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

Forschungsnotiz — Messen, Deuten, Anwenden

Autor: Thomas Schüller · forschung@tschueller.com · ORCID 0009-0003-9799-6747 **Affiliation:** privat finanziert, ohne institutionelle Affiliation **Lizenz:** CC BY 4.0 · **Version:** 0.1 · **Datum:** 2026-05-17 **Strang:** FORSCHUNG / Wissenschaftsphilosophie / W1 **Werkzeug:** BILANX · **Dokumenttyp:** Forschungsnotiz (Template v1.0) **Status:** abgeschlossener Test mit dokumentierter Bewertung **Besonderheit:** **Bilanz der Werkstatt** — der Korpus ist der Datensatz

---

1. These

Einunddreißig Module, sechs Stränge, ein Verfahren: Jedes Modul stellt ein enges Modell auf, rechnet es durch und dokumentiert das Ergebnis — auch dann, wenn es unbequem ist. Dieses Modul zieht die Bilanz.

Es hat eine Besonderheit. Die Werkstattlinie verlangt: **keine Aussage ohne Rechnung.** Ein wissenschaftsphilosophisches Modul könnte dieser Regel ausweichen, indem es über Wissenschaft redet statt zu rechnen. Es tut es nicht: **Der eigene Korpus ist der Datensatz.** Alle Zahlen stammen aus den Ergebnisdateien der Module, nicht aus der Erinnerung.

Das geprüfte engere Modell M1 lautet: *„Ein Modell ist entweder wahr oder falsch. Ein widerlegtes Modell ist wertlos. Messen liefert Tatsachen; Deuten kommt danach und ist beliebig."*

Die Behauptung dieses Moduls: Alle drei Teilsätze sind falsch — und der eigene Korpus liefert die Belege.

2. Annahmen

1. Die Ergebnisdateien der Module sind vollständig und maschinell auswertbar.
2. Die Kennzeichnung „M1 widerlegt" wurde in jedem Modul nach demselben Kriterium gesetzt.
3. Gültigkeitsgrenzen, die in Modulen gerechnet wurden, sind vergleichbar zusammenstellbar.
4. Unterbestimmtheit lässt sich als Struktur beschreiben, auch wenn die Gebiete verschieden sind.
5. Die Bilanz eines Verfahrens durch dieses Verfahren selbst ist möglich, aber nicht neutral (siehe Abschnitt 8).

3. Mathematisches Modell

M1 — „wahr oder falsch”: Für jedes Modell M gilt $M \in \{\text{wahr}, \text{falsch}\}$, und aus $M = \text{falsch}$ folgt: M ist wertlos.

M2 — Bilanz. Für N geprüfte Module wird ausgezählt, in wie vielen das enge Modell in **allen** Szenen fiel (n_{voll}), in wie vielen es teilweise stand hielt (n_{teil}), und wie viele kartierenden statt widerlegenden Zweck hatten (n_{kart}). M1 ist widerlegt, sobald $n_{\text{teil}} > 0$ — denn dann gilt dasselbe Modell in einer Szene und fällt in der nächsten.

M3 — Unterbestimmtheit. Eine Abbildung $f: X \to Y$ von Ursachen auf Daten heißt unterbestimmt, wenn $\dim(\ker f) > 0$, also mehrere Ursachen dieselben Daten erzeugen. Bei der Metamerie ist dies exakt: $\dim(\ker f) = 61 - 3 = 58$.

M4 — Gültigkeitsbereich. Statt $M \in \{\text{wahr}, \text{falsch}\}$ gilt: Zu M gehört ein Bereich B und ein Fehlermaß $\varepsilon(x)$, sodass $$x \in B \iff \varepsilon(x) < \varepsilon_{\text{tol}}.$$ Beim Pendel: $B = \{\theta_0 < 23°\}$ bei $\varepsilon_{\text{tol}} = 1\,\%$.

