Has science gone bad?

Konstantinos Pelechrinis*

School of Information Sciences, University of Pittsburgh, Pennsylvania, USA

A heated debate has been at the forefront of scientific circles during the last few years. Rogue studies being published with falsified data, abuse and misuse of statistical analysis tools, and a general lack of trust in scientific methods have been regular topics not only within the scientific community but in the popular press as well. In this paper I try to point to some crucial elements of the whole process that generate positive feedback to the problem, hence further intensifying it. I will also make a proposal on how the workflow of academic publishing could be changed to alleviate some of the problems.

1. The problem

A few years back the *Economist* focused extensively on the issue of "bad norms and practices" in science.[1] The main argument/observation was that researchers do not validate published results any more. This leads to the publication of studies that are irreproducible and potentially wrong. It is not clear whether this is a result of intentional or unintentional manipulation of the data and/or methods, widely known as *P*-hacking (Nuzzo, 2014). Despite the awareness raised by similar articles in the popular press, the situation has become worse to the point that the American Statistical Association recently published guidelines on how *P*-values should be used in the scientific literature (Baker, 2016).

With such complex problems it is hard—or even impossible—to pinpoint the root cause(s). Nevertheless, there are factors that seem highly plausible for explaining part of the problem. Ioannidis has extensively studied it and proclaimed that "it can be proven that most claimed research findings are false" (Ioannidis, 2005). In reaching this conclusion he considered the statistical power of an experimental design, the effect of size and the prestudy probability of a relationship being true, among other factors. Ioannidis' work provides a principled, statistical way to think about the issue and it sheds light on major methodological issues with experimental design. On the contrary, in this article I want to emphasize two different, behavioral factors that have received considerably less attention but which, I believe, can explain part of the situation. In particular, in the rest of this paper I shall examine the motives behind pursuing a PhD degree today as well as the publication bias that dominates the world of academic journals. My disclaimer is that my experience is mainly within the greater area of computer science and engineering, but doubtless similar phenomena appear in other areas, possibly to a different extent.

2. Why get a doctorate today?

A doctoral degree has been long thought as the channel through which science advances and new developments emerge. During most of the 20th century the degree was associated with a scientific hunt for knowledge that was mainly geared towards highly motivated individuals that were intrigued by solving hard, fundamental, life-changing problems. Around the turn of the century though, things started changing. A much larger number of students graduate with a doctoral degree. Based on a report from US National Science Foundation (NSF) (Fiegener, 2011), between 2000 and 2010 there has been a 28% increase in the number of doctorates awarded in science and engineering fields (the report observes that there was a reduction between 2009 and 2010 but it is extremely small, 0.9%. In computer science the increase is even higher, reaching almost 100% (see Figure 1)! While this by itself is not necessarily bad we have to bring into our minds a few things.

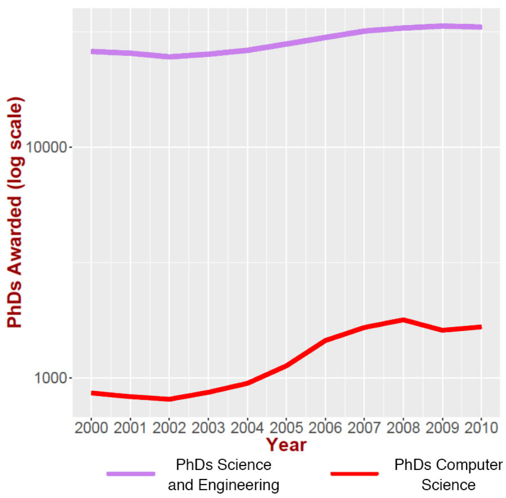

Figure 1. During the decade 2000–2010 there has been an increase in the number of PhDs awarded in the science and engineering fields (source: Fiegener, 2011).

* Corresponding author. Email: kpele@pitt.edu

1 E.g., http://tinyurl.com/ovwctme and http://tinyurl.com/lx96am6

Journal of Biological Physics and Chemistry **16** (2016) 75–78
Received 21 March 2016; *accepted* 31 March 2016

doi: 10.4024/06PE16F.jbpc.16.01

Computing and scientific advances were made possible during the second half of the last century because many of the fundamental scientific problems were solved after the Second World War and until the late 1990s. After a while the "new problems" that appear are mainly variations of existing ones. Similarly, the solutions offered are minor tweaks of existing ones. Of course one can argue, and rightly so, that new problems and breakthroughs nevertheless always arise. For example, compressive sensing, which was developed in the mid-2000s, is certainly one such fundamental breakthrough, which will revolutionize a large number of applications that rely on sampling. But in general their number is diminishing and there is no reason to think that more fundamental problems have all of a sudden appeared like the heads of a hydra. Hence, with this in mind, the increase in the number of doctoral degrees appears strange at first glance. If there are not more fundamental problems than before, why this increase?

