Journal of Scientific Exploration, Vol. 35, No. 2, pp. 422–432, 2021 0892-3310/21

BOOK REVIEW

Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth by Stuart Ritchie. Metropolitan Books (Henry Holt), 2020. 353 pp. $29.99 (hardcover). ISBN 978-1250222695.

Reviewed by Henry H. Bauer

Virginia Polytechnic Institute & State University
hhbauer@vt.edu; www.henryhbauer.homestead.com

https://10.31275/20211975

This book does a splendid job of describing and documenting the dysfunctional features of contemporary science mentioned in the book's subtitle. Were I still teaching, I would have my students read this book as the basis for many productive class discussions. The margins of my copy overflow with notes, comments, and cues for further reading. The 80 pages of endnotes, for some 260 pages of text, are the best and most interesting documentation that I can recall ever finding in such a book. At any rate, I recommend this book wholeheartedly; I doubt that anyone interested in the nature of contemporary science will fail to be informed and to find stimulation for further thought and reading.

The Preface already promises that this will be a page-turner. Many will be astonished and disheartened by the fully documented cases of outwardly distinguished academics whose work was largely or completely fraudulent, as with Diederik Stapel (pp. 4–5 and later).

Ritchie quite appropriately sees replication as the essence of science (p. 5): "If it won't replicate, then it's hard to describe what you've done as scientific at all." Note that this is an *empirical* statement, not the Popperian criterion that theories must be falsifiable in principle if they are to be regarded as scientific. If a claimed observable phenomenon cannot be repeated, then we cannot know that it was real, that it happened even once, when first claimed. That's the continuing dilemma

for parapsychology, cryptozoology, for anomalistics in general. Ritchie points out that the scientific community failed to handle appropriately the issue of replication in the case of Stapel, and also with Daryl Bem's claimed evidence of precognition. Overall, peer review and journal publication practices have not saved science from "a dizzying array of incompetence, delusion, lies, and self-deception" (p. 7).

Part I of the book, "Ought and Is," illustrates that science does not work as we think it should, for the most fundamental reason that scientists are human beings engaged in a social activity (Chapter 1, How Science Works; Chapter 2, The Replication Crisis). Science is socially constructed, as is often said, but only in the sense that the subjective inputs from individuals interact to yield something more objective, or at least less subjective (p. 14).

Unwarranted dogmatism pollutes science. The book, *Thinking Fast and Slow*, whose author Daniel Kahneman had won a Nobel Prize for economics, garnered almost universal rave reviews, but half a dozen years later, Kahneman confessed that he had been wrong in describing studies of priming as being unquestionably true (p. 28). "Power posing" was another fad widely greeted as an important psychological insight, promoted by "the second-most-watched TED talk ever" and a best-selling book (p. 29) before being debunked. So too with the Stanford Prison Experiment and the studies of obedience by Stanley Milgram, which continue to be widely cited by pundits and others who have not become aware of how badly flawed these studies were (pp. 29–30). The trouble is that once the media have welcomed as true an important claimed discovery, it continues to be mistakenly taken to be true by huge swaths of society: The debunking never gets as enthusiastic and prominent coverage as the initial claims of remarkable discovery. "The studies that failed to replicate continue to be routinely cited both by scientists and other writers: Entire lines of research, and bestselling popular books, were being built on their foundation" (p. 32); and "these are just the ones we know about" (p. 34). The lack of replication pervades science as a whole, but it is most troublesome for society on matters of medical research and practice (p. 38 ff.).

Part II of the book gives details of "Faults and Flaws": Chapter 3 on fraud, Chapter 4 on bias, Chapter 5 on negligence, and Chapter 6 on hype. These details should be read by everyone, and should be required

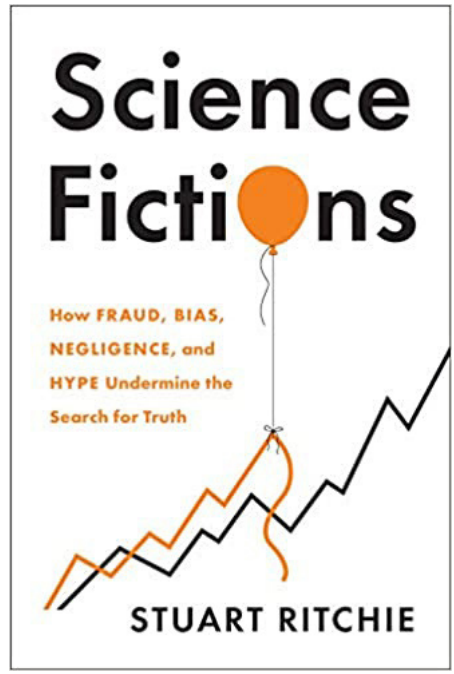
reading for all scientists, researchers, and physicians.

