

***Surrogate Science:
How Fisher, Neyman-Pearson, and Bayes Were Transformed into the Null Ritual***

Gerd Gigerenzer

If statisticians agree on one thing, it is that scientific inference should not be made mechanically. Despite virulent disagreements on other issues, Ronald Fisher and Jerzy Neyman, two of the most influential statisticians of the 20th century, were of one voice on this matter. Good science requires both statistical tools and informed judgment about what model to construct, what hypotheses to test, and what tools to use. Practicing statisticians rely on a “statistical toolbox” and on their expertise to select a proper tool. Social scientists, in contrast, tend to rely on a single tool.

In this talk, I trace the historical transformation of Fisher’s null hypothesis testing, Neyman-Pearson decision theory, and Bayesian statistics into a single mechanical procedure that is performed like compulsive hand washing: the null ritual. In the social sciences, this transformation has fundamentally changed research practice, making statistical inference its centerpiece. The essence of the null ritual is:

1. Set up a null hypothesis of “no mean difference” or “zero correlation.” Do not specify the predictions of your own research hypothesis.
2. Use 5% as a convention for rejecting the null. If significant, accept your research hypothesis. Report the result as $p < .05$, $p < .01$, or $p < .001$, whichever comes next to the obtained p-value.
3. Always perform this procedure.

I use the term “ritual” because this procedure shares features that define social rituals: (i) the repetition of the same action, (ii) a focus on special numbers or colors, (iii) fears about serious sanctions for rule violations, and (iv) wishful thinking and delusions that virtually eliminate critical thinking. The null ritual has each of these four characteristics: mindless repetition; the magical 5% number, fear of sanctions by editors or advisors, and delusions about what a p-value means, which block researchers’ intelligence. Starting in the 1940s, writers of bestselling statistical textbooks for the social sciences have silently transformed rivaling statistical systems into an apparently monolithic method that could be used mechanically. The idol of a universal method for scientific inference has been worshipped and institutionalized since the “inference revolution” of the 1950s. Because no such method has ever been found, surrogates have been created, most notably the quest for significant p-values. I show that this form of surrogate science fosters delusions and argue that it is one of the reasons of “borderline cheating” which has done much harm, creating, for one, a flood of irreproducible results in fields such as psychology, cognitive neuroscience and tumor marker research.

Today, proponents of the “Bayesian revolution” are in a similar danger of chasing the same chimera: an apparently universal inference procedure. A better path would be to promote an understanding of the various devices in the “statistical toolbox.” I discuss possible explanations why a toolbox approach to statistics has been so far successfully prevented by journal editors, textbook writers, and social scientists.