Measurement Blind Spots

300 hypotheses, 30 refined, 3 investigated

1 November 2026

Measurement Blind Spots

300 hypotheses about what experimental science fails to see, 30 developed in detail, and three investigated with working mathematics.

Compiled 1 November 2026. Every entry is a proposition, not a report of a measurement. No experiment described in this collection has been performed.

What this is

Science advances by measuring. It also accumulates blind spots: quantities an instrument cannot observe, parameters that are fitted rather than measured, approximations that are exact in the regime where they were tested and wrong in the regime where they are used, and disagreements between two accepted methods that are resolved by convention rather than by evidence.

This collection is a systematic hunt for those blind spots. It contains:

- **300 hypotheses**, each stating the established baseline, the proposed mechanism, the observation that would discriminate, an available data source or experiment, the prior art, and an explicit refutation condition. They span physics, astronomy, chemistry, electrochemistry, materials, biology, medicine, engineering, Earth science, and the shared statistical machinery underneath all of them.
- **30 refined hypotheses** developed further, each separating what is a mathematical identity, what is an artefact of the analysis, and what is a genuine falsifiable prediction.
- **3 full investigations** with derivations, validated numerical code, and results: a finite-window identifiability problem in decay fitting, a hidden boundary conductance in thermal metrology, and a proof that Kramers–Kronig consistency does not imply linearity in impedance spectroscopy.

The seven blind-spot classes

Every hypothesis is tagged with one:

Class	Count	What it means
`boundary-condition`	64	The measurement is dominated by an interface, edge, or finite-size term that the model omits
`hidden-parameter`	54	A quantity is fitted, assumed, or inherited from a protocol rather than measured
`time-dependent`	37	Drift, aging, relaxation, or protocol timing sets the result
`model-disagreement`	27	Two accepted methods give different answers and both are internally consistent
`model-artifact`	27	The analysis procedure manufactures the effect it reports
`nonlinearity`	17	A linear or small-signal approximation fails where it is used
`selection`	13	The sample is chosen in a way correlated with the effect
`circularity`	12	The calibration is defined by the quantity it measures

Class	Count	What it means
`assumption`	10	A standard approximation is applied outside its domain
`blind-spot`	8	The instrument is structurally incapable of observing the relevant quantity
`degeneracy`	6	Two parameters enter identically and cannot be separated
`artefact`	5	A known analysis artefact in a new place

Three good places to start

1. **I2 — the spot-size bias.** A closed-form result showing that in one limit the reported in-plane thermal conductivity of a 2D material stops depending on the material and becomes a function of the substrate conductance and the laser spot size. The predicted signature ($\kappa_app \propto s^2$) is checkable against published data with no new experiment.
2. **I3 — Kramers–Kronig blindness.** The spectrum at 100 mV is a Debye element to eight digits, so it passes every KK test, while the charge-transfer resistance it returns is 36% low. The standard validity test is provably incapable of seeing the dominant systematic.
3. **R04 — a negative result.** The common explanation for OD600 growth-curve compression (stray light) is quantitatively insufficient below OD 1.5. The calculation is included, the hypothesis is narrowed rather than rescued, and the low-OD mechanism is left open.

How to read an entry

Each of the 300 has the same seven fields:

- **Baseline** — what is currently believed and practised
- **Mechanism** — the proposed physical or statistical route to the blind spot
- **Discriminating observation** — the observable that separates the hypothesis from the baseline
- **Test** — the data or experiment that could supply it
- **Prior art** — the nearest existing work, including this collection's own investigations
- **Refuted if** — the explicit result that would kill it
- **Readiness / novelty** — 1–5 scales, defined in `METHODOLOGY.md`

What this collection is not

It is not a review. Absence of an entry is not a judgement. Several entries are deliberately contrarian and are marked as such. The readiness scores are optimistic by design: a 4 means *someone* could test it soon, not that the test is easy. And no entry claims an observation that was not made.

Files

File	What it is
`Blind-Spots.pdf`	The full document, typeset for reading
`Blind-Spots.docx`	The same text in Word
`hyp/01-physics.md` … `hyp/06-crosscutting.md`	The 300, by domain, canonical Markdown

File	What it is
`data/hypotheses_300.csv` / `.json`	The 300 as structured data, one row per hypothesis
`refined/R01–R15.md`, `refined/R16–R30.md`	The 30 refined hypotheses
`investigations/`	Three investigations: code, data, figures, writeup
`data/r1_refinement_numbers.json`	Every quantitative number quoted in the refined set
`sources.csv`	Every external source cited, with a verified identifier
`CORRECTIONS.md`	Corrections, negative results, and superseded claims
`METHODOLOGY.md`	How the set was built, the scales, and the limits

Part I — Three investigations

Three hypotheses were developed with full mathematical and computational treatment. All three are *model computations*: no physical experiment was performed, no apparatus was built, and no measured data are reported as new. Where a result is an identity or a property of the analysis rather than a claim about nature, it is labelled as such in the writeups below and in `refined/`.