Symbol	Bedeutung	Einheit
n_{voll}	Module, in denen M1 ganz fiel	—
$\dim(\ker f)$	Dimension des Nullraums	—
B	Gültigkeitsbereich	je nach Modell
ε_{tol}	Fehlertoleranz	—

4. Parameter

Szenario	Setup
A Bilanz	alle 27 Module mit auswertbarem Datensatz, Stand 2026-05-17
B Unterbestimmtheit	drei Fälle aus getrennten Gebieten: Farbmetrik, Wahrnehmung, Quantenphysik
C Gültigkeitsbereiche	sechs Modelle mit gerechneter Grenze aus verschiedenen Modulen

Toleranz: $\varepsilon = 0$ — ein Gegenbeispiel genügt.

5. Vorhersage

Szenario	Erwartung
A	wenn das Verfahren ergebnisoffen war, hielt M1 in einigen Fällen stand
B	dieselbe Struktur in verschiedenen Gebieten, aber verschiedene Art der Unterbestimmtheit
C	widerlegte Modelle mit angebbarem, nützlichem Bereich

6. Falsifikationskriterium

M1 ist widerlegt, sobald (a) ein Modell in einer Szene gilt und in einer anderen fällt, (b) Messen die Ursache nicht eindeutig bestimmt, oder (c) ein widerlegtes Modell nachweislich weiter trägt.

7. Konkurrenzmodelle

- **K1 — wahr oder falsch (M1):** geprüftes engeres Modell; die naive Vorstellung von Wissenschaft.
- **K2 — naiver Falsifikationismus:** ein Gegenbeispiel erledigt eine Theorie. Zu grob, wie Szene C zeigt.
- **K3 — Duhem-Quine:** Theorien werden nie einzeln geprüft, sondern im Verbund. Wird hier nicht behandelt, ist aber verwandt.
- **K4 — Kuhn:** Paradigmenwechsel statt schrittweiser Widerlegung. Die Werkstatt arbeitet auf kleinerer Skala.
- **K5 — Instrumentalismus:** Modelle sind Werkzeuge, nicht Weltbeschreibungen. Erklärt Szene C gut, wird aber Szene B nicht gerecht.
- **K6 — Gültigkeitsbereich mit Fehlermaß:** die hier vertretene Position — Modelle haben Bereiche und angebbare Kosten.

8. Versagensbereich

- **Die Bilanz ist nicht neutral.** Dieses Modul wertet ein Verfahren mit demselben Verfahren aus. Das ist kein Zirkel im logischen Sinn, aber es ist auch keine unabhängige Prüfung. Eine externe Auswertung könnte zu anderen Gewichtungen kommen.
- **Die Fallauswahl in Szene C ist illustrativ, nicht zufällig.** Sie zeigt, was gezeigt werden soll. Bei einer Bilanz ist das unvermeidlich, aber es bleibt eine Auswahl. **Die Zählung in Szene A ist dagegen vollständig.**
- **Die Kennzeichnung „M1 widerlegt" ist eine Setzung.** Sie wurde in jedem Modul nach dessen eigenem Kriterium vergeben. Bei zwei Modulen (Spektrum, Saite) diente sie einem kartierenden Zweck, was die Zählung erklärt, aber auch zeigt, dass das Flag nicht überall dasselbe bedeutet.
- **Nur Module mit maschinell auswertbarem Datensatz sind erfasst** — 27 von 31. Die vier frühen Module ohne strukturierte Ergebnisdatei fehlen.
- **Nicht behandelt:** Duhem-Quine-These, Theoriebeladenheit der Beobachtung, Inkommensurabilität, Wissenschaftssoziologie, Reproduzierbarkeitskrise.
- **Die Deutungsfrage bleibt offen.** Was der Fall ist, wenn mehrere Deutungen dieselben Vorhersagen machen, gehört ins geplante Modul W2.

