Book Review

Stuart Ritchie (2020). *Science Fictions: How fraud, bias, negligence, and hype undermine the search for truth*. Metropolitan Books. ISBN 1250222699.

Science has been the way we have learned the secrets of nature and converted them into useful practices that have advanced human understanding of how it works and how we can use knowledge to build, manipulate, and understand the complexities of our world and even the cosmos. We have trusted both the people who do the work of science and the methods that are used to check research results for accuracy. When things go as intended, the process works and errors are detected either in the process leading to publication or later when reported results are tested by independent researchers.

About 5 - 6 years ago, we first heard that a large number of published research papers failed replication. Initially, most of this was related to psychology, but we learned that other fields were showing similar problems. The discoveries turned out to be true and the problems were not limited to replication, but to many aspects of the process from investigation to publication. In *Science Fictions*, Ritchie examines each of the primary entities involved in scientific research: universities, researchers, funding agencies, peer reviewers, journals, publishers, and later researchers who cite the published findings, and various science news outlets. Each has an interest in the process. The various parties understand how they are measured and have learned to manipulate aspects of the total process to benefit from the natural competition that they experience. We expect people and their institutions to want to compete fairly, but the process is vulnerable to manipulation. Some of the things that they do to enhance their products (papers) are minor (not publishing null results, etc.) and do not corrupt science to the point of failure. Unfortunately, some go into outright fraud, causing real damage to science as an institution and to participants in the steps from funding to publication and later citations.

> Nearly 90 per cent of chemists said that they'd had the experience of failing to replicate another researcher's result; nearly 80 per cent of biologists said the same, as did almost 70 per cent of physicists, engineers, and medical scientists. Only a slightly lower percentage of scientists said they'd had trouble in replicating their own results.

The extent of misconduct is difficult to grasp, partly because there are so many forms of such practices that are scattered among the entities that figure into the end result. Some egregious cases discussed in the book are blatant and deserving of serious repercussions. Ritchie gives examples of incidents that have led to false results and sometimes have been taken as fact for years, causing harm to individual and institutional reputations. When the research relates to medical treatments or practices, false information is obviously of great concern.

The extent of the truly reprehensible misconduct, such as falsifying data, is not clear from the book, but is presumably less common than tweaking results to improve statistical significance or to enhance the apparent strength of outcomes. One might think that researchers would not use altered images or data copied from a different study or other forms of misrepresentation that are rather easy for another scientist to detect, but we find examples listed.

Ritchie's topics are well covered, with examples and discussions of the factors associated with the

motivation for each category of misconduct. The best way to understand what is in the book is to review some of the salient topics. I will briefly list them below:

Conduct primarily related to researchers

Retraction - There is a lengthy discussion of the impact of retracting a paper by the author or journal. A retraction is a badge of dishonor that researchers understand, yet a few submit worthless papers on a regular basis. Retractions: honest mistakes account for about 40%; 2% of scientists account for 25%; and fortunately only 0.04% of papers are retracted. I learned that there is a Retraction Watch Database–not the place to improve your career options.

p-hacking - This practice has been more frequently discussed in recent years. It is the attempts of a researcher to achieve significance ($p < 0.05$) by running underpowered studies repeatedly until one happens to reach the standard for significance. This is much the same as spinning a roulette wheel until the ball finally lands on the number that is selected. Ritchie gives us an excellent description of p-hacking and explains how it results in bad science.

Fraud - The book provides detailed discussions of data being altered, copied, and fabricated.

Negligence - While this is not the result of intentional data manipulation, it does reflect poorly on the researcher and the reviewers. It may happen as the result of a typo in a spreadsheet, or data that was omitted. One amusing error is mentioned in connection with autocorrect software: "... gene names like SEPT 2 and MARCH 1 were mistakenly converted to dates."

Salami-slicing - Publishing multiple findings from a single study as separate papers. While it is not fraudulent, it adds to the mess of too many papers and publication counts being distorted. Researchers want to maximize their h-index; more papers can mean more citations, without any scientific benefit.

Self-citation - Some authors try to game the "system" by including lots of references to their own papers, thereby adding to the citation count. As with most of the practices Ritchie cited, this one is yet another example of how any incentive will cause at least some people to do things that are unethical or worse.

File-drawering - Some studies simply vanish by not being published. It could happen at either the researcher or journal level. The reason is often that the study produced null results. While this is a mild transgression, it contributes to publication bias.

Conduct primarily related to journals and publishers

Publishing only large effects - Researchers are driven to such things as p-hacking because they have learned that they must show significance or their work will not be published. Journals want big effects and new findings and will selectively publish these. This motivates researchers to push for precisely those end points, knowing that failure to publish means wasted effort and no credit for their work.