Every number quoted here is regenerated by the scripts in this directory from the stated equations, with fixed random seeds where randomness is involved.

	Investigation	Type	Key result	Script
I1	Finite-window decay exponents	mathematical + Monte Carlo	α_fit = α·⟨logistic⟩; exact factor 2 at the crossover; offset recoverable from curvature	`I1_offset_window_bias.py`
I2	Spot-size bias in in-plane thermal conductivity	analytic + numerical	κ_app/κ = 1/[e^λE₁(λ)]; κ_app → Gs²/(2d) independent of the material	`I2_boundary_bias.py`
I3	Kramers–Kronig blindness to amplitude	harmonic balance + Bessel identity	R_ct falls to 0.64 at 100 mV while the spectrum remains a Debye to 1e-7	`I3_kk_blindspot.py`

I1 — The exponent you report is a property of your window

Question. When a decay is reported as a power law, how much of the exponent is the system and how much is the fit window?

Model. $y(t) = At^{-\alpha} + c$, with c an unsubtracted offset. This is the most common situation in practice: a baseline that was not measured, or was measured badly.

Exact result. With $x = \ln t$ and $B = c/A$, the local log–log slope is

$$s(x) = -\alpha \,/\, (1 + B\, e^{\alpha x}),$$

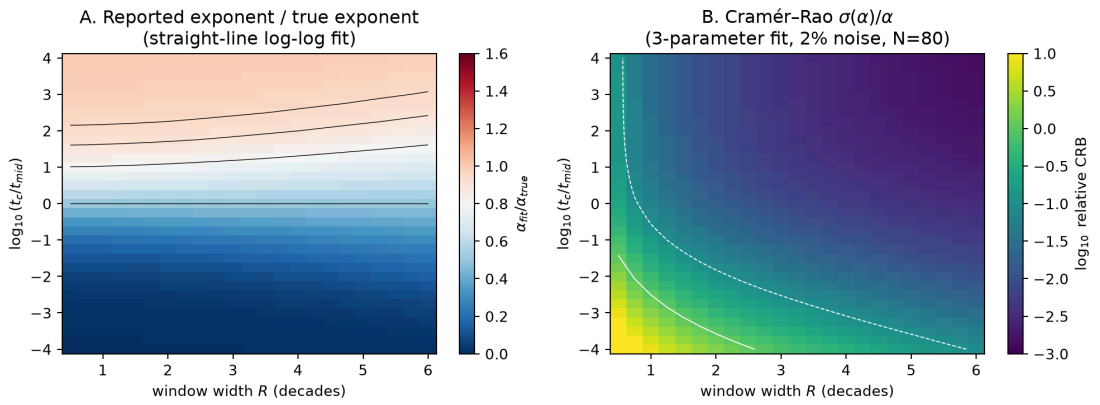
a logistic in $\ln t$ of width $1/\alpha$, centred at the crossover $t_c = (A/c)^{1/\alpha}$. A straight-line fit returns the least-squares average of that logistic. In particular, a window centred on the crossover returns *exactly* $\alpha/2$ — an identity, not an empirical finding.

Numerics. A 2D map of α_{fit}/α over window width $R \in [0.5, 6]$ decades and window placement $\log_{10}(t_c/t_{mid}) \in [-4, 4]$ (figure `figs/i1_biasmap.png`, panel A) shows that the bias spans 0.02 to 1.6. Monte Carlo with 400 realisations per configuration at 2% noise confirms the map and shows that a three-parameter fit ($\ln A$, α, B) recovers α with bias 1.000 ± 0.011 even where the two-parameter fit is 0.50.