As to fraud, it is chilling and hair-raising to read of the failure of so distinguished an institution as Sweden's Karolinska Institute to deal with dishonest medical practices in its own bailiwick: artificial tracheas that damaged innumerable people (p. 48 ff.). Ritchie makes the excellent point that science's long record of trustworthiness "might, perversely, be what prevents it from spotting the bad actors in its midst" (p. 54). The very top journals, *Science* and *Nature*, had published fraudulent claims about cloning and about induced pluripotent (~stem) cells. It seems virtually inevitable that many *yet undiscovered* instances of deliberate fraud are present in less prestigious publications (p. 60). In that connection it should be noted that is there is a continuing spate of "predatory" journals established that ask authors to pay "processing" charges for prompt, supposedly peer-reviewed publication online, so that the publications are available to everyone, whereas the long-established professional journals are available only to people with access to academic and research libraries or to those who are prepared to pay not-insignificant amounts for individual published articles. The situation here is somewhat muddied as more and more of the traditional professional journals also offer authors the option of paying to have their articles immediately available online as "open access" on the journals' websites (p. 219).

Distinguishing fraud from honesty and genuine reports is complicated by the fact that bad actors can sometimes achieve genuine accomplishments: Woo Suk Hwang, the well-regarded Korean researcher who perpetrated many frauds including the claim of cloning a human embryo (p. 55), did in fact succeed in cloning an Afghan Hound (p. 57).

When statistical analysis is fundamentally involved, there are several approaches to detecting fraud; for example, "If a dataset looks too neat, too tidily similar across different groups," or if there are too few missing data-points (p. 63), or if the distribution of numbers might not be what is mathematically expected (p. 64). A lengthy endnote describes Benford's law, an empirical fact for which there seems to be no satisfactory explanation, that "the first significant digit of the numbers in many data sets is far more likely to be low than high" (p. 275).

A general reason why fraud in science is so unexpected is it that



the motives seem inscrutable. Thus, Ritchie describes a case in which it would have been easier for the perpetrator of fraud to have actually carried out the study rather than doing the intricate work needed to carry out the fraud (p. 65). In some cases it may be that a fraud reflects the perpetrator's utter belief that his views are correct and that producing the needed data would just be too much trouble (p. 71).

At any rate, the book is spot-on in pointing out how widespread and deep the damage can be from fraud, tarnishing the reputations of many innocent colleagues and polluting the scientific literature for a long time, because retractions do not become quickly or universally known so that fraudulent articles continue to be mistakenly cited as trustworthy (p. 74 ff.)

Chapter 4 grapples with the subtle, insidious factor of bias, which is a universal human trait. One source may be the desire to get clear, exciting results to support a pet theory, or to defeat a rival's claims (p. 83). Another is certainly the emphasis on publishing positive results only. This has the disadvantage that failed attempts to replicate published work are hidden from view (the so-called file-drawer problem), whereby work that cannot be replicated may continue to be cited as though it were meaningful. That is indeed a problem, but the suggestion that publication should be based not on the results of a study but on the soundness of its methodology (p. 85) may go too far; most people, after all, are interested in learning only of noteworthy results, and it is entirely rational for a journal to publish only what it believes may attract subscribers and readers.

Ritchie gives an excellent explanation (p. 86 ff.) of the need for some

objective criterion for deciding whether a given result is meaningful or only accidental. The widely used "$p \leq .05$" criterion is arbitrary as well as quite weak; usually the effect size is more usefully meaningful, and it is also important to note the size of the sample being tested. The explanation of meta-analysis is also excellent, including a discussion of the use of funnel plots to detect possible publication bias: Such bias can distort the literature because published work will be slanted toward large effect sizes. That can be particularly damaging in medicine, by misleading as to the appropriate use of biomarkers (p. 94) and the efficacy of drugs (p. 95).

