

Souvenir G: The Current State of Play in Psychology

Failed replications, we hear, are creating a “cold war between those who built up modern psychology and those” tearing it down with failed replications (Letzter 2016). The severe tester is free to throw some fuel on both fires.

The widespread growth of preregistered studies is all to the good; it's too early to see if better science will result. Still, credit is due to those sticking their necks out to upend the status quo. I say it makes no sense to favor preregistration and also deny the relevance to evidence of optional stopping and outcomes other than the one observed. That your appraisal of the evidence is altered when you actually see the history supplied by the registered report is equivalent to worrying about biasing selection effects when they're not written down; your statistical method should pick up on them.

By reviewing the hypotheses and analysis plans in advance, RRs (registered reports) should also help neutralize P-hacking and HARKing (hypothesizing after the results are known) by authors, and CARKing (critiquing after the results are known) by reviewers

with their own investments in the research outcomes, although empirical evidence will be required to confirm that this is the case. (Munafò et al. 2017, p. 5)

The papers are provisionally accepted before the results are in. To the severe tester, that requires the author to explain how she will pinpoint blame for negative results. I see nothing in preregistration, in and of itself, to require that. It would be wrong-headed to condemn CARKing; post-data criticism of assumptions and inquiries into hidden biases might be altogether warranted. For instance, one might ask about the attitude toward the finding conveyed by the professor: what did the students know and when did they know it? Of course, they must not be ad hoc saves of the finding.

The field of meta-research is bursting at the seams: distinct research into changing incentives is underway. The severe tester may be jaundiced to raise qualms, but she doesn't automatically assume that research into incentivizing researchers to behave in a fashion correlated with good science – data sharing, preregistration – is itself likely to improve the original field. Not without thinking through what would be needed to link statistics up with the substantive research problem. In some fields, one wonders if they would be better off ignoring statistical experiments and writing about plausible conjectures about human motivations, prejudices, or attitudes, perhaps backed by interesting field studies. It's when researchers try to test them using sciency methods that the project becomes pseudoscience.