

Discriminating Experiments: Severe Tests and Operational Falsification

Juan Segurola

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Segurola. All Rights Reserved.

Correspondence: info@3Dresyns.com · www.3Dresyns.com

INF-007 · Scientific Inference and Experimental Reasoning · Version v0.2

Abstract

An experiment is most informative when plausible explanations make different predictions under the conditions actually tested. Merely repeating a measurement, increasing sample size or perturbing a system does not guarantee discrimination if all live hypotheses predict the same outcome. This article develops the general logic of discriminating experiments. It begins from the rival set established in INF-004, identifies predictions that diverge among alternatives, and asks whether the proposed measurement or intervention has sufficient capability to reveal those divergences. The article uses severe-testing and strong-inference ideas as design principles while rejecting a simplistic view in which one failed prediction mechanically falsifies a theory. Auxiliary hypotheses, measurement limits, boundary conditions and state dependence must be made explicit. The output of a discriminating experiment can be elimination of one rival, retention of several, or insufficient information. VPP-specific diagnostic experiment design remains with ER-184.

Keywords: discriminating experiment; severe test; falsification; rival hypotheses; experimental design; test capability; intervention

1. Experiments should separate alternatives, not merely perturb systems

Platt's strong-inference programme emphasised multiple hypotheses and experiments capable of excluding alternatives [1]. The core insight is still useful: an experiment adds little inferential value if every live explanation predicts the same result. A discriminating experiment instead chooses a perturbation or measurement for which plausible rivals diverge.

2. Start from the rival set

INF-004 owns the decision about which explanations remain live. INF-007 takes that set as input. Each rival should be linked to a prediction conditional on stated auxiliary assumptions. The design target is not "collect more data" in the abstract but obtain observations whose distribution or qualitative pattern differs enough among rivals to change the inferential state.

3. Severity and capability

Mayo's severe-testing framework emphasises that a result is probative against an error or discrepancy only if the procedure had a high capability to reveal that discrepancy were it present [2]. INF uses this as an operational design principle, not as a universal statistical doctrine. A negative result from an insensitive assay is weak discrimination; a negative result from a validated high-capability test can be strongly informative about hypotheses that required a detectable effect.

4. Intervention can create contrast

Observation alone may leave rivals observationally equivalent. Intervention can change the causal structure or system state in a way that generates different predictions. Modern causal inference formalises this distinction between observing a variable and intervening on it [3,4]. Yet intervention is not automatically decisive: non-compliance, compensatory pathways, off-target effects and measurement limitations can preserve ambiguity.

5. Auxiliary assumptions must be exposed

A predicted difference depends on more than the focal hypothesis. It can require correct calibration, adequate exposure range, stable specimen state, correct boundary conditions and valid measurement timing. A failed prediction therefore challenges the conjunction of focal and auxiliary assumptions. The design should identify which auxiliaries are being relied on and, where possible, include controls that separate focal failure from measurement or implementation failure.

6. A discriminating-experiment matrix

Rival	Predicted response	Required capability	Interpretation if response absent
A	Large directional change under perturbation P	Resolve change above method uncertainty	A weakened only if state and method capability are verified
B	No change under P but response under Q	Q must isolate the mechanism unique to B	Absence under both P and Q may leave implementation ambiguity
C	Response tracks nuisance variable N	Measure N independently and perturb it if possible	C weakened if response persists when N is controlled

7. Falsification is operational and bounded

INF does not use “falsified” to mean that one discrepant point permanently destroys a theory. Operational falsification means that, under specified assumptions and adequate test capability, an observed result is incompatible with a rival prediction to a degree sufficient to remove or materially weaken that rival for the present inferential question. Broader theory revision may require additional evidence.

8. Design efficiency versus discrimination

A statistically efficient design for estimating one parameter is not necessarily the best design for distinguishing model structures. Discrimination may require sampling at conditions where rival predictions diverge most strongly rather than where the preferred model is easiest to estimate. This difference is central when structural uncertainty dominates parameter uncertainty.

9. Minimum reporting discipline

- State the live rival explanations before the experiment.
- For each rival, state the predicted observation under the proposed intervention or measurement.
- Quantify or otherwise justify test capability for the predicted difference.
- List auxiliary assumptions required for the prediction.
- Include controls that diagnose implementation or measurement failure where feasible.
- Allow “insufficient discrimination” as an acceptable outcome.
- Do not generalise a successful discriminating test beyond the state and domain actually exercised.

References

- [1] Platt, J.R. Strong Inference. *Science* 146(3642), 347-353 (1964). DOI: 10.1126/science.146.3642.347.
- [2] Mayo, D.G. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press, 2018. DOI: 10.1017/9781107286184.
- [3] Pearl, J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511803161.
- [4] Hernán, M.A.; Robins, J.M. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020. DOI: 10.1201/9781315374932.

Data and materials statement

No new experimental dataset was generated for this article. The article is a methodological analysis based on the cited literature and the frozen ownership architecture of the INF series.