content unchanged. Contrast that amount of red-tape with the over 400 pages of the APA style guide most journals require complying with when submitting.

Figure 2. Number of words required for transparency versus stylish writing.



Nevertheless, if push came to shove, we could maintain the existing amount of red-tape by substituting some arguably less vital aspects of style requirements from the APA guide with those we have propose here. For example, we could eliminate the following 23 words in the “Metrication” chapter of the *Publication Manual* to make space for our proposed 21 words enhancing the credibility of a paper’s findings: “**Spacing**. Never use a space between a prefix and a base unit. Examples: kg, kilogram [...] Do not use a period after a symbol.” (p.130)

b) Exceptions

Debates about change tend to focus on exceptions: “what if this happens?” “what if that happens?” Exceptions, fortunately, are exceptional. They are atypical.

If we accept that the status-quo is not perfect (and in that there appears to be little contention), changes need not be assessed in terms of their perfection, merely in terms of their improvement. A good policy can then be enacted as a default, and waivers be granted for exceptions.

In the early 1900s the United States passed “pure food” regulations that required manufactures to disclose what was in the food they were selling. The practical concerns manufacturers expressed towards transparency in food then mirror those that researchers express towards transparency in research today.

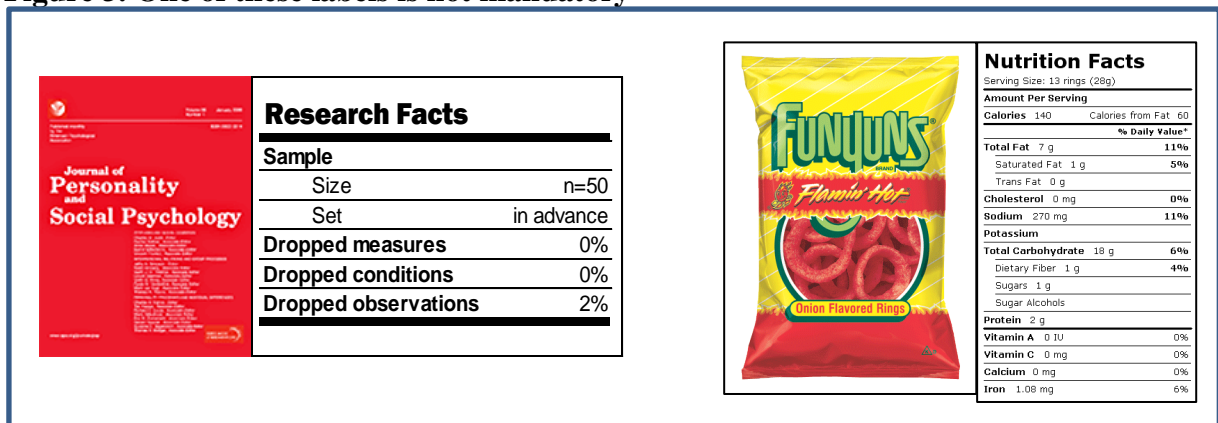
Not unlike objections to require sample size to be $n > 20$, for example, ice-cream manufacturers objected to the arbitrary threshold of 16% of buttery fat to qualify as ice-cream.¹

Not unlike objections that it is easier to comply with disclosure for researchers in some fields than in others, sardines importers complained that labeling the exact type of oil in their cans was harder on them than on domestic producers, etc.²

Fortunately for all of us today, the federal government then knew better than to let the perfect be enemy of the good. If only perfect solutions were implemented, we would still be drafting the Pure-Food act of 1906.

We hope that editors will emulate the pragmatic politicians of the 1900s and implement disclosure requirements in our journals before a perfect solution with no detractors is arrived at. In the meantime, those of us who realize transparency is a necessary condition for evidence to be scientific can start adding 21 words to our papers.

Figure 3. One of these labels is not mandatory



¹ “Do not want to make alcohol conspicuous,” *New York Times*, September 21st, 1906, pp.5

² “Grocers complain of Wiley,” *New York Times*, November 20th, 1904, pp.7