FN-Wissenschaft-Bilanx-EN-v0.1-2026-05-17

Research Note — Measuring, Interpreting, Applying

Author: Thomas Schüller · forschung@tschueller.com · ORCID 0009-0003-9799-6747 **Affiliation:** privately funded, no institutional affiliation **Licence:** CC BY 4.0 · **Version:** 0.1 · **Date:** 2026-05-17 **Strand:** FORSCHUNG / Philosophy of Science / W1 **Tool:** BILANX · **Document type:** Research note (Template v1.0) **Status:** completed test with documented assessment **Special feature: balance sheet of the workshop** — the corpus is the data set

---

1. Thesis

Thirty-one modules, six strands, one procedure: each module sets up a narrow model, computes it through and documents the result — even when it is inconvenient. This module draws the balance.

It has a peculiarity. The workshop line demands: **no claim without calculation.** A philosophy-of-science module could evade this rule by talking about science instead of computing. It does not: **the corpus itself is the data set.** All figures come from the modules' result files, not from memory.

The tested baseline model M1 reads: *"A model is either true or false. A refuted model is worthless. Measuring delivers facts; interpreting comes afterwards and is arbitrary."*

This module's claim: all three clauses are false — and the corpus supplies the evidence.

2. Assumptions

1. The modules' result files are complete and machine-readable.
2. The label "M1 refuted" was assigned in each module by the same criterion.
3. Validity limits computed in modules are comparably compilable.
4. Underdetermination can be described as a structure even where the fields differ.
5. Assessing a procedure by that same procedure is possible but not neutral (see section 8).

3. Mathematical model

M1 — "true or false": For every model M, $M \in \{\text{true}, \text{false}\}$, and from $M = \text{false}$ it follows that M is worthless.

M2 — balance. For N tested modules it is counted in how many the narrow model fell in **all** scenarios (n_{full}), in how many it partly held (n_{part}), and how many had a mapping rather than refuting purpose (n_{map}). M1 is refuted as soon as $n_{\text{part}} > 0$ — for then the same model holds in one scenario and fails in the next.

M3 — underdetermination. A map $f: X \to Y$ from causes to data is underdetermined if $\dim(\ker f) > 0$, i.e. several causes produce the same data. For metamerism this is exact: $\dim(\ker f) = 61 - 3 = 58$.

M4 — domain of validity. Instead of $M \in \{\text{true}, \text{false}\}$: to M belongs a domain B and an error measure $\varepsilon(x)$ such that $$x \in B \iff \varepsilon(x) < \varepsilon_{\text{tol}}.$$ For the pendulum: $B = \{\theta_0 < 23° \}$ at $\varepsilon_{\text{tol}} = 1\,\%$.

Symbol	Meaning	Unit
n_{full}	modules where M1 fell entirely	—
$\dim(\ker f)$	dimension of the null space	—
B	domain of validity	model-dependent
ε_{tol}	error tolerance	—

4. Parameters

Scenario	Setup
A balance	all 27 modules with evaluable data set, as of 2026-05-17
B underdetermination	three cases from separate fields: colorimetry, perception, quantum physics
C domains of validity	six models with computed limits from various modules

Tolerance: $\varepsilon = 0$ — one counterexample suffices.

5. Prediction

Scenario	Expectation
A	if the procedure was open-ended, M1 held in some cases
B	the same structure in different fields, but different kinds of underdetermination
C	refuted models with a statable, useful domain

6. Falsification criterion

M1 is refuted as soon as (a) a model holds in one scenario and fails in another, (b) measuring does not determine the cause uniquely, or (c) a refuted model demonstrably continues to carry.

7. Competing models

- **K1 — true or false (M1):** tested baseline; the naive picture of science.
- **K2 — naive falsificationism:** one counterexample finishes a theory. Too coarse, as scenario C shows.
- **K3 — Duhem-Quine:** theories are never tested singly but in bundles. Not treated here, but related.
- **K4 — Kuhn:** paradigm shifts rather than stepwise refutation. The workshop operates on a smaller scale.
- **K5 — instrumentalism:** models are tools, not descriptions of the world. Explains scenario C well but does not do justice to scenario B.
- **K6 — domain of validity with error measure:** the position taken here — models have domains and statable costs.

