

## **Excursion 2 Tour II: Falsification, Pseudoscience, Induction\***

*Blurb of Tour.* Tour II visits Popper, falsification, corroboration, Duhem's problem (what to blame in the case of anomalies) and the demarcation of science and pseudoscience (2.3). While Popper comes up short on each, the reader is led to improve on Popper's notions (live exhibit (v)). Central ingredients for our journey are put in place via souvenirs: a framework of models and problems, and a post-Popperian language to speak about inductive inference. Defining a severe test, for Popperians, is linked to when data supply novel evidence for a hypothesis: family feuds about defining novelty are discussed (2.4). We move into Fisherian significance tests and the crucial requirements he set (often overlooked): isolated significant results are poor evidence of a genuine effect, and statistical significance doesn't warrant substantive, e.g., causal inference (2.5). Applying our new demarcation criterion to a plausible effect (males are more likely than females to feel threatened by their partner's success), we argue that a real revolution in psychology will need to be more revolutionary than at present. Whole inquiries might have to be falsified, their measurement schemes questioned (2.6). The Tour's pieces are synthesized in (2.7), where a guest lecturer explains how to solve the problem of induction now, having redefined induction as severe testing.

### **Excursion 2 Tour II: keywords**

Corroboration, Demarcation of science and pseudoscience, Falsification, Duhem's problem  
Novelty, Biasing selection effects, Simple significance tests, Fallacies of rejection, NHST,  
Reproducibility and replication

### Notes on 2.3

There are four key, interrelated themes from Popper:

**(1) Science and Pseudoscience.** For a theory to be scientific it must be testable and falsifiable.

**(2) Conjecture and Refutation.** We learn not by enumerative induction but by trial and error: conjecture and refutation.

**(3) Observations Are Not Given.** If they are at the “foundation,” it is only because there are apt methods for testing their validity. We dub claims observable *because* or to the extent that they are open to stringent checks.

**(4) Corroboration Not Confirmation, Severity Not Probabilism.** Rejecting probabilism, Popper denies scientists are interested in highly probable hypotheses (in any sense). They seek bold, informative, interesting conjectures and ingenious and severe attempts to refute them.

These themes are in the spirit of the error statistician. Considerable spade-work is required to see *what to keep and what to revise*, so bring along your archeological shovels.

#### **The Severe Tester Revises Popper’s Demarcation of Science (Live Exhibit**

**(vi)):** What he should be asking is not whether a theory is unscientific, but *When is an inquiry into a theory, or an appraisal of claim H, unscientific?* We want to distinguish meritorious modes of inquiry from those that are BENT. If the test methods enable ad hoc maneuvering, sneaky face- saving devices, then the inquiry – the handling and use of data – is unscientific. Despite being logically falsifiable, theories can be rendered immune from falsification by means of questionable methods for their testing.

**Greater Content, Greater Severity.** The severe tester accepts Popper’s central intuition in (4): if we wanted highly probable claims, scientists would stick to low-level observables and not seek generalizations, much less theories with high explanatory content. A highly explanatory, high-content theory, with interconnected tentacles, has a higher probability of having flaws discerned than low-content theories that do not rule out as much. Thus, when the bolder, higher content, theory stands up to testing, it may earn higher overall severity than the one with measly content. It is the fuller, unifying, theory developed in the course of solving interconnected problems that enables severe tests.

**Methodological Probability.** Probability in learning attaches to a method of conjecture and refutation, that is to testing: it is *methodological probability*. An error probability is a special case of a methodological probability. We want methods with a high probability of teaching us (and machines) how to distinguish approximately correct and incorrect interpretations of data. That a theory is plausible is of little interest, in and of itself; what matters is that it is *implausible* for it to have passed these tests were it false or incapable of adequately solving its set of problems.

**Methodological falsification.** We appeal to methodological rules for when to regard a claim as falsified.

- Inductive-statistical falsification proceeds by methods that allow  $\sim H$  to be inferred with severity. A first step is often to infer an anomaly is real, by falsifying a “due to chance” hypothesis.

- Going further, we may corroborate (i.e., infer with severity) effects that count as falsifying hypotheses. A *falsifying hypothesis* is a hypothesis inferred in order to falsify some other claim. Example: the pathological proteins (prions) in mad cow disease infect without nucleic acid. This falsifies: all infectious agents involve nucleic acid.

Despite giving lip service to testing and falsification, many popular accounts of statistical inference do not embody falsification – even of a statistical sort.

However, the falsifying hypotheses that are integral for Popper also necessitate an evidence-transcending (inductive) statistical inference. If your statistical account denies we can reliably falsify interesting theories because doing so is not strictly deductive, it is irrelevant to real-world knowledge.

### **The Popperian (Methodological) Falsificationist Is an Error Statistician**

When is a statistical hypothesis to count as falsified? Although extremely rare events may occur, Popper notes:

*such occurrences would not be physical effects, because, on account of their immense improbability, they are not reproducible at will ... If, however, we find reproducible deviations from a macro effect ... deduced from a probability estimate ... then we must assume that the probability estimate is falsified. (Popper 1959, p. 203)*

In the same vein, we heard Fisher deny that an “isolated record” of statistically significant results suffices to warrant a reproducible or genuine effect (Fisher 1935a, p. 14).

### **In a sense, the severe tester ‘breaks’ from Popper by solving his key**

**problem:** Popper’s account rests on severe tests, tests that would probably falsify claims if false, but he cannot warrant saying a method is probative or severe, because that would mean it was reliable, which makes Popperians squeamish. It would appear to concede to his critics that Popper has a “whiff of induction” after all. But it’s not inductive enumeration. Error statistical methods (whether from statistics or informal) can supply the severe tests Popper sought.

### **A scientific inquiry (a procedure for finding something out) for a severe tester:**

- blocks inferences that fail the minimal requirement for severity:
- *must be able to embark on a reliable probe to pinpoint blame for anomalies (and use the results to replace falsified claims and build a repertoire of errors).*

The parenthetical remark isn’t absolutely required, but is a feature that greatly strengthens scientific credentials.

The reliability requirement is: infer claims just to the extent that they pass severe tests. There’s no sharp line for demarcation, but when these requirements are absent, an inquiry veers into the realm of questionable science or pseudoscience.