

COMMENTARY

Is open science rewarding A while hoping for B?

Paul E. Spector 

School of Information Systems and Management, Muma College of Business, University of South Florida and Tampa General Hospital, Tampa, FL, USA
Email: pspector@usf.edu

Reading Guzzo, Schneider, and Nalbantian's (2022) paper on the unintended consequences of open science policies reminds me of the classic Steven Kerr (1995) paper "On the Folly of Rewarding A, While Hoping for B" in which he makes the point that often we fail to consider the reward systems when we attempt to influence behavior. Such is the case, it seems to me, with the open science movement. Its solution to questionable research practices is to create a system of journal policies that focuses on individual researcher behavior rather than the reward systems that produce it. It is based on an implicit assumption that open science policies prevent questionable research practices and even outright research fraud. Without reforming the reward structure, however, such an outcome is highly unlikely.

There are two classes of behavior that open science approaches are designed to reduce: research fraud and questionable research practices. Both arise under a system in which career outcomes are determined by publication in a limited number of exclusive academic outlets. Many departments, particularly in business schools, require publication in a small list of elite "A" journals. Those journals have very narrow requirements that include, as Guzzo et al. note, theory and hypotheses that are confirmed by data. Failure to confirm hypotheses in a paper makes it difficult if not impossible to publish in the places that make one competitive on the academic job market, earn tenure, and reap other rewards (e.g., financial bonuses and reduced teaching loads).

Researcher career rewards are based on publishing in the "right" places. Publication rewards (acceptance in top journals) are dependent on not only having theory, hypothesis, and confirmation, but also in convincing reviewers that a given submitted article makes a significant contribution. This means not only filling some perceived gap in the literature but being seen as novel in some way. Finding results that seem counterintuitive can be a good novelty strategy, but they are hard to realize without using a few research tricks. This reward structure—what it takes for career and publication success—puts researchers under tremendous pressure to survive professionally in a cut-throat and hypercompetitive environment where journal acceptance rates in top outlets are about 5%, and authors must fight reviewers over trivial issues through multiple rounds of review. Should we be surprised that so many researchers, facing the dilemma of either gaming the system or finding a different line of work, choose the questionable route? I find myself skeptical that a new set of requirements on researchers will produce a different result.

Research fraud

One of the leading explanations for criminal fraud was proposed by Donald Cressey (1953) and is today known as the Cressey Fraud Triangle. Based on interviews with a sample of incarcerated fraudsters, Cressey concluded that fraud is the byproduct of three elements.

- Pressure: A strong need for money or other rewards (e.g., publication in rigorous peer-reviewed outlets).
- Opportunity: The ability to commit the fraud successfully. (e.g., sufficient research skills to fabricate a dataset and conduct fraudulent analysis).
- Rationalization: A willingness to justify actions (e.g., convincing yourself that no one will be harmed or that “everyone else” is doing it).

There is little in open science procedures that addresses these three elements. Pre-registration, data sharing, and other open science ideas might be good practices, but they will not reduce pressure for publication, reduce the opportunity for dishonesty, nor impact rationalization. An individual who is willing to commit scientific fraud under the old system can easily pre-register and then create datasets that produce desired results and even share synthetic data and fictitious procedures. Although the need for transparency might make some extra work for the scientific fraudster, it does not prevent fraud.

Questionable research practices

Guzzo et al. explain how open science is intended to reduce questionable research practices like HARKing (hypothesizing after results are known) and p-hacking (iterative analyses designed to create statistical significance, or what might be called a Type 1 error hunt). Such practices are not considered outright fraud because they are based on legitimately collected data, but their use can be problematic nevertheless. Such practices arise in the same environment as research fraud and can be considered from the perspective of the Cressey Fraud triangle. Researchers engage in questionable research practices because of the pressure to publish, their research skills that make it possible, and rationalization. In fact, it is easier to rationalize questionable research practices than fraud because their use is widely accepted and even encouraged. Who among us hasn't been asked by editors and/or reviewers to add/change hypotheses, include control variables, and add/change analyses?

The open science movement has certainly raised awareness about the problems with questionable research practices, but when faced with pressure to publish in A-list journals, what is the academic researcher to do? You have spent more than a year working on a study, investing hundreds of hours, only to find that your expected results were not achieved. Do you give up on the project at this point, or do you switch into data mining mode and conduct a series of analyses to milk the dataset of significant results? Knowing that top (or any) journals wouldn't be welcoming to a paper that describes your exploratory approach, you craft a set of hypotheses based on your significant results and submit the paper as testing a set of theory-based hypotheses.

The reforms we need

Until we reform the reward system for both journals and universities, open science will have limited impact on how researchers conduct themselves. There is too much pressure to engage in questionable research practices, and in the extreme, scientific fraud. Questionable research practices are so widespread that they are easy to rationalize. I once overheard a professor instructing a doctoral student in p-hacking the data from his experimental dissertation study if he wanted to publish in a journal, explaining that this is how it works in the real world.

Reform universities

One of the main functions of a tenure-earning faculty member, especially in a Research 1 university, is to contribute to the knowledge base of their field and to have impact. Unfortunately, this is typically operationalized almost exclusively by number of publications. Even worse in many

departments, especially in business schools, there is a journal list that faculty must publish in to gain tenure. These lists can be quite short; the coveted UT Dallas list contains just 24 journals that represents all the business disciplines. Competition for these journals is fierce, and success would be difficult without taking advantage of all the tricks of the trade.

