



Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Keeping Researchers Honest: The Case for Sealed-Envelope-Submissions

Martin Dufwenberg and Peter Martinsson

Working Paper n. 533

This Version: December 16, 2014

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
<http://www.igier.unibocconi.it>

The opinions expressed in the working papers are those of the authors alone, and not those of the Institute, which takes non institutional policy position, nor those of CEPR, NBER or Università Bocconi.

Keeping Researchers Honest: The Case for Sealed-Envelope-Submissions*

Martin Dufwenberg[♦] & Peter Martinsson[♥]

December 16, 2014

Abstract: Because journals favor clear stories researchers' may gain by engaging in scientific misconduct, ranging from shady practices like running more sessions hoping for significance to outright data fabrication. To set researchers' incentives straight, we propose sealed-envelope submissions, where editors' and referees' evaluations are based only on the interest of the research question and on the proposed empirical method.

1. Problem

Many worry about fraud in research. By a recent estimate “two-thirds of retracted life-science papers were stricken from the scientific record because of misconduct” (Corbyn 2013, p. 21; cf. Fang, Steen & Casadevall 2013). Couzin-Frankel (2013, p. 68) quotes an anonymous researcher: “We did this experiment a dozen times, got this answer once, and that’s the one we decided to publish.” Tip of an iceberg? Anecdotes? It is not in a researcher’s interest to disclose a shady practice, making it hard to find direct evidence on how widespread cheating is and on how misleading published results may be.¹

It is easier to judge the problem by reflecting on the incentives involved. Arguably there is great cause for concern. Suppose journals wish to “cast results as a story that they believe others

* We have benefited from comments by Georg Kirchsteiger and Larry Samuelson, as well as participants at a seminar organized by Bocconi’s Experimental Laboratory for the Social Sciences. Financial support from Vetenskapsrådet (Swedish Research Council) is gratefully acknowledged.

[♦] Bocconi University, IGIER, U of Gothenburg, U of Arizona, CESifo; martin.dufwenberg@unibocconi.it

[♥] University of Gothenburg; peter.martinsson@economics.gu.se

¹ Economists tried. See List, Bailey, Euzent & Martin’s (2001) survey of unethical behavior, using randomized response techniques that encourage honest responses despite the sensitive topic. Brodeur, Lé, Sangnier & Zylberberg (2013, p. 1) report that, 2005-11, three top economics journals (*AER*, *JPE*, *QJE*) published empirical findings with p-values that exhibit “a valley between 0.25 and 0.10 and a bump slightly below 0.05” seen to indicate that many researchers “*inflate* the value of ... almost-rejected tests by choosing a ‘significant’ specification.”

will want to read” (Couzin-Frankel 2013, p. 68). By backward induction, given the large rewards (grants, tenure, careers!) for publishing well, researchers may gain by tweaking findings (cf. Lacetera & Zirulia 2011). There is a spectrum of possibilities, from shady practices like running more sessions hoping for significance to outright data fabrication.

Proposals to rectify the problem appeared, though efficacy is doubtful. Whistle-blowing by peers involves “significant risks, and the path is rarely simple” (Young, Ledford & Van Noorden 2013, p. 454). Having senior mentors teach integrity may be useful (Neaves 2012), but the possibility of aligned incentives between junior and senior scholars suggest that relying on such honesty may be wishful thinking. Study registration and pre-analysis plans – besides being burdensome to formulate (if more data patterns than the actual one need addressing) and besides denying researchers freedom to incorporate information they didn’t think of before seeing the data – does not solve the following issue: if certain results are more publishable than others researchers will still have incentives to fabricate such results while flagging for them beforehand (cf. Humphreys, Sanchez de la Sierra & van der Windt 2013).

We propose a different approach, where the empirical results are submitted in a sealed envelope to the journals.

2. Solution

The problem may be overcome by a drastic change in how articles are submitted and evaluated for publication at journals. We call it a sealed-envelope submissions proposal:

Journals should insist that submitted articles do not reveal any empirical results. All the data, along with the statistical analyses, should be submitted in a sealed envelope. The editors and referees should evaluate the submission based only on the interest of the chosen research question and on the relevance of the chosen empirical method. After making their accept-reject decision, the editors may then open the envelope.