The book does not cite the popular saw that "there are lies, damned lies, and statistics," or the version that "one can prove anything by statistics"; but several examples are given of how inappropriate or incompetent statistical analysis brings misleading publications into the literature. One way is through "*p*-hacking": Since the criterion of $p \leq .05$ is so vital for getting articles published, all sorts of fudging may be used to bring results below that level (p. 99). That this is quite common is suggested by studies that found a remarkably high proportion of published articles with *p*-values *just* below .05. A perhaps more subtle misdeed is HARKing, **H**ypothesizing **A**fter the **R**esults are **K**nown: If the original hypothesis is not substantiated, the data are searched for *any* significant correlation, and the article later submitted for publication pretends that this was the original hypothesis; .05 is 1 in 20, so one is likely to find an apparently significant correlation for every 20 attempts. Another pitfall for the unwary is *overfitting* of the data to an empirical curve (p. 108).

A variety of inappropriate or incompetent statistical approaches abound in the literature, found for instance even in the work of Brian Wansink, who had achieved more than nearly three decades of a truly distinguished career status (p. 98), until independent analyses of the data led to his resignation (p. 102). Clinical trials in medicine are increasingly suspect since they are paid for by self-interested drug companies (p. 110 ff.). "You might wonder how doctors and their patients are supposed to trust a medical literature that's permeated with bias . . . . I have no idea" (p. 112). When biases become shared among a whole community, it constitutes a dangerous *groupthink*, for example sustaining the belief that amyloid plaques are the cause of Alzheimer's

disease, against all the evidence that preventing or removing such plaques is entirely ineffective (p. 114). Political ideology is yet another possible source of bias.

The overall lesson, obvious but almost never learned, is that when statistical analysis is an integral part of research, independent professional statisticians ought to be consulted—*and their advice followed—in the initial planning of research protocols* as well as in the analysis of results (p. 209).

Chapter 5 deals with negligence. Some amusing and disheartening examples are given, and the common-sense point is made that published numbers should not obviously make no sense. It would always be useful to ask: How could such a conclusion come to be known? Was the sample large enough for adequate statistical power (p. 134)?

It is again disheartening to read that "large-scale reviews have . . . found under-powered research . . . [to be] rife in medical trials, biomedical research more generally, economics, brain imaging, nursing research, behavioral ecology, and . . . psychology" (p. 137). Claims to identify genes responsible for IQ scores, depression, schizophrenia, and a variety of other (usually behavioral) traits have almost invariably turned out to result from under-powered statistics (p. 141). It seems to have been forgotten that the founder of *p*-value statistics, Ronald Fisher, pointed out a century ago that "complex traits must be massively *polygenic* . . . related to many thousands of small-effect genes" (p. 142).

Institutions and individuals routinely exaggerate the import of claimed findings in order to gain public attention and funding. Chapter 6 describes several egregious instances of this hype, including the claimed discovery by NASA researchers of a life-form in which arsenic atoms replace phosphorus, something that seems impossible a priori; and indeed it was a mistaken claim. How very common hype has become is shown by a number of studies that found in scientific papers a striking increase since the 1970s in the use of such words as "innovative," "promising," "unique," "unprecedented." (Examining the frequency of word-usage in non-technical fields is facilitated by Google's "N-gram" application that searches a vast number of digitized books from as far back as when books were first printed.)

Hype, even when merely exaggeration rather than plainly false, risks causing damage to the credibility of scientists, scientific institutions, and science itself. Common instances include unwarranted advice, for example as to diet or exercise, and ballyhooing supposed medical advances on the basis of results only in cells in test tubes or in experimental animals. Just as with fraud and with innocent mistakes, the damage continues long after the scientific knowledge itself has been corrected, as the corrections are not hyped as emphatically as the initial claims had been. The detailed examples given in Chapter 6 deserve to be read by everyone. The most egregious hype is often associated with so-called emerging fields: stem cells, genetics, epigenetics, machine learning, brain imaging, the microbiome (the countless millions of microbes that infest our bodies) (pp. 160–161).