Part of the answer is related to the job market. An increasing number of students choose to pursue an advanced degree for the chance to compete for a better-paying job in industry (this is true at least in the USA). While there is not anything bad with this in principle, there are side effects that are deleterious. These students are not motivated by the same principles as students who pursue a doctoral degree with the goal of advancing knowledge. Given their different motivation, these students do not show the same persistence and, hence, they settle for solving problems that are either unrealistic or superficial or both, which results in huge numbers of publications that are bogus and of unknown importance and relevance. It is not uncommon for students to "prepare" papers to be published without taking great care of the experimental design, data collection and methodological analysis. They just want to publish their work, graduate and join the workforce. And there are always journals and conferences interested in such studies!

Of course, someone who wants to play the devil's advocate here would say it is the rôle of a PhD student's advisor to guide and ensure the quality of the work. This is in principle very true but, again, the way the academic system operates does not leave many options to advisors. The number of PhD students graduated is one of the elements that can make or break an advisor's promotion case, who therefore desires to graduate as many PhDs as possible. This is evidenced by the increase in the numbers of PhDs in general (i.e., you need to keep up with the general trend). This essentially guarantees that papers of unknown quality will find their way to journals and conferences because students need to graduate at some point.

So why is this a plausible cause of the various problems related to irreproducibility of scientific work? Simply because students are not interested in identifying the truth. They are just interested in publishing. Actually, the bigger the headlines their work makes the better for them. Hence, they are looking for highly improbable results, which Ioannidis has actually shown to be more prone to being false positives and, hence, irreproducible (Ioannidis, 2005). Unfortunately, the doctoral degree has become the new "advanced Master's" degree. I am seeing an alarmingly increasing number of students who, after completing their Master's degree, turn to doctoral studies simply because they did not find the job of their dreams. And it is clear from my interaction with them that these students are producing—on average because there are always exceptions—less significant work than the student with the goal of advancing knowledge.

3. Publication bias

Even more crucial for the current situation is the academic peer review process, in particular the so-called publication bias, meaning that journal editors have a predisposition to accept studies that support a *statistically significant* result (Rothstein, Sutton and Borenstein, 2006). This phenomenon has been studied with the finding that papers with statistically significant results are at least three times more likely to be published compared to studies with null results, even when controlling for the quality of the experimental design (Easterbrook et al., 1991; Dickersin et al., 1987).[2] Furthermore, various scientometric studies have reported that this bias is becoming more pronounced with time (Fanelli, 2011).

To dramatize the phenomenon let me consider a fictitious scenario where a group of scientists is studying the impact of eating hot dogs on the academic performance in SATs.[3] It turns out that after feeding a random set of high school students for six months during their senior year the only impact the hot dogs had was on the students' weight. No correlation between the SAT performance of the treated group and that of the control group was identified. The scientists, happy with their discovery (implying that we do not have to make our kids fat to get into college!), submit their work for publication. Soon enough they get a notification from the journal, which regrets to inform them that their work will not be published since no causal link has been identified. Of

[2] Presumably publication of these results was itself not immune from publication bias.
[3] Scholastic aptitude tests, widely used in primary and secondary schools.

course, this is masked by using different words; e.g., "The study does not advance the scientific field since no substance that can increase the SAT scores is identified", as if identifying what does *not* cause something is not important. After a while another group learnt about the study that was never published and they decide to repeat the experiment. Again, the only effect identified was on the weight of the students and again the scientists had no luck publishing the results. Fast forward one year and the same journal receives a paper claiming a causal link between eating hot dogs and acing the SATs. Gladly, after a short review cycle, the paper is accepted and praised in the popular press. Now you are well informed that instead of spending money on tutoring for SATs just spend it on hot dogs (and make sure to keep some savings for health insurance later). But are you?

