Day Three: Falsification, Pseudoscience, Induction: Popper (Excursion 2 Tour II)

We move from Popperian falsification to significance tests, and from Popper’s demarcation to current-day problems of pseudoscience and irreplication.

An excerpt from our Museum guide gives a broad-brush sketch of the start of II

Karl Popper had a brilliant way to “solve” the problem of induction: Hume was right that enumerative induction is unjustified, but science is a matter of deductive falsification. Science was to be demarcated from pseudoscience according to whether its theories were testable and falsifiable. A hypothesis is deemed severely tested if it survives a stringent attempt to falsify it. Popper’s critics denied he could sustain this and still be a deductivist….. (SIST, p. 75)

2.3 Popper, Severity, and Methodological Probability

Here's Popper's summary (drawing from Popper, *Conjectures and Refutations*, 1962, p. 53):

- [Enumerative] induction…is a myth. It is neither a psychological fact...nor one of scientific procedure
- The actual procedure of science is to operate with conjectures…
- Repeated observation and experiments function in science as tests of our conjectures or hypotheses, i.e., as attempted refutations
- [It is wrongly believed that using the inductive method can] *serve as a criterion of demarcation between science and pseudo-science*. …None of this is altered in the least if we say that induction makes theories only probable.

The themes are well known, but I want to revisit them for getting at key issues in the statistical crisis of science

Conjecture and Refutation. The problem of induction is a problem only if it depends on an unjustifiable procedure such as enumerative induction. *(SIST p. 76)*

- Popper denies scientists were in the habit of inductively enumerating.
- It doesn't even hold up on logical grounds: "another instance of an A that is a B" assumes a conceptual classification scheme.
- The actual procedure for learning in science is to operate with conjectures in which we then try to find weak spots and flaws.

- Deductive logic is needed to draw out the remote logical consequences that we actually have a shot at testing (ibid., p. 51).

Methodological Probability*.* A valuable idea to take from Popper is that probability in learning attaches to a method of conjecture and refutation, to testing: it is *methodological probability* (p. 80)

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

“epistemology” vs “variability” shoehorn: The choices for probability commonly offered are stark: “in here” (beliefs ascertained by introspection) or “out there” (frequencies in long-runs, or chance mechanisms).

We reject this (Excur 1, Souv. (D)).

Falsification is rarely deductive (outside of philosophy classes: all swans are white)

Though there are analogous cases, I will argue, in testing assumptions

How can good science be all about falsification? My answer is: we can erect reliable rules for falsifying claims with severity (corroborate their denials, p. 81)

Kuru (which means “shaking”) SIST p. 81

Widespread among the Fore people of New Guinea,1960s.

- In around 3-6 months, victims go from unsteady gait to severe tremors, outbursts of laughter to inability to swallow and death.
- Kuru, and (what we now know to be) related diseases, e.g., Mad Cow, Crutzfield Jacobs, scrapie) are “spongiform” diseases: the brains of their victims have small holes giving them a spongy appearance.
- Kuru clusters within families, in particular among women and their children, or elderly parents.
- They began to suspect transmission was through mortuary cannibalism by the maternal kin (this was a main source of meat permitted women, and was also a way of honoring the dead).

- Ending these cannibalistic practice eradicated the disease

No one expected revolutionary implications: that discovering its cause would falsify an established theory that only viruses and bacteria could be infectious (central dogma of biology):

H: All infectious agents have nucleic acid.

Any infectious agent free of nucleic acid would be *anomalous* for *H*–meaning it goes against what *H* claims.

The anomaly: kuru is transmitted by a protein alone, by changing a normal protein shape into an abnormal fold.

Stanley Prusiner called the infectious protein a prion.

A separate step is required to decide when *H*'s anomalies should count as falsifying *H*.

