Excursion 3 Tour I
We move from Popper through a gallery on "Data Analysis in the 1919 Eclipse tests of the General Theory of Relativity (GTR)" (3.1) which leads to the main gallery on the origin of statistical tests (3.2) by way of a look at where the main members of our statistical cast are in 1919: Fisher, Neyman and Pearson. From the GTR episode, we identify the key elements of a statistical test–the steps in E.S. Pearson's opening description of tests in 3.2. The classical testing notions–type I and II errors, power, consistent tests–are shown to grow out of requiring probative tests. The typical (behavioristic) formulation of N-P tests came later. The severe tester breaks out of the behavioristic prison. A first look at the severity construal of N-P tests is in [Exhibit (i)](). Viewing statistical inference as severe testing shows how to do all N-P tests do (and more) while a member of the Fisherian Tribe (3.3). We consider the frequentist principle of evidence FEV and the divergent interpretations that are called for by Cox's taxonomy of null hypotheses. The last member of the taxonomy–substantively based null hypotheses–returns us to the opening episode of GTR.

key terms
GTR, eclipse test, ether effect, corona effect, PPN framework, statistical test ingredients, Anglo-Polish collaboration, Lambda criterion; Type I error, Type II error, power, P-value, unbiased tests, consistent tests uniformly most powerful (UMP); severity interpretation of tests, severity function, water plant accident; sufficient statistic; frequentist principle of evidence FEV; sensitivity achieved, [same as attained power (att power)], Cox's taxonomy (embedded, nested, dividing, testing assumptions), Nordvedt effect, equivalence principle (strong and weak)