

### Souvenir L: Beyond Incompatibilist Tunnels

What people take away from the historical debates is Fisher (1955) accusing N-P, or mostly Neyman, of converting his tests into acceptance sampling rules more appropriate for five-year plans in Russia, or making money in the USA, than for science. Still, it couldn't have been too obvious that N-P distorted his tests, since Fisher tells us only in 1955 that it was Barnard who explained that, despite agreeing mathematically in very large part, there is this distinct philosophical position. Neyman suggests that his terminology was to distinguish what he (and Fisher!) were doing from the attempts to define a unified rational measure of belief on hypotheses. N-P both denied there was such a thing. Given Fisher's vehement disavowal of subjective Bayesian probability, N-P thought nothing of crediting Fisherian tests as a step in the development of "inductive behavior" (in their 1933 paper).

The myth of the radical difference in either methods or philosophy is a myth. Yet, as we'll see, the hold it has over people continues to influence the use and discussion of tests. It's based almost entirely on sniping between Fisher and Neyman from 1935 until Neyman leaves for the USA in 1938. Fisher didn't engage much with statistical developments during World War II. Barnard describes Fisher as cut off "by some mysterious personal or political agency. Fisher's isolation occurred, I think, at a particularly critical

<sup>5</sup> Pearson said that a statistician has an  $\alpha$  and a  $\beta$  side, the former alludes to what they say in theory, the latter to what they do in practice. In practice, even Neyman, so often portrayed as performance-oriented, was as inferential as Pearson.

time, when opportunities existed for a fruitful fusion of ideas stemming from Neyman and Pearson and from Fisher" (Barnard 1985, p. 2). Lehmann observes that Fisher kept to his resolve not to engage in controversy with Neyman until the highly polemical exchange of 1955 at age 65. Fisher alters some of the lines of earlier editions of his books. For instance, Fisher's disinterest in the attained *P*-value was made clear in *Statistical Methods for Research Workers* (SMRW) (1934a, p. 80):

... in practice we do not want to know the exact value of *P* for any observed value of [the test statistic], but, in the first place, whether or not the observed value is open to suspicion.

If *P* is between .1 and .9 there is certainly no reason to suspect the hypothesis tested. If it is below .02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05.

Lehmann explains that it was only "fairly late in life, Fisher's attitude had changed" (Lehmann 2011, p. 52). In the 13th edition of SMRW, Fisher changed his last sentence to:

The actual value of *P* obtainable ... indicates the strength of the evidence against the hypothesis. [Such a value] is seldom to be disregarded. (p. 80)

Even so, this at most suggests how the methodological (error) probability is thought to provide a measure of evidential strength – it doesn't abandon error probabilities. There's a deeper reason for this backtracking by Fisher; I'll save it for Excursion 5. One other thing to note: F and N-P were creatures of their time. Their verbiage reflects the concern with "operationalism" and "behaviorism," growing out of positivistic and verificationist philosophy. I don't deny the value of tracing out the thrust and parry between Fisher and Neyman in these excursions. None of the founders solved the problem of an inferential interpretation of error probabilities – though they each offered tidbits. Their name-calling: "you're too mechanical," "no *you* are," at most shows, as Gigerenzer and Marewski observe, that they all rejected mechanical statistics (2015, p. 422).

The danger is when one group's interpretation is the basis for a historically and philosophically "sanctioned" reinterpretation of one or another method. Suddenly, rigid rules that the founders never endorsed are imposed. Through the Incompatibilist philosophical tunnel, as we are about to see, these reconstruals may serve as an effective way to dismiss the entire methodology – both F and N-P. After completing this journey, you shouldn't have to retrace this "he said/they said" dispute again. It's the methods, stupid.