The book has a fine summary at the end of Chapter 6:

> even though caution, restraint and skepticism are basic virtues of science, we have a system that incentivizes the precise opposite. Scientists are pushed into publishing as many papers as possible, and hyping them up to the high heavens, by an academic system that's become an impediment to getting science right. (p. 172)

Part III of the book, "Causes and Cures," adds little to that summary. Chapter 7 spells out "Perverse Incentives" that have already been indicated throughout earlier chapters: Seeking prestige and wealth and careers, institutions and individuals both practice excessive publication and hype. A misdeed not mentioned earlier is the faking of peer review (p. 185 ff.) or of an individual's "h-index" (p. 187 f.), which was originally designed, like the "impact factor" (p. 190) for journals, to measure quality rather than quantity. Neither does what was hoped, as could have been foreseen under Goodhart's Law: "When a measure becomes the target, it ceases to be a good measure" (p. 192). It ought to be obvious that objective numerical measurements and calculations can never substitute for human judgment of *quality* or *value*.

I regret being unable to say anything good about Chapter 8, "Fixing Science." "The problems science faces are systemic, indicating an entire culture gone awry" (p. 190). Indeed. And that culture has gone awry not by itself but through the influence on it of the wider

society, in which wealth and power and influence and status are the desired outcomes, not at all what science supposedly seeks—the best achievable, most objective understanding of the material world. Science and scientists depend on the wider society for their resources, and the values and aims of that wider society thereby come to govern scientific activity. Trustworthy science depends on the honesty and integrity of researchers (p. 21), but that is not what contemporary society rewards. The fact that science publishing has become dominated by commercial businesses like Elsevier and Springer makes it whistling in the wind to suggest changes in publication practices toward making reliability and truthfulness the prime objectives.

Science is a human and social activity, as Ritchie pointed out early in the book. The overarching consequence is that there is really only one feasible way to cure contemporary dysfunctions: to remove the incentives that cause the regrettable behavior. But the present system of incentives is the same as in society at large: Success means to achieve status, influence, wealth, power. Researchers can become successful only by attaining what the wider society values. Ritchie himself recognizes that: "All we need to fix science is to give people the right motivation" (p. 234).

The book quite properly describes as desiderata the Mertonian Norms of "universalism," "communality," and "organized skepticism"; but these had been identified by sociologist Robert Merton around 1940, and had been reasonably achievable in practice only in the "good old days" when pure science was an ivory-tower activity largely independent of outside influences, a cottage industry of independent intellectual entrepreneurs, of comparatively little interest to the wider society and depending hardly at all on the wider society for needed resources; "pure" scientific research was then carried on with all or almost all the needed resources provided by the researchers' universities. World War II marked a sea change and a turning point (Bauer 2017, Chapter 1), as outside funding of research increased exponentially, culminating in the dysfunctional present state of affairs in which, as John Ziman (1994) pointed out, the Mertonian Norms had to be augmented with "originality," the incentive to produce something positively original, and where there is no longer any meaningful distinction between "pure" and "applied" research.

QUIBBLES AND CAVEATS

I recommend this book unreservedly. Its criticisms of contemporary scientific practices are appropriate, sound, superbly documented. It should be seen as a compliment to the book's meaty discussions that some quibbles and caveats and points of disagreement seem called for.

Although what Ritchie says does apply to science overall, there is a pervasive emphasis on statistical analysis, almost all of which is scarcely relevant to physical science, where replicability and proof and disproof are often much more straightforward. This bears noting because society's high regard for the trustworthiness of scientific facts, knowledge, and understanding rests primarily on the achievements of astronomy, physics, and chemistry, and it is important to realize that nothing like the certainty attainable in those can be matched by medical science or by the behavioral and social sciences where only knowledge of a probabilistic, statistical nature is attainable.

Regarding statistics, the Bayesian approach is given too short shrift. That the Bayesian prior is inherently subjective (p. 207 f.) doesn't matter; one can start with an arbitrary prior probability of 0.5, and as evidence accumulates the Bayesian method brings the calculated probability closer and closer to the objectively sustainable ("true") one. Matthews (1998, 1999) has explained why the Bayes method is even *in principle* preferable to the frequentist *p*-value approach.

Ritchie is spot-on in describing science as socially constructed in the sense that subjective interactions lead to less-subjective consensus; but Ritchie is not a postmodernist relativist constructionist of the former Edinburgh cult. But he might well have added that *time* is a crucial element in making science more reliable.

As to "in recent years it's become increasingly, painfully obvious that peer review is far from the guarantee of accuracy and reliability it's cracked up to be" (p. 15), actually that was obvious half a century ago to many practicing researchers; a wide-ranging discussion was published four decades ago (Peters & Ceci, 1982).

