

Published in final edited form as:

*Cortex*. 2014 February ; 51: 107–108. doi:10.1016/j.cortex.2013.09.011.

## Incentivising reproducible research

Katherine S. Button<sup>1,2,\*</sup> and Marcus R. Munafò<sup>2</sup>

<sup>1</sup>School of Social and Community Medicine, University of Bristol, United Kingdom

<sup>2</sup>School of Experimental Psychology, University of Bristol, United Kingdom

Wagenmakers and Forstmann argue that journals (and journal editors) should actively solicit replication attempts to confirm important findings (Wagenmakers & Forstmann, 2013). They point out, correctly in our opinion, that current incentive structures in science discourage scientists from adopting practices that are advantageous for science as a whole, because they leave individual scientists disadvantaged relative to their peers. Replication is one example of this, although there are others, such as the general lack of enthusiasm for publishing null results. Much of this arises from the current “publish or perish” culture within science.

These issues are not new concerns about low statistical power and publication bias predate our own (Button et al., 2013a) by at least 50 years (Cohen, 1962; Sterling, 1959). Critically, exhortations that scientists should strive to change their practices often appear to have little effect; previous studies of statistical power, for example, have not led to an increase in statistical power of studies (Sedlmeier & Gigerenzer, 1989). We are therefore sanguine about the prospects that our recent work will have real long-term impact, despite the interesting debate and commentary that it has generated (Ashton, 2013; Bacchetti, 2013; Button et al., 2013b; Quinlan, 2013). At the same time, much of how science is conducted has changed over the last 50 years (Hubel, 2009). An obvious example is the enormous number of statistical tests that it is now possible to perform in minutes, which previously would have been impossible when statistics were calculated by hand. Multiple testing of this kind, when not reported and corrected for, invalidates key assumptions of null hypothesis significance testing (Sterne & Davey Smith, 2001).

Concerns about low statistical power and publication bias were raised 50 years ago, and it is unlikely that changes in our ability to repeatedly interrogate our data will have improved matters. Despite this, most technological advances are of course to the general advantage of science. However, other changes in how science is conducted are structural and relate to the incentives to which scientists respond. While new approaches can be initiated from the bottom-up by individual scientists, this is likely to be slow and potentially hindered by the short-term consequences of, for example, publishing fewer papers. It is therefore dangerous to assume that science is ultimately self-correcting (Ioannidis, 2012), in which case there is a clear need for change to be initiated from the top down. Journals, funding agencies and other key stakeholders have a critical role to play by incentivising research activity likely to benefit the field as a whole. The recommendation by Wagenmakers and Forstmann is one

\*Corresponding author. kate.button@bristol.ac.uk (K.S. Button).

example of this, while the Registered Reports format now available at Cortex is another. What both illustrate is the need for creative thinking and leadership on the part of those in a position to encourage change.

## References

- Ashton JC. Experimental power comes from powerful theories: the real problem in null hypothesis testing. *Nature Reviews Neuroscience*. 2013; 14(8):585.
- Bacchetti P. Small sample size is not the real problem. *Nature Reviews Neuroscience*. 2013; 14(8):585.
- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*. 2013a; 14(5):365–376.
- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Confidence and precision increase with high statistical power. *Nature Reviews Neuroscience*. 2013b; 14(8):585.
- Cohen J. The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal and Social Psychology*. 1962; 65:145–153. [PubMed: 13880271]
- Hubel DH. The way biomedical research is organized has dramatically changed over the past half-century: are the changes for the better? *Neuron*. 2009; 64(2):161–163. [PubMed: 19874784]
- Ioannidis JPA. Why science is not necessarily selfcorrecting. *Perspectives on Psychological Science*. 2012; 7(6):645–654. [PubMed: 26168125]
- Quinlan PT. Misuse of power: in defence of small-scale science. *Nature Reviews Neuroscience*. 2013; 14(8):585.
- Sedlmeier P, Gigerenzer G. Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*. 1989; 105(2):309–316.
- Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance *e* or vice versa. *Journal of the American Statistical Association*. 1959; 54(285):30–34.
- Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? *British Medical Journal*. 2001; 322(7280):226–231. [PubMed: 11159626]
- Wagenmakers, EeJ; Forstmann, BU. Rewarding high power replication research. *Cortex*. 2013