9. Simulation

Numerische Auswertung in `bilanx_simulation.py`:

1. Szene A: Auszählung aller Module mit Datensatz nach Kategorie; Zusammenstellung der Fälle, in denen M1 stand hielt; Zählung der dokumentierten Eigenkorrekturen.
2. Szene B: drei Fälle von Unterbestimmtheit mit Struktur, Dimension und Art der Auflösbarkeit.
3. Szene C: sechs Modelle mit Status, Gültigkeitsgrenze, Kriterium und weiterbestehendem Nutzen.

Ergebnisse in `bilanx_results.json`, Plot als `bilanx_szenen.svg`.

10. Ergebnis

10.1 A — Die Bilanz

Kategorie	Anzahl	Anteil
vollständig widerlegt	**22**	**82 %**
teilweise gestützt	3	11 %
kartierend statt widerlegend	2	7 %
gesamt	**27**	

Wo M1 stand hielt:

Modul	Der Fall
09 Farbmischung	additive RGB-Mischung
10 Spiegel	ebener Spiegel — perfekte Abbildung
12 Instrumente	zweistufiges Mikroskop
Audio 01 Spektrum	FFT bin-zentriert
Audio 02 Saite	Nylon-Saite harmonisch

Dokumentierte Eigenkorrekturen: 9 (Q1: 4, L4: 2, M2: 2, N1: 1)
Papers mit explizitem Versagensbereich: 52

Bewertung M1: widerlegt.

10.2 B — Unterbestimmtheit

Gebiet	Struktur	Dimension	Art
Farbmetrik (M1)	61 → 3	**58**	exakt berechenbar
Wahrnehmung (L5/L6)	1 Reiz → 2 Deutungen	2	auflösbar
Quantenphysik (Q1/Q2)	1 Formalismus → 5 Deutungen	5	**prinzipiell**

Bewertung M1: widerlegt.

10.3 C — Gültigkeitsbereiche

Modell	Status	Gültig bis
Harmonisches Pendel	streng falsch	**23°** (< 1 % Fehler)
Geometrische Optik	streng falsch	$\lambda \ll$ Strukturgröße
Klassische Mechanik	streng falsch	$\lambda/L < 10^{-3}$
Newtons Korpuskeln	widerlegt 1850	Bahnform korrekt
Lokaler Realismus	ausgeschlossen	0°, 90°, 180° exakt
Ebener Spiegel	**gilt**	immer, im ebenen Fall

5 von 6 sind streng falsch — und tragen weiter.

Bewertung M1: widerlegt.

11. Bewertung

M1 ist in allen drei Teilsätzen widerlegt — und die Belege kommen aus der eigenen Arbeit.

Szene A: Die Werkstatt hat nicht immer recht bekommen. Das ist der wichtigste Einzelbefund dieser Bilanz. In 22 von 27 Modulen fiel das enge Modell vollständig, in fünf nicht.

Ein Verfahren, das immer dasselbe Ergebnis liefert, prüft nichts. Hätten alle 27 Module ihr M1 widerlegt, wäre das ein Grund zum Misstrauen gewesen — dann hätte die Auswahl der Modelle vermutlich schon das Ergebnis vorweggenommen. **Dass fünf Modelle ganz oder teilweise stand hielten, ist der Beleg für Ergebnisoffenheit.**

Und diese fünf haben eine gemeinsame Eigenschaft, die beim Auszählen sichtbar wurde: **Es sind die idealen Grenzfälle.** Der ebene Spiegel bildet tatsächlich perfekt ab — die Abweichungen beginnen bei Krümmung. Die additive RGB-Mischung folgt tatsächlich der einfachen Beschreibung — die Metamerie zeigt erst, wo es endet. Die Nylon-Saite ist harmonisch — die Klavier-Bass-Saite nicht.