The reason for chosing such an obviously false positive example is that the more improbable a hypothesis, the more intriguing it is for journal editors. Of course, the blind trust on *P*-values has its share of blame here. A *P*-value expresses the probability of obtaining a result as extreme as the one observed simply by chance. When a hypothesis is tested at a specific significance level, typically 0.05, a rejection of the null hypothesis does *not* make the alternative true, it just expresses the fact that there is a rather low probability that the evidence points against the null by pure chance. However, what this means at a meta-study level is that if you fix the significance level at 0.05 and you repeat the study 20 times, then if there is indeed no effect (i.e., hot dogs do not lead to better SAT scores) then even by chance you expect that one of these repetitions will reject the null and identify a (false) positive effect![4] However, the 19 studies that failed to reject the null will never get published, while the 1 false positive study not only will be published but it will be treated as the indisputable truth! Figure 2 visualizes this process. Based on this argument it does not come as a surprise that most of the published studies cannot be replicated.

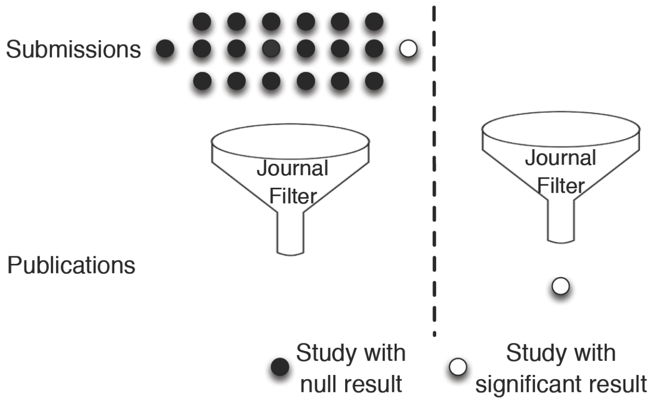

Figure 2. Studies that have not identified any significant result (i.e., P-value > 0.05) will not go through the "journal filter" of publication bias. They do not appear at the other end of the filter-funnel.

4. Changing the peer review workflow

In general, it seems that no small tweak will lead to a dramatic and long-lasting impact on the problem. Here I put forward a proposal for the way that the peer review workflow should be changed to accommodate the discovery of the truth (see Figure 3). Of course, the way scientific journals operate is undoubtedly hard to change.

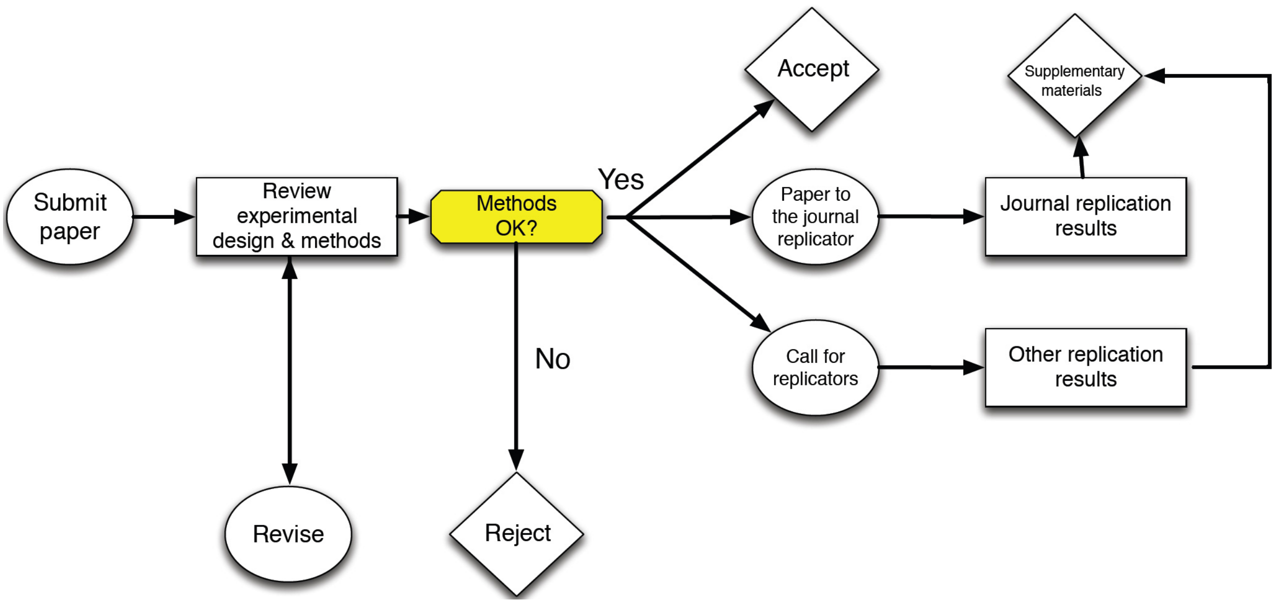

Figure 3. Replication of studies needs to be an inherent part of the review process.