It's necessary to adequately pinpoint blame for anomalies: *Duhemian* problems

The anomalous effect is actually a hypothesis, as Popper admits:

We say that a theory is falsified only if we have accepted basic statements which contradict it…This condition is necessary, but not sufficient; for we have seen that non-reproducible single occurrences are of not sigf]nificance to science…We shall take it as falsified **only if we discover a reproducible effect which refutes the theory. …we only accept the falsification if a low level empirical hypothesis which describes such an effect is proposed and corroborated.** (Popper 1959, 86)
(see also pp 82-3 in SIST)

(~ to Fisher's "we need not an isolated result")

We need a <u>falsifying hypothesis</u> (typically statistical): a hypothesis inferred to falsify some other claim

This is problematic for Popper because Popperians say they never have to justify a claim as true or reliable

He cannot say a method is probative or severe because that would mean it was reliable: “a whiff of induction”, though not inductive enumeration

That means you can’t be a falsificationist and remain a strict deductivist: the falsifying hypothesis involves an evidence transcending (inductive) statistical inference

The discovery of prions led to a “revolution” in molecular biology, Prusiner gets a Nobel prize in 1997.

Two Options Given Hume's Problem: (p. . 45)

- We obtain knowledge non-inductively (a way other than enumerative induction).

- We obtain knowledge by induction and thus by a logically invalid and rationally unjustifiable procedure.

Hume chooses option 2 and concludes that knowledge is just belief based on habit.

Popper wants to take the first option

He claims it's "rational" to accept falsifying hypotheses even though we don't infer them

(p. 51) Hume said our beliefs in regularities are irrational—he’s right if he means we can’t prove them true or probable; however, if belief includes our critical acceptance of scientific theories, a tentative acceptance combined with an eagerness to revise the theory if we design a test it cannot pass, then (Hume was wrong, there’s nothing irrational in accepting such a theory, in preferring it or in relying on for practical purposes on well-tested theories.

Why?

Because no more rational course of action is open to us. p. 51 “Assume we have made it our task to live in this unknown world of ours; …”
nothing "safer" than accepting the "best tested" theory.
(Is this convincing?)

Likewise, you can't really solve Duhem, you accept or "prefer" (as Popper said) the large scale research program or paradigm as a whole.

I only recently came across a remark: (SIST, p. 85)

> [W]e can be reasonably successful in attributing our refutations to definite portions of the theoretical maze. (For we *are* reasonably successful in this–a fact which must remain inexplicable for one who adopts Duhem's and Quine's view on the matter) (1962, p. 243).

That doesn't mean Popper's account solves it.

Falsifying the central dogma of biology (infection requires nucleic acid) involved no series of conjunctions from *H* down to observations, but moving from *the bottom up*, as it were.

- prions are not eradicated with techniques known to kill viruses and bacteria (e.g., UV irraditation, boiling, hospital disinfectants, hydrogen peroxide, and much else).
- If it were a mistake to regard prions as having no nucleic acid, then at least one of these known agents would have eradicated it
- prions are deactivated with substances known to kill proteins.

Popper concluded (36):

1. It is easy to obtain confirmations, if we look for them.
2. **Confirmations should only count if they are the result of risky predictions, if without the theory, we should have expected an event incompatible with the theory.**
3. Every good theory is a prohibition, the more it forbids, the better it is.
4. A theory that is not refutable by any conceivable event is not scientific. Irrefutability is not a virtue.
5. Every genuine test of a theory is an attempt to falsify it or refute it. There are degrees of testability.
6. **Confirming evidence should not count except when it is the result of a genuine test of the theory—it must be able to be presented as a serious but unsuccessful attempt to falsify it (corroborating evidence).**

7. Some testable theories when found false are upheld by their admirers, e.g., by introducing *ad hoc* some auxiliary assumptions or reinterpreting it *ad hoc* so that it escapes refutation. The price paid is to destroy or lower the scientific status of the theory.

But #2, and #6 include a positive side….how can Popper argue for them?
#2 Is Popper’s definition of novelty (prediction)

(Note souvenirs pp 86-7)

With #7, we get to demarcation:

Exhibit (vi) Revisiting Popper's Demarcation

Popper always puts the weight on the theory to be scientific and testable; I think he should put it on the inquiry or test.

- 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 is unscientific.
- Despite being logically falsifiable, theories can be *rendered immune from falsification*.
- Adhering to a falsified theory no matter what is poor science.