8. Domain of failure

- **The balance is not neutral.** This module assesses a procedure with that same procedure. That is no circle in the logical sense, but it is also no independent check. An external assessment could weight things differently.
- **The case selection in scenario C is illustrative, not random.** It shows what is to be shown. For a balance sheet this is unavoidable, but it remains a selection. **The count in scenario A, by contrast, is complete.**
- **The label "M1 refuted" is a stipulation.** It was assigned in each module by that module's own criterion. In two modules (spectrum, string) it served a mapping purpose, which explains the count but also shows that the flag does not mean the same everywhere.
- **Only modules with a machine-readable data set are covered** — 27 of 31. The four early modules without structured result files are missing.
- **Not treated:** the Duhem-Quine thesis, theory-ladenness of observation, incommensurability, sociology of science, the replication crisis.
- **The interpretation question remains open.** What is the case when several interpretations make the same predictions belongs to the planned module W2.

9. Simulation

Numerical evaluation in `bilanx_simulation.py`:

1. Scenario A: count of all modules with data sets by category; compilation of cases where M1 held; count of documented self-corrections.
2. Scenario B: three cases of underdetermination with structure, dimension and kind of resolvability.
3. Scenario C: six models with status, validity limit, criterion and continuing use.

Results in `bilanx_results.json`, plot as `bilanx_szenen.svg`.

10. Result

10.1 A — the balance

Category	Count	Share
fully refuted		

	22	**82 %**
partly upheld	3	11 %
mapping rather than refuting	2	7 %
total	**27**	

Where M1 held:

Module	The case
09 colour mixing	additive RGB mixing
10 mirrors	plane mirror — perfect imaging
12 instruments	two-stage microscope
Audio 01 spectrum	FFT bin-centred
Audio 02 string	nylon string harmonic

Documented self-corrections: 9 (Q1: 4, L4: 2, M2: 2, N1: 1) **Papers with an explicit domain of failure:** 52

Assessment M1: refuted.

10.2 B — underdetermination

Field	Structure	Dimension	Kind
colorimetry (M1)	61 → 3	**58**	exactly computable
perception (L5/L6)	1 stimulus → 2 interpretations	2	resolvable
quantum physics (Q1/Q2)	1 formalism → 5 interpretations	5	**in principle**

Assessment M1: refuted.

10.3 C — domains of validity

Model	Status	Valid up to
harmonic pendulum	strictly false	**23°** (< 1 % error)
geometrical optics	strictly false	$\lambda \ll$ structure size
classical mechanics	strictly false	$\lambda/L < 10^{-3}$
Newton’s corpuscles	refuted 1850	path shape correct
local realism	excluded	0°, 90°, 180° exactly
plane mirror	**holds**	always, in the plane case

5 of 6 are strictly false — and continue to carry.

Assessment M1: refuted.

11. Assessment

M1 is refuted in all three clauses — and the evidence comes from the work itself.

Scenario A: the workshop did not always turn out right. This is the most important single finding of the balance. In 22 of 27 modules the narrow model fell entirely, in five it did not.

A procedure that always yields the same result tests nothing. Had all 27 modules refuted their M1, that would have been grounds for suspicion — the selection of models would presumably have anticipated the result. **That five models held wholly or partly is the evidence for open-endedness.**

And these five share a property that became visible in the counting: **they are the ideal limiting cases.** The plane mirror does image perfectly — deviations begin with curvature. Additive RGB mixing does follow the simple description — metamerism only shows where it ends. The nylon string is harmonic — the piano bass string is not.

That is already the answer to M1: **the same model holds in one scenario and fails in the next.** The mirror case makes it clearest — "perfect imaging" is true in the plane case and false in the curved one. One truth value does not suffice to describe that.

Scenario B: a structure that imposed itself. Three fields with nothing to do with one another — colorimetry, perceptual psychology, quantum physics — show the same basic figure: **different causes produce identical data.**

The workshop did not look for this. It first appeared in **M1 (CATENA)**, when metamerism turned out to be a 58-dimensional null space. Then in the **perception strand**, where under balanced competition two interpretations fit equally well. Then in **Q1 and Q2**, where five interpretations of quantum mechanics make identical predictions. Each time independently, each time in a different module.

The difference between the cases matters as much as the similarity. For metamerism the underdetermination is exactly computable — 58 dimensions, and one can state which spectra are indistinguishable. For perception it is **resolvable**: motion, context or a second eye settle the matter. For the quantum interpretations it is **in principle** — no better measurement helps, because all interpretations predict the same.