The lack of a separate bibliography is a minor inconvenience; the citations in the endnotes use a style common in the humanities, where once a reference has been cited in full it is subsequently repeated only by something like "Bauer, Dogmatism, pp. . . .", leaving a reader to search

earlier notes for full details of the source. It is not true, incidentally, that retracted articles always remain available on a journal's website (p. 66); it might be just the abstract or the original citation that remains (for example, Goodson, 2014).

That Andrew Wakefield's claims were fraudulent about autism and the MMR (measles mumps rubella) vaccine (pp. 76, 225) remains controversial (Canary Party, 2015). Such terms as "vaccine skeptic" (p. 78 ff.) or "anti-vaxxer" are pejorative and unwarranted; in most cases, "vaccine skeptics" are not opposed to *vaccination as such* but rather point to the unsatisfactory and often damaging so-called "side" effects of some vaccines, for example Cervarix and Gardasil against HPV (Holland et al., 2018).

The book is wrong in accepting the view that HIV is sexually transmitted (p. 239) and that HIV causes AIDS (p. 244); and in accepting the mainstream views on climate change (pp. 13, 34). Those last points drive home the danger warned against by President Eisenhower (1961): "in holding scientific research and discovery in respect, as we should, we must also be alert to the equal and opposite danger that public policy could itself become the captive of a scientific–technological elite."

It is instructive and also dismaying that so well-read and insightful an author as Stuart Ritchie has been misled in this way on topics where the mainstream consensus is clearly contradicted by profuse evidence in the professional literature (Bauer, 2012, pp. 18–29). Active researchers like Ritchie have no practical alternative to trusting what researchers in other fields conclude; and even someone so aware as Ritchie of the dysfunctions of contemporary science has not yet realized that public policies on important matters can be as badly mistaken nowadays as a century ago when accepted expert opinion about eugenics led to the forced sterilization of tens of thousands of Americans during about half of the 20th century.

Such public tragedies may be avoidable only if the scientific community learns to examine without preconception the possible merits of minority views, and if a truly impartial authority is established to provide governments, and also non-governmental institutions, with as objective and unbiased as possible an assessment of the relative merits of mainstream views and minority dissenting views—for instance a Science Court (Bauer, 2017, Chapter 12).

REFERENCES

Bauer, H. H. (2012). *Dogmatism in science and medicine: How dominant theories monopolize research and stifle the search for truth.* McFarland.

Bauer, H. H. (2017). *Science is not what you think: How it has changed, why we can't trust it, how it can be fixed.* McFarland.

Canary Party. (2015, October 1). *Timeline of events in the William Thompson CDC whistleblower scandal.* https://canaryparty.org/commentary/timeline-of-events-in-the-william-thompson-cdcwhistleblower-scandal

Eisenhower, D. D. (1961, January 17). Presidential farewell speech. *Public Papers of the Presidents, Dwight D. Eisenhower, 1960*, pp. 1035–1040. http://avalon.law.yale.edu/20th_century/eisenhower001.asp

Goodson, P. (2014, September 23). Questioning the HIV-AIDS hypothesis: 30 years of dissent. *Frontiers in Public Health, 2*, 154. https://doi.org/10.3389/fpubh.2014.00154. Retraction at https://www.frontiersin.org/articles/10.3389/fpubh.2014.00154/full

Holland, M., Rosenberg, K. M., & Iorio, E. (2018). *The HPV vaccine on trial: Seeking justice for a generation betrayed.* Skyhorse.

Matthews, R. A. J. (1998). Facts versus factions: The use and abuse of subjectivity in scientific research. *European Science and Environment Forum Working Paper*. https://hivskeptic.files.wordpress.com/2010/10/matthews.pdf
Also published in 2000 in J. Morris (Ed.), *Rethinking risk and the precautionary principle* (pp. 247–282), Butterworth-Heinemann.

Matthews, R. A. J. (1999). Significance levels for the assessment of anomalous phenomena. *Journal of Scientific Exploration, 13*(1), 1–7.

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5*(2), 187–195. https://doi.org/10.1017/S0140525X00011183

Ziman, J. (1994). *Prometheus bound: Science in a dynamic steady state.* Cambridge University Press.