[4] More realistically, repeating the study 200 times might yield 10 false positives.

Negative results need to be more widely accepted (there are journals that publish only negative results but they are few and far between). It is good to allow the scientific community to decide whether a hypothesis examined and a negative result are important. This requires a hard reset of mindset and a more rigorous reviewing process. I guess we should not be delusional at best if we were to think that three reviewers can speak for the whole community.

The reviewers should be responsible only for checking domain knowledge facts and the analysis. We should add another rôle to the review process—the *replicator*, responsible for replicating the original experiment (or theoretical derivation) reported in the submitted manuscript. His or her name will be visible to the original authors, who could suggest possible replicators (similarly to the suggestion of reviewers). One logistic question is who will pay for the replica experiment. Well, this is the really hard question but it may be that state funding agencies would cover such costs (e.g., they could fall under the same category as the travel grants for conferences that the NSF regularly awards). Of course, other practical problems will arise; for instance, some studies require years to be completed (e.g., medical longitudinal studies on long-term impacts of drugs). However, we can definitely start with studies that span a shorter interval; because it cannot be done for all studies does not mean that we should not do it for the ones we can! Of course this means that *full* turnaround time might get slower, but in the meantime the paper can be made available to the scientific community with the appropriate disclaimers as it goes through the various stages (e.g., not peer-reviewed, peer-reviewed but not replicated etc.).

What will happen if the experiment cannot be replicated? Well, the paper can still be published (assuming that the reviewers found the experimental design and methods used in the original work to be sound) but the replication study will also be published, potentially as supplementary materials for the original study, thus giving the replicator credit as well as the original authors, and institutions need to recognize replication not as a simple service but as a form of publication—because in a sense it is. To move one step further, the journal could also call for "replication" contributions; that is, invite researchers in the respective fields to take upon the task of replicating specific published studies. Various policies of course need to be in place to handle things such as conflict of interest, but similar policies are already in place for other purposes.

I doubt that any journal will take this suggestion into consideration with the current mindset but I put it forward in the belief that it encompasses the true nature of science: scientists making discoveries, and their peers validating or refuting them until we get a set of studies on the same hypothesis that we can meta-analyse and reach a final conclusion. But for now all these ideas sound like scientific utopia.

5. Parting thoughts

This essay is not an effort to point the finger to "bad" scientists. After all, I am sure that it is hard to not have committed unintentional errors. False positives appear by chance, and there is little that one can do about it. However, the way the system operates intensifies the problem, making scientists look for these "black swans"—even when they are not naturally black but just painted. Promotions and research funding—the two most important things in the career of an academic scientist nowadays—require publications. Getting a doctoral degree to improve one's chances for a well-paid industry job requires publications. And publications require playing with the journal rules; i.e., finding black swans.

References

Baker, M. Statisticians issue warning over misuse of *P* values. *Nature* **571** (2016) 151.

Dickersin, K., Chan, S., Chalmers, T.C, Sacks, J.H.S. and Smith, H. Publication bias and clinical trials. *Controlled Clin. Trials* **8** (1987) 343–353.

Easterbrook, P., Gopalan, R., Berlin, J. and Matthews, D. Publication bias in clinical research. *The Lancet* **337** (1991) 867–872.

Fanelli, D. Negative results are disappearing from most disciplines and countries. *Scientometrics* **90** (2011) 91–904.

Fiegener, M.V. Number of doctorates awarded in the United States declined in 2010. *Info Brief (National Center for Science and Engineering Statistics)* NSF-12-303 (November 2011).

Ioannidis, J.P.A. Why most published research findings are false. *PLoS Med.* **2** (2005) 124.

Nuzzo, R. Scientific method: Statistical errors. *Nature* **506** (2014) 150–152.

Rothstein, H.R., Sutton, A.J. and Borenstein, M. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Hoboken: Wiley (2005).