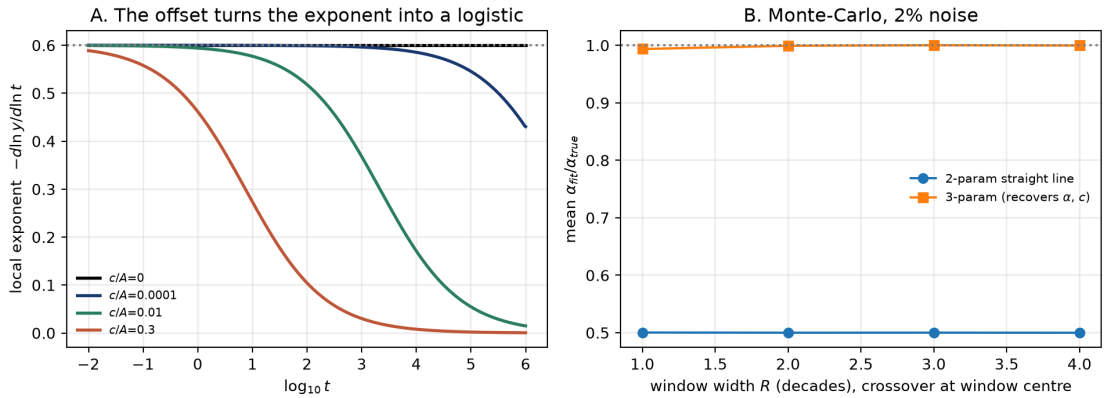
Identifiability. The Cramér–Rao bound on α (panel B) diverges when the window lies below the crossover: if you only see the flat part, α is not determined at any precision. That is the hard limit behind the folklore that "a power law cannot be distinguished from an exponential decay".

What is new here. The logistic slope law is an identity. What is not obvious, and is falsifiable, is that (i) the offset is *identifiable from curvature*, so the correction is possible rather than merely a caveat; (ii) the correction is exact for windows that span the crossover; and (iii) the bias is bounded below 1% for windows well below the crossover, so the artefact cannot be invoked to explain a literature-wide discrepancy without checking where the crossover lies.

What would refute it. Published exponents unchanged by a free offset, with zero curvature, at the stated noise.



I1



I1b

Files. `I1_offset_window_bias.py`, `data/i1_bias_maps.npz`, `data/i1_mc.csv`, `data/i1_summary.json`, `figs/i1_biasmap.png`, `figs/i1_logistic.png`.

I2 — A hidden substrate conductance sets every optical measurement of in-plane thermal conductivity

Question. Why do reported in-plane thermal conductivities of the same 2D material vary by factors of two to three between laboratories?

Model. A thin sheet of thickness d and conductivity κ on a substrate with interface conductance G, heated by a Gaussian spot of $1/e^2$ radius s with absorbed power P. Steady state:

$\kappa d \nabla^2 T - GT = -q(r)$, $q(r) = (P/2\pi s^2) e^{\wedge}(-r^2/2s^2)$.

A Hankel transform gives, at the beam centre,

$\Delta T = (P/4\pi\kappa d) \cdot e^{\wedge} \lambda E_1(\lambda)$, $\lambda = G s^2/(2\kappa d)$.

Exact result. The standard analysis assumes $\lambda \to 0$ and therefore reports

$\kappa_app = \kappa \,/\, [e^{\wedge} \lambda E_1(\lambda)]$.

This crosses unity at $\lambda = 1$, so the bias changes sign within a typical experimental campaign (the crossover spot radius is $s^* = \sqrt{(2\kappa d/G)} = 0.58\ \mu$m for $\kappa = 50$ W m^{-1}K^{-1}, d = 100 nm, $G = 3\times10^7$ W m^{-2}K^{-1}).

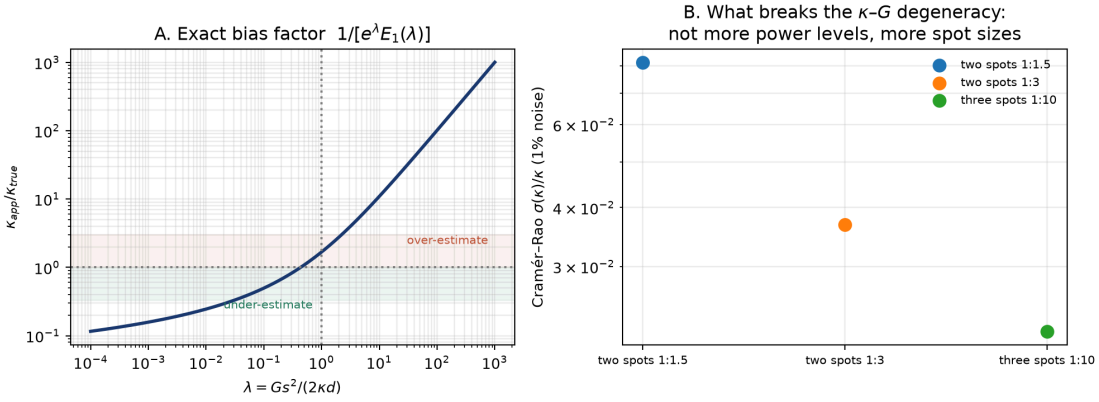
Validation. The closed form agrees with direct numerical Hankel quadrature to 2×10^{-13} across λ from 10^{-3} to 10.