<u>Accepting faulty papers</u> - While there are unscrupulous journals that will publish anything for a fee, there are examples of even the top journals publishing material that did not replicate:

> The highest profile of these involved a large consortium of scientists who chose 100 studies from three top psychology journals and tried to replicate them. The results, published in Science in 2015, made bitter reading: in the end, only 39 per cent of the studies were judged to have replicated successfully. Another one of these efforts, in 2018, tried to replicate twenty-one social-science papers that had been published in the world's top two general science journals, Nature and Science. This time, the replication rate was 62 per cent. Further collaborations that looked at a variety of different kinds of psychological phenomena found rates of 77 per cent, 54 per cent, and 38 per cent. Almost all of the replications, even where successful, found that the original studies had exaggerated the size of their effects.

<u>Fake peer-review</u> - This happens at the journal level and is a corruption of the intent of peer review. Papers may be passed along as peer-reviewed, but the peer review is either nonexistent or perfunctory.   This is one example where replication would be unlikely.

<u>Publishing for citations</u> - The journals want to maximize their impact factor and this means garnering citations. The problem is that this excludes null findings and outcomes that are not going to be cited, so the end result is publication bias. In some cases the journals seek citations via citation cartels.

> In the face of a growing number of citation cartels, Thomson Reuters, the company that calculates impact factors, has begun to exclude certain journals from its rankings because of their 'anomalous citation' practices.

<u>Predatory journals</u> - These journals will publish anything for a price, making it possible for a researcher to publish poorly designed studies or even totally fabricated data and results.

Repairing the problems

Most readers of *Science Fictions* will probably either be researchers or those in the sequence that leads to publication. These people will already be familiar with most of the items Ritchie documented and will also know that various practices have been adopted to help resolve the abuses that have become increasingly problematic for the conduct of good science.

<u>Pre-registration</u> - This has been implemented in some journals. Ritchie also identified a more rigorous version of it in which the pre-registration is subjected to peer review. The journal commits to publish the paper, irrespective of the results. These approaches are supposed to reduce publication bias and the motivation to overstate effect sizes. Ritchie points out that some pre-registered papers simply vanish and some researchers make changes to their analysis after it begins.

<u>Open science and open journals</u> - In the case of open journals, the review process is done online and in the open. In the case of open science, the entire process is done online with full view of any interested parties. There was no specific mention of whether this could be used in a broad sense, such as in studies of medicines and related devices.

<u>Preprints</u> - These increase openness and help speed the process of getting information out to others.

<u>Automated checks to expose corrupt data</u> - Various algorithms have been developed to evaluate papers and data to look for likely errors or manipulation.

> The great thing about statcheck, the GRIM test, and Carlisle's method is that they can all be performed using just the summary data that are routinely provided in papers: things like p-values, means, sample sizes and standard deviations.

There are various other remedies discussed. It remains to be seen how many of these become standard practices, but it is likely that those that survive will help the overall integrity of the system.

Missing

The discussions of publication bias were focused on factors such as reaching significance, large effect sizes, prospects for citations, etc. Less was discussed about political and ideological biases, although there is mention of the heavy liberal weight in numbers of professors in social sciences. In some research domains, there are politically charged subjects (intelligence, global warming, education, etc.). Noah Carl, Michael A. Woodley of Menie (2019) lists intelligence researchers who have suffered serious consequences for addressing forbidden topics. Some researchers have had inordinate difficulties in finding publishers who will even consider their books. One instance involved James Flynn, who was well known to be strongly on the left side of the political spectrum, but when he wrote a book in favor of free speech on campuses, he was denied publication by Emerald Press, even after it was listed in Emerald's September 2019 catalogue (James Flynn, 2019). In the more distant past, Chris Brand's book The g Factor: General Intelligence and Its Implications was withdrawn by Wiley Publishing because it dared to discuss a politically taboo topic (Constance Holden, 1996). This category of publication bias is not simply bias, it is censorship. I would have liked to see it included in Ritchie's otherwise broad and inclusive account of bad practices.

References

Constance Holden (1996). Wiley Drops IQ Book After Public Furor. *Science* Vol. 272, Issue 5262, pp. 644.

James Flynn (2019). My Book Defending Free Speech Has Been Pulled. *Quillette*.

Noah Carl, Michael A. Woodley of Menie (2019). A scientometric analysis of controversies in the field of intelligence research. *Intelligence* 77, 101397.

Robert L. Williams
bvv@cox.net