Souvenir K: Probativism

> [A] fundamental tenet of the conception of inductive learning most at home
> with the frequentist philosophy is that inductive inference requires building
> up incisive arguments and inferences by putting together several different
> piece-meal results … the payoff is an account that approaches the kind of
> full-bodied arguments that scientists build up in order to obtain reliable
> knowledge and understanding of a field. (Mayo and Cox 2006, p. 82)

The error statistician begins with a substantive problem or question. She
jumps in and out of piecemeal statistical tests both formal and quasi-formal.
The pieces are integrated in building up arguments from coincidence, inform-
ing background theory, self-correcting via blatant deceptions, in an iterative
movement. The inference is qualified by using error probabilities to determine
not "how probable," but rather, "how well-probed" claims are, and what has
been poorly probed. What's wanted are ways to measure how far off what
a given theory says about a phenomenon can be from what a "correct" theory
would need to say about it by setting bounds on the possible violations.

An account of testing or confirmation might entitle you to confirm, support,
or rationally accept a large-scale theory such as GTR. One is free to reconstruct
episodes this way – after the fact – but as a forward-looking account, they fall far
short. Even if somehow magically it was known in 1960 that GTR was true, it
wouldn't snap experimental relativists out of their doldrums because they still
couldn't be said to have understood gravity, how it behaves, or how to use one
severely affirmed piece to opportunistically probe entirely distinct areas.

---

[11] Some recommend "equivalence testing" where H_0: $\mu \geq \mu_0$ or $\mu \leq -\mu_0$ and rejecting both sets
bounds on μ. One might worry about low-powered tests, but it isn't essentially different from
setting upper bounds for a more usual null. (For discussion see Lakens 2017, Senn 2001a, 2014,
R. Berger and Hsu 1996, R. Berger 2014, Wellek 2010).

Learning from evidence turns not on appraising or probabilifying large-scale theories but on piecemeal tasks of data analysis: estimating backgrounds, modeling data, and discriminating signals from noise. Statistical inference is not radically different from, but is illuminated by, sexy science, which increasingly depends on it. Fisherian and N-P tests become parts of a cluster of error statistical methods that arise in full-bodied science. In Tour II, I'll take you to see the (unwarranted) carnage that results from supposing they belong to radically different philosophies.