

The margins of reproducibility

Charles Day

In 2016 *Nature* published the results of polling scientists about their views on a central tenet of science: reproducibility. Shockingly, 70% of the 1576 respondents said they had tried and failed to reproduce another scientist's experiments; 52% said that irreproducibility constituted a significant crisis.

Physicists and engineers were the most confident in the reproducibility of published results. Perhaps because of that confidence, they were also the least likely, at a rate of 24%, to have taken concrete steps to improve reproducibility. Given that some of the most notorious cases of irreproducibility in science have been perpetrated by physicists, it's important to examine the causes and propose remedies.

The most flagrant cases are the most straightforward to address. When Jan Hendrik Schön fabricated the results of experiments on organic semiconductors in the early 2000s, he surely knew that what he was doing was wrong. I doubt ethics training, even if he had received any, would have made a difference. What could have checked him is foreknowledge of the shame and sanctions he brought on himself.

Schön set out to deceive. Other researchers have sincerely believed in their theories or experiments—and then persisted in advocating their conclusions even after contrary evidence has convinced almost everyone else of the conclusions' invalidity. Forty years ago, Stanley Pons and Martin Fleischmann trumpeted their measurement of a temperature increase in an electrochemical cell, which they attributed to the fusion of deuterium to make tritium or helium-3. The widespread failure to reproduce the results or to plausibly explain them convinced most physicists that cold fusion is spurious. Pons, Fleischmann, and others who study low-energy nuclear reactions remain unswayed.

Sometimes mavericks are right. Alfred Wegener's theory of continental drift was eventually confirmed. Sincere belief in the face of opposing evidence is challenging to classify as unethical, except when attributable to the deliberate and dishonest mishandling or misinterpretation of data. A bitter dispute about supercooled water's structure just above the temperature at which it freezes came down to an unphysical assumption buried in a subroutine. (See "The war over supercooled water" by Ashley Smart, *PHYSICS TODAY* online, 22 August 2018.)

The largest number of cases of irreproducible research likely arise from scientists' pushing at the margins of what is technically and statistically feasible. In the July 2008 issue of *PHYSICS TODAY* (page 12), I reported on a survey of quasars that found evidence in their UV spectra of a hot phase of intergalactic plasma. The finding was newsworthy because the plasma

could account for 40% of the baryons missing at low redshifts and suspected to lurk in filamentary structures connecting clusters of galaxies. What of the remaining 60%? Astronomers anticipated finding the baryons in even hotter, x-ray-emitting plasma. "New missions, such as NASA's *Constellation-X* and ESA's *X-ray Evolving Universe Spectroscopy*, are expected to find some of them," I wrote.

Neither mission has launched. I was surprised, therefore, to encounter papers in 2018 and 2019 that claimed to have discovered the even hotter plasma in data gathered by two spacecraft, ESA's *XMM-Newton*¹ and NASA's *Chandra X-Ray Observatory*,² that were in orbit when I wrote the 2008 story. Both studies abutted the limits of statistical significance and relied on challengeable assumptions. They are not invalid, however. Both groups fully described their methods, and thanks to the widespread practice in space-based astronomy of making data publicly available, astronomers are free to repeat the analysis.

Ernest Rutherford's glib remark, "If your experiment needs statistics, you ought to have done a better experiment," is not helpful when the better experiment is years away. What's needed, I think, is better instruction in data analysis. Mine came indirectly from Allyn Tennant, then a postdoc, who urged me to buy and read Philip Bevington's book, *Data Reduction and Error Analysis for the Physical Sciences* (1969). The book remained my *vade mecum* throughout my career as an astronomer.

In their November 2004 *PHYSICS TODAY* article, "Ethics and the welfare of the physics profession" (page 42), Kate Kirby and Frances Houle cited a survey of physics graduate students and postdocs that was conducted by Roman Czerwinski of the American Institute of Physics (publisher of *PHYSICS TODAY*). It found that only half had received training on acceptable ways to interpret and analyze data.

Physicists should continue to publish marginally significant results. But when they do so, the statistical analysis should be sound, transparent, and reproducible.



References

1. F. Nicastro et al., *Nature* **558**, 406 (2018).
2. O. E. Kovács et al., *Astrophys. J.* **872**, 83 (2019).

PT