Das ist bereits die Antwort auf M1: **Dasselbe Modell gilt in einer Szene und fällt in der nächsten.** Der Spiegel-Fall macht es am deutlichsten — „perfekte Abbildung" ist im ebenen Fall wahr und im gekrümmten falsch. Ein Wahrheitswert reicht nicht aus, um das zu beschreiben.

Szene B: eine Struktur, die sich aufgedrängt hat. Drei Gebiete, die nichts miteinander zu tun haben — Farbmetrik, Wahrnehmungspsychologie, Quantenphysik —, zeigen dieselbe Grundfigur: **Verschiedene Ursachen erzeugen identische Daten.**

Die Werkstatt hat das nicht gesucht. Es tauchte zuerst in **M1 (CATENA)** auf, als die Metamerie sich als 58-dimensionaler Nullraum herausstellte. Dann im **Wahrnehmungs-Strang**, wo bei ausgeglichener Konkurrenz zwei Deutungen gleich gut passen. Dann in **Q1 und Q2**, wo fünf Deutungen der Quantenmechanik identische Vorhersagen machen. Jedes Mal unabhängig, jedes Mal in einem anderen Modul.

Der Unterschied zwischen den Fällen ist dabei ebenso wichtig wie die Ähnlichkeit. Bei der Metamerie ist die Unterbestimmtheit exakt berechenbar — 58 Dimensionen, und man kann angeben, welche Spektren sich nicht unterscheiden lassen. Bei der Wahrnehmung ist sie **auflösbar**: Bewegung, Kontext oder ein zweites Auge entscheiden die Sache. Bei den Quantendeutungen ist sie **prinzipiell** — keine bessere Messung hilft, weil alle Deutungen dasselbe vorhersagen.

Die Merge-Regel gilt auch für diesen Befund: Dass drei Gebiete dieselbe **Form** zeigen, heißt nicht, dass sie dasselbe **Phänomen** zeigen. Ohne diese Unterscheidung würde aus einer Beobachtung eine Weltanschauung.

Und damit ist der zweite Teilsatz von M1 erledigt: **Messen liefert nicht einfach Tatsachen.** Es liefert Daten, die mit mehreren Zuständen der Welt verträglich sind. Deuten ist nicht das, was danach kommt — es ist Teil davon, aus Daten Aussagen zu machen.

Szene C: die falsche Frage und die richtige. Fünf von sechs aufgeführten Modellen sind streng genommen falsch. Alle fünf werden weiter verwendet, und zwar zu Recht.

Das harmonische Pendel ist falsch — die Rückstellkraft geht mit \(\sin\theta\), nicht mit \(\theta\). Es trägt bis 23° mit unter 1 % Fehler, und Pendeluhren funktionieren. Die geometrische Optik ist falsch — Licht sind keine Strahlen. Der gesamte Linsenbau rechnet damit. Die klassische Mechanik ist falsch — sie trägt von der Brückenstatik bis zur Raumfahrt.

„Wahr oder falsch" hat ein Urteil als Antwort. „Bis wohin trägt es und was kostet die Abweichung" hat eine Zahl. 23 Grad. 10^{-3}. 39,5 Grad. Die zweite Frage ist die brauchbarere, und sie ist die einzige, die sich rechnen lässt.

Newtons Korpuskeltheorie ist der lehrreichste Eintrag der ganzen Tabelle. Sie war falsch — Foucault maß 1850 nach. Aber sie war nicht wertlos, sondern **unverzichtbar**: Weil sie eine präzise, entgegengesetzte Vorhersage machte (Licht schneller im dichteren Medium statt langsamer, Faktor 1,78), wurde die Frage überhaupt entscheidbar. Eine vage richtige Theorie hätte das nicht geleistet.

Ein widerlegtes Modell kann mehr leisten als ein unscharfes richtiges. Damit ist auch der dritte Teilsatz von M1 erledigt.