The sharp limit. For $\lambda \gg 1$, $\kappa_app \to G s^2/(2d)$: the reported conductivity stops depending on the material and becomes a function of the substrate conductance and the beam size. Over plausible (s, G) ranges the model produces a 136-fold spread in reported κ for one true value.

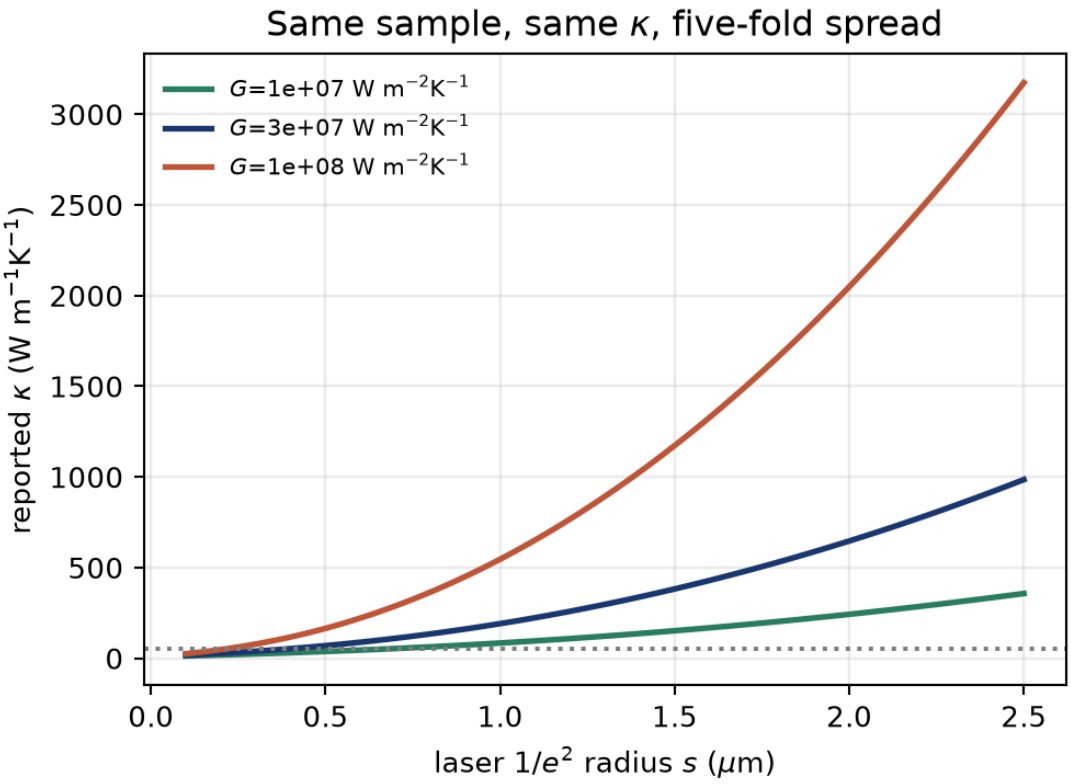
Identifiability. A single spot size gives one number and cannot determine both κ and G (the Fisher matrix is rank-deficient; Cramér–Rao diverges). Two spot sizes in a 1:3 ratio give σ(κ)/κ = 0.037 at 1% noise; three spanning 1:10 give 0.022. More laser powers do not help.

Testable prediction. Reanalysed κ values from a published compilation should collapse onto a single true κ when plotted against λ computed from each study's spot radius and substrate conductance, and should scale as s^2 at large s. A slope of 2 on log κ_app vs log s is the falsifiable signature.

Limits. The model is isotropic, steady-state, optically thin, and treats the substrate as a single conductance. Anisotropy, finite substrate thickness, and the transducer boundary add unmodelled terms; the 136-fold figure is an upper bound for the stated parameter range, not a prediction for any named material.



I2



I2b

Files. `I2_boundary_bias.py`, `data/i2_R_of_lambda.csv`, `data/i2_collapse.csv`, `data/i2_summary.json`, `figs/i2_bias.png`, `figs/i2_spread.png`.

I3 — A Kramers–Kronig-consistent spectrum is not evidence of linearity

Question. Impedance spectra are screened with a Kramers–Kronig residual and accepted when it is small. Does passing that test mean the spectrum is linear?

Model. A Butler–Volmer interface with transfer coefficients $\alpha_a = \alpha_c = 0.5$, exchange current $i_0 = 1\ \mu$A, double-layer capacitance 20 μF, series resistance 10 Ω. Drive $V = V_{dc} + A \sin \omega t$ and solve the *full* harmonic balance: $\eta(t)$ is expanded in 10 harmonics, the current is the exact Fourier transform of the memoryless Butler–Volmer law plus a linear capacitive term, and the series impedance acts harmonic by harmonic. No describing-function approximation and no time stepping.