Universities need to rethink how they evaluate faculty research performance. We need to go beyond numbers of publications and where the work appears as these are deficient criteria. Publishing in A-list journals does not guarantee that work will have impact on a field, and given questionable research practices in wide use, such work might not even replicate. A broader view of criteria include.

- **Importance of the problem:** Often introductions of papers include discussion of the nature of the problem and why it is important. This goes beyond just a gap in the literature, as sometimes gaps exist because questions are trivial. It gets to the issue of why anyone should care about the problem the researcher is addressing. Will the application of this line of research improve organizations or people's lives in organizations? Faculty research performance should include consideration of importance.
- **Extent to which work is programmatic.** If a researcher is seriously pursuing a problem, it will take more than one study to address it. Multiple studies are needed to demonstrate that a finding can be replicated, and they are needed to rule out alternative explanations, suggest explanatory mechanisms, and identify boundary conditions (a point made by Guzzo et al.). A coherent set of studies spread across a variety of journals, perhaps none of which are A-list, can have more impact and importance to the field than a series of one offs in UT Dallas list journals, especially if none of those one offs can be replicated.
- **Appearance in reasonable outlets for the problem.** Where the work appears should be based on its nature. If you are studying teamwork in surgical teams, the best outlets might include medical journals. Should we discount a paper that appeared in the *New England Journal of Medicine* because it isn't on "the list"? Publishing in a wide range of journals across disciplines should be seen as a plus, but too often it is not. We should look for broad impact and not just impact in a tiny subdiscipline.
- **Research impact:** Ultimately the goal of research is to have impact on the world. Publication should be the means to the end and not the end itself. Unfortunately, the reward structure focuses faculty on publication and not beyond. How else do you explain why so many researchers fail to respond to requests for copies of their work? As much attention should be paid to citation and other indicators of how research impacts the field as to publication itself.
- **Impact beyond the academy:** There are many ways to have impact beyond academia. Much of the work from business and I-O psychology can potentially impact the broader society, and that impact should be encouraged and rewarded.

Universities need to consider that research performance is complex, requiring the assessment of several criteria to adequately capture it. Not everyone will have high performance in all areas, for example, one person's work might have primarily an academic impact whereas another's will inform practice. We need rational flexibility in evaluating individual performance in its entirety and not just count the number of papers on a journal list.

Reform journals

Journals share the blame for questionable research practices in that they encourage and, in some cases, demand them. The most elite journals in the organizational sciences have evolved a rigid, one-size-fits-all approach to journal reports. They must include a complex theoretical narrative and a set of hypotheses that are linked (often quite loosely) to one or more theories, find support

for at least some of those hypotheses, and offer some level of novelty. They are unfriendly to research that doesn't fit the mold, including work that is inductive, qualitative, and simple. Work that gives the illusion of rigor is valued over work that is directly relevant to practice. Reforms include.

- **End the collusion.** Journals are as responsible, if not more, than authors for questionable research practices. Journal editors need to put an end to reviewer-instigated questionable research practices. When reviewers ask for new or modified hypotheses, editors should push back. Reviewers should be told that this is inappropriate, and authors should be told to ignore such suggestions. When reviewers ask for new analyses or control variables, again editors should push back. If editors of the major journals would start doing this, it would not be long before everyone got the idea that these are unacceptable practices.
- **Be more flexible.** Journals need to be open to a variety of approaches and methods. There should be room for both inductive and deductive research—every paper should not have hypotheses and theory. Journals also should pay more attention to the importance of the problem and less to whether results are statistically significant or not. The importance of a problem should not be determined by how the results came out. A better balance between papers finding versus not finding statistically significant findings will reduce the allure of p-hacking. A balance between inductive and deductive papers will remove the pressure to HARK.

Concluding thoughts

The open science movement is an attempt to improve science by reforming the practices of researchers. The practices suggested are hard to fault if applied in a flexible and thoughtful way as noted by Guzzo et al. Whether journals require specific actions or not, we should all do our best to avoid HARKing by only testing a priori hypotheses and p-hacking by preplanning our analyses. We should be sure to thoroughly explain our procedures, either in research reports or supplemental materials. And we should respond to requests for our data and materials so others can better understand what we did. These are all sound practices that benefit our science.

But I share Guzzo et al.'s (2022) concerns that if these open science practices become institutionalized and bureaucratized, they might further limit the sorts of research that gets published, pushing us into an even smaller one-size-fits-all box. To some extent, it feels like imposing open science practices on researchers is a case of blaming the victim. If we really want to reform our science, we need to change the reward structures to remove incentives to game the system.

References

- Cressey, D. R. (1953). *Other people's money: A study in the social psychology of embezzlement*. The Free Press.
- Guzzo, R. A., Schneider, B., & Nalbantian, H. R. (2022). Open science, closed doors: The perils and potential of open science for research in practice. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *15*(4), 495–515.
- Kerr, S. (1995). On the folly of rewarding A, while hoping for B. *Academy of Management Executive (1993–2005)*, *9*(1), 7–14. <http://www.jstor.org/stable/4165235>