FACILITATING FALSEHOOD

The Costs of Researcher Freedom

Figure 15: Flexible Research Design

Why do researchers get away with sloppy science? In part because, far too often, no one is watching and no one is there to stop them. We think of freedom as a good thing, but in the realm of scientific experimentation, *uncontrolled researcher freedom* makes it easy for scientists to err in all the ways described above.[74] The fewer the constraints on scientists' research designs, the more opportunities for malfeasance—and, as it turns out, a lot of scientists will go astray, deliberately or accidentally. For example, lack of constraints allows researchers to alter their methods midway through a study—changing hypotheses, stopping or recommencing data collection, redefining variables, "fine-tuning" statistical models—as they pursue publishable, statistically significant results. Researchers often justify midstream alteration of research procedures as flexibility or openness to new evidence[75]—but in practice such "flexibility" frequently subserves scientists' unwillingness to accept a negative result.

Researchers sometimes have good reasons to alter a research design before a study is complete—for example, if a proposed drug in a clinical trial appears to be causing harm to the experimental subjects.[76] (Though scientists can take even this sort of decision too hastily.[77]) But researchers also stop some clinical trials early on the grounds that a treatment's benefits are already apparent and that it would be wrong to continue denying that treatment to the patients in the control group. Such truncated clinical trials pose grave ethical hazards: as one discussion put it, truncated trials "systematically overestimate treatment effects" and can violate "the ethical research requirement of scientific validity."[78] Moreover, a 2015 article in the *Journal of Clinical Epidemiology* indicated that "most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules."[79] In other words, most decisions to discontinue were done on the fly, without regard for the original research design. The researchers changed methodology midstream.

Simmons and his co-authors demonstrated their point by running an experiment to see if listening to selected songs will make you, literally, younger. Their flexible research design produced data that revealed an effect of 18 months, with $p = .040$.

A now-famous 2011 article by Simmons, Nelson, and Simonsohn estimated that providing four "degrees of researcher freedom"—four ways to shift the design of an experiment while it is in progress—can lead to a 61% false-positive rate. Or, as the subtitle of the article put it, "Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." Simmons and his co-authors demonstrated their point by running an experiment to see if listening to selected songs will make you, literally, younger. Their flexible research design produced data that revealed an effect of 18 months, with $p = .040$.[80]

Absence of Openness

Lack of openness also contributes to the reproducibility crisis. Investigators far too rarely share data and methodology once they complete their studies. Scientists ought to be able to check and critique one another's work, but many studies can't be evaluated properly because researchers don't make their data and procedures available to the public. We've seen that small changes in research design can have large effects on researchers' conclusions. Yet once scientists publish their research, those small changes vanish from the record, and leave behind only the statistically significant result. For example, the methods used in meta-analyses to harmonize cognitive measures across data sets "are rarely reported."[81] But someone reading the results of a meta-analysis can't understand it properly without a detailed description of the harmonization methods and of the codes used in formatting the data.

Moreover, data sets often come with privacy restrictions, usually to protect personal, commercial, or medical information. Some restrictions make sense—but others don't. Sometimes unreleased data sets simply vanish—for example, those used in environmental science.[82] Data sets can disappear because of archival failures, or because of a failure to plan how to transfer data into new archival environments that will provide reliable storage and continuing access. In either case, other researchers lose the ability to examine the underlying data and verify that it has been handled properly.

> In February 2017, a furor that highlighted the problem of limited scientific openness erupted in the already contentious field of climate science. John Bates, a climate scientist who had recently retired from the National Oceanic and Atmospheric Administration (NOAA), leveled a series of whistleblowing accusations at his colleagues.[83] He focused on the failure by Tom Karl, the head of NOAA's National Centers for Environmental Information, to archive properly the dataset that substantiated Karl's 2015 claim to refute evidence of a global warming hiatus since the early 2000s.[84] Karl's article had been published shortly before the Obama administration submitted its Clean Power Plan to the 2015 Paris Climate Conference, and it had received extensive press coverage.[85] Yet Karl's failure to

archive his dataset violated NOAA's own rules—and also the guidelines of *Science*, the prestigious journal that had published the article. Bates' criticisms touched off a political argument about the soundness of Karl's procedures and conclusions, but the data's disappearance meant that no scientist could re-examine Karl's work. Supporters and critics of Karl had to conduct their argument entirely in terms of their personal trust in Karl's professional reliability. Practically, the polarized nature of climate debate meant that most disputants believed or disbelieved Karl depending upon whether they believed or disbelieved his conclusions. Science should not work that way—but without the original data, scientific inquiry could not work at all.

Both scientists and the public should regard skeptically research built upon private data. Gelman responded appropriately, if sarcastically, to Wansink's refusal to share his data on privacy grounds:

> *Some people seem to be upset that Wansink isn't sharing his data. If he doesn't want to share the data, there's no rule that he has to, right? It seems pretty simple to me: Wansink has no obligation whatsoever to share his data, and we have no obligation to believe anything in his papers. No data, no problem, right?*[86]

"Wansink has no obligation whatsoever to share his data, and we have no obligation to believe anything in his papers. No data, no problem, right?" – Andrew Gelman