The merge rule applies to this finding too: that three fields show the same **form** does not mean they show the same **phenomenon**. Without that distinction an observation would turn into a worldview.

And with that the second clause of M1 is settled: **measuring does not simply deliver facts.** It delivers data compatible with several states of the world. Interpreting is not what comes afterwards — it is part of turning data into claims.

Scenario C: the wrong question and the right one. Five of the six models listed are strictly false. All five continue to be used, and rightly so.

The harmonic pendulum is false — the restoring force goes as \(\sin\theta\), not \(\theta\). It carries to 23° with under 1 % error, and pendulum clocks work. Geometrical optics is false — light is not rays. All lens design computes with it. Classical mechanics is false — it carries from bridge statics to spaceflight.

"True or false" has a verdict as its answer. "How far does it carry and what does the deviation cost" has a number. 23 degrees. 10^{-3}. 39.5 degrees. The second question is the more usable one, and it is the only one that can be computed.

Newton's corpuscular theory is the most instructive entry in the whole table. It was false — Foucault measured it in 1850. But it was not worthless, it was **indispensable**: because it made a precise,

opposite prediction (light faster in the denser medium instead of slower, factor 1.78), the question became decidable at all. A vaguely correct theory would not have achieved that.

A refuted model can achieve more than a fuzzy correct one. With that the third clause of M1 is settled too.

What the workshop learned as a procedure. The template requires a **domain of failure** in section 8 — filled in 52 papers. This is no gesture of modesty. After this balance it is clear why: **that is where the information sits.** A model without a stated limit is not more modest but poorer — one does not know when to trust it.

The same holds for the nine documented self-corrections. Four of them in Q1, where an unsuitable measure, a faulty significance treatment and a measurement of the wrong quantity came to light one after another. Two in Q2, where a wrong angle choice brought the result to zero. They are reported rather than silently fixed, because a procedure that hides its errors is not testable.

And what remains open. This balance concerns the procedure, not the subject matter. The question of what is the case when several interpretations make the same predictions — whether one of them is true and we cannot know it, or whether the question is wrongly posed — remains unanswered. It belongs to the planned module **W2**.

The workshop thus ends not with a result but with a sharpened question. That befits its procedure: **not the answer is the valuable part, but the statement of where it ceases to hold.**

12. Next test

Module W2 (The Interpretations) as the actual conclusion: what follows from the existence of empirically equivalent interpretations? There Copenhagen, many-worlds, Bohm, GRW and QBism would be compared not as opinions but by their structural costs.

Possible extension v0.2: the Duhem-Quine thesis; theory-ladenness of observation (the perception strand supplies material for this); reproducibility as a workshop question; an external assessment of the corpus by other criteria.

Appendix A — Cross-references

- Practice explanation: `PRAXIS-Bilanx-EN.md` · Companion paper: `BEGLEITPAPIER-Bilanx-EN.md`
- Tool: `BILANX.html`
- Code: `bilanx_simulation.py` · Plot: `bilanx_szenen.svg` · Data: `bilanx_results.json`
- **Data source:** all module result files in the master bundle `WERKSTATT-GESAMT_2026-05-17.zip`
- **Particularly used:** M1 signal chain (metamerism), M2 optical mechanics (Newton), M3 oscillation (pendulum), L5/L6 (bistability), Q1/Q2 (interpretations), 10 mirrors (ideal limiting case)

Appendix B — External stocktaking

- **Popper (1934/59)** — falsification as a demarcation criterion.
- **Duhem (1906), Quine (1951)** — theories are tested in bundles.
- **Kuhn (1962)** — paradigms and scientific revolutions.

- **Lakatos (1970)** — research programmes with a hard core and protective belt.
- **Cartwright (1983), *How the Laws of Physics Lie*** — models as tools with domains of application.

What distinguishes this note: it does not argue about science in general but assesses a concrete, fully documented corpus — with figures drawn from the data sets and reproducible. The cases where the procedure did **not** yield the expected result are counted in.

Appendix C — Sources

- Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchinson (German 1934, *Logik der Forschung*).
- Duhem, P. (1906). *La théorie physique: son objet et sa structure*. Chevalier & Rivière.
- Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review* 60:20–43.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press.

Word count: ca. 2180 · Pages: 10 · Version 0.1 · 2026-05-17