- Some areas have so much noise and/or flexibility that they can’t or won’t distinguish warranted from unwarranted explanations of failed predictions.

Demarcation (SIST, p. 89)

A scientific inquiry or test must:

- block inferences that fail the minimal requirement for severity
- 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 ability or inability to pin down the source of failed replications–a familiar occupation these days–speaks to the scientific credentials of an inquiry.

2.4 Novelty and Severity (SIST p. 89)

> According to modern logical empiricist orthodoxy, in deciding whether hypothesis *h* is confirmed by evidence *e*, …we must consider only the statements *h* and *e*, and the logical relations between them. It is quite irrelevant whether *e* was known first and *h* proposed to explain it, or whether *e* resulted from testing predictions drawn from *h*. (Musgrave 1974, p. 2)

Statistician Keynes likewise held that the "…question as to whether a particular hypothesis happens to be propounded before or after examination of [its instances] is quite irrelevant (Keynes, 1921/1952, p. 305).

Logics of confirmation ran into problems because they insisted on purely formal or syntactical criteria of confirmation that, like deductive logic, "should contain no reference to the specific subject matter" (Hempel, 1945, p. 9) in question.

The Popper-Lakatos school attempts to avoid these shortcomings by means of novelty requirements:

> *Novelty Requirement*: for data to warrant a hypothesis *H* requires not just that
> (i) *H* agree with the data, but also (ii) the data should be novel or surprising or the like.

For decades Popperians squabbled over how to define novel predictive success.

There's

- *temporal novelty*–the data were not already available before the hypothesis was erected (Popper, early);

- (2) *theoretical novelty*–the data were not already predicted by an existing hypothesis (Popper, Lakatos),

- (3) *use-novelty*–the data were not used to construct or select the hypothesis.

Temporal novelty is untenable: known data (e.g., the perihelion of Mercury, anomalous for Newton) are often strong evidence for theories (e.g., GTR).

Popper ultimately favored theoretical novelty: *H* passes a severe test with ***x***, when *H* entails ***x***, and ***x*** is theoretically novel–according to a letter he sent me. (We replace "entails" with something like "accords with".)

However, theoretical novelty, prevents passing *H* with severity, so long as there's already a hypothesis that predicts the data or phenomenon ***x*** (they're not clear which). John Worrall (1978a, pp. 330-1)

Why should the first hypothesis that explains ***x*** be better tested?

I take the most promising notion of novelty to be a version of *use-novelty*:

H passes a test with data ***x*** severely, so long as ***x*** was not used to construct *H* (Worrall 1989).

Data can be known, so long as it wasn't used in building *H*, presumably to ensure *H* accords with ***x***.

While the idea is in sync with the error statistical admonishment against "peeking at the data" and finding your hypothesis in the data–it's far too vague as it stands.

Watching this debate unfold in philosophy, I realized none of the notions of novelty were either sufficient or necessary for a good test (Mayo 1991).

There is as much opportunity for bias to arise in interpreting or selectively reporting results, with a known hypothesis, as there is in starting with data and artfully creating a hypothesis.

- Nor is violating use-novelty a matter of the implausibility of *H*.
- On the contrary, popular psychology thrives by seeking to explain results by means of hypotheses expected to meet with approval, at least in a given political tribe.
- Preregistration of the detailed protocol is supposed to cure this. (We come back to this.)

Should use-novelty be *necessary* for a good test?

Violations of use-novelty need not be pejorative.

A trivial example: count all the people in the room and use it to fix the parameter of the number in the room.

- think of confidence intervals: we use the data to form the interval estimate; the same data warrant the hypothesis constructed!
- Likewise using the same data to arrive at and test assumptions of statistical models can be entirely reliable.
- What matters is not novelty, in any of the senses, but severity in the error statistical sense.

- We should look at particular ways it can fail:
  Biasing selection effects: when data or hypotheses are selected or generated (or a test criterion is specified), in such a way that the minimal severity requirement is violated, seriously altered, or incapable of being assessed.