Was die Werkstatt als Verfahren gelernt hat. Das Template verlangt in Abschnitt 8 einen **Versagensbereich** — in 52 Papers ausgefüllt. Das ist keine Bescheidenheitsgeste. Nach dieser Bilanz ist klar, warum: **Dort sitzt die Information.** Ein Modell ohne angegebene Grenze ist nicht bescheidener, sondern ärmer — man weiß nicht, wann man ihm trauen darf.

Dasselbe gilt für die neun dokumentierten Eigenkorrekturen. Vier davon in Q1, wo ein untaugliches Maß, eine falsche Signifikanzbehandlung und eine Messung der falschen Größe nacheinander auffielen. Zwei in Q2, wo eine falsche Winkelwahl das Ergebnis auf null brachte. Sie sind ausgewiesen statt stillschweigend behoben, weil ein Verfahren, das seine Fehler verbirgt, nicht prüfbar ist.

Und was offen bleibt. Diese Bilanz betrifft das Verfahren, nicht die Sache. Die Frage, was der Fall ist, wenn mehrere Deutungen dieselben Vorhersagen machen — ob eine davon zutrifft und wir es nicht wissen können, oder ob die Frage falsch gestellt ist —, bleibt unbeantwortet. Sie gehört ins geplante Modul **W2**.

Die Werkstatt endet damit nicht mit einem Ergebnis, sondern mit einer geschärften Frage. Das entspricht ihrem Verfahren: **Nicht die Antwort ist der wertvolle Teil, sondern die Angabe, wo sie aufhört zu gelten.**

12. Nächster Test

Modul W2 (Die Deutungen) als eigentlicher Abschluss: Was folgt daraus, dass empirisch äquivalente Deutungen existieren? Dort wären Kopenhagen, Viele-Welten, Bohm, GRW und QBism nicht als Meinungen, sondern nach ihren strukturellen Kosten zu vergleichen.

Mögliche Erweiterung v0.2: Duhem-Quine-These; Theoriebeladenheit der Beobachtung (der Wahrnehmungs-Strang liefert dafür Material); Reproduzierbarkeit als Werkstattfrage; eine externe Auswertung des Korpus durch andere Kriterien.

Anhang A — Querverweise

- Praxis-Erklärung: `PRAXIS-Bilanx.md` · Begleitpapier: `BEGLEITPAPIER-Bilanx.md`
- Werkzeug: `BILANX.html`
- Code: `bilanx_simulation.py` · Plot: `bilanx_szenen.svg` · Daten: `bilanx_results.json`
- **Datenquelle:** alle Modul-Ergebnisdateien im Master-Bundle `WERKSTATT-GESAMT_2026-05-17.zip`
- **Besonders verwendet:** M1 Signalkette (Metamerie), M2 Optische

Mechanik (Newton), M3 Schwingung (Pendel), L5/L6 (Bistabilität), Q1/Q2 (Deutungen), 10 Spiegel (idealer Grenzfall)

Anhang B — Externer Bestandscheck

- **Popper (1934/59)** — Falsifikation als Abgrenzungskriterium.
- **Duhem (1906), Quine (1951)** — Theorien werden im Verbund geprüft.
- **Kuhn (1962)** — Paradigmen und wissenschaftliche Revolutionen.
- **Lakatos (1970)** — Forschungsprogramme mit hartem Kern und Schutzgürtel.
- **Cartwright (1983), *How the Laws of Physics Lie*** — Modelle als Werkzeuge mit Anwendungsbereich.

Was diese Notiz davon unterscheidet: Sie argumentiert nicht über Wissenschaft im Allgemeinen, sondern wertet einen konkreten, vollständig dokumentierten Korpus aus — mit Zahlen, die aus den Datensätzen stammen und reproduzierbar sind. Die Fälle, in denen das Verfahren **nicht** das erwartete Ergebnis lieferte, sind mitgezählt.

Anhang C — Quellen

- Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchinson (dt. 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.

Wortzahl: ca. 2200 · Seiten: 10 · Version 0.1 · 2026-05-17