Our diagnosis of the problem was based on a backward induction argument, and so is now our solution. We trace the key problem – data fabrication – to the conditioning of editorial decisions on the nature of data. Make editorial decisions blind to the nature of researchers’ data and the incentives to manipulate the data go away.

3. Galileo

Related proposals were floated in the past, but concerned avoiding publication bias or project selection rather than eliminating incentives for misconduct (Sterling 1959, Rosenthal 1966, Walster & Cleary 1970, Feige 1975, Dufwenberg 2014). If results are published only if they tell a clear story (*e.g.*, through statistically significant effects), outlier data get over-represented in published work.² These proposals seem to have been largely forgotten or neglected, probably because one can brush off the problem and say that, as long as one is aware of the bias one can adjust one's outlook accordingly. Published data is still real data.

It is much harder to brush off scientific misconduct with an analogous argument. If data is made up, if chosen estimation methods are conditioned on significance, or if reporting is done with spin, how can one tell what's real from what is make-believe? Fake data is *not* real data. Depending on the degree of misconduct, conclusions may vary from dubious to useless. We believe the problem is serious because researchers' incentives are so strong. Furthermore, the risks involved for cheaters may be rather small. "There is no cost to getting things wrong; the cost is not getting them published," as psychologist Brian Nosek put it when consulted for a recent article on the topic (*The Economist*, 2013). With our proposal, editorial decisions become independent of the nature of the data, so no one can gain by manipulating findings.

Researchers have reacted to incentives since Galileo, by many considered as the father of science, denounced heliocentricism. While it is easy to sympathize with his decision, modern-day incentives encourage less laudable researcher conduct. The sealed-envelope submission proposal holds promise to set those incentives straight!

References

- Brodeur, A., M. Lé, M. Sangnier, Y. Zylberberg (2013), "Star Wars: The Empirics Strike Back," Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit, IZA, Bonn, No. 7268.
- Corbyn, Z (2013), "Misconduct is the Main Cause of Life-Sciences Retractions," *Nature* 490, 21.
- Couzin-Frankel, J. (2013), "The Power of Negative Thinking," *Science* 342, 68-69.

² Bias either enters directly through editors' decisions, or because researcher do not bother to write up null findings (cf. Franco, Malhotra & Simonovits 2014).

- Dufwenberg, M. (2014), “Maxims for Experimenters”, forthcoming in *Methods of Modern Experimental Economics*, Frechette, G. & A. Schotter (Eds.), Oxford University Press.
- Fang, F.C., R.G. Steen, & A. Casadevall (2013), “Misconduct Accounts for the Majority of Retracted Scientific Publications,” *Proceedings of the National Academy of Sciences* 109, 17028-17033.
- Feige, E. (1975), “The Consequences of Journal Editorial Policies and a Suggestion for Revision,” *Journal of Political Economy* 83, 1291-1296.
- Franco, A., N. Malhotra & G. Simonovits (2014), “Publication Bias in the Social Sciences: Unlocking the File Drawer,” *Science* 345, 1502-1505.
- Humphreys, M. R. Sanchez de la Sierra & P. van der Windt (2013), “Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration,” *Political Analysis* 21, 1-20.
- Lacetera, N. & L. Zirulia (2011), “The Economics of Scientific Misconduct,” *Journal of Law, Economics and Organization* 27, 568-603, 2011.
- List, J., C. Bailey, P. Euzent & T. Martin (2001), “Academic Economists Behaving Badly? A Survey on Three Areas of Unethical Behavior,” *Economic Inquiry* 39, 162-170.
- Neaves, W. (2012), “The Roots of Research Misconduct,” *Nature* 488, 121-122.
- Rosenthal, R. (1966), *Experimenter Effects in Behavioral Research*, New York: Appleton-Century-Croft.
- Sterling, T. (1959), “Publication Decision and Possible Effects on Inferences Drawn from Tests of Significance – Or Vice Versa,” *Journal of the American Statistical Association* 54, 30-34.
- The Economist* (2013), “Trouble at the Lab,” Issue 19th-25th October.
- Walster, G. & T. Cleary (1970), “A Proposal for a New Editorial Policy in the Social Sciences,” *American Statistician* 24, 16-19.
- Young, E., H. Ledford & R. Van Noorden (2013), “3 Ways to Blow the Whistle,” *Nature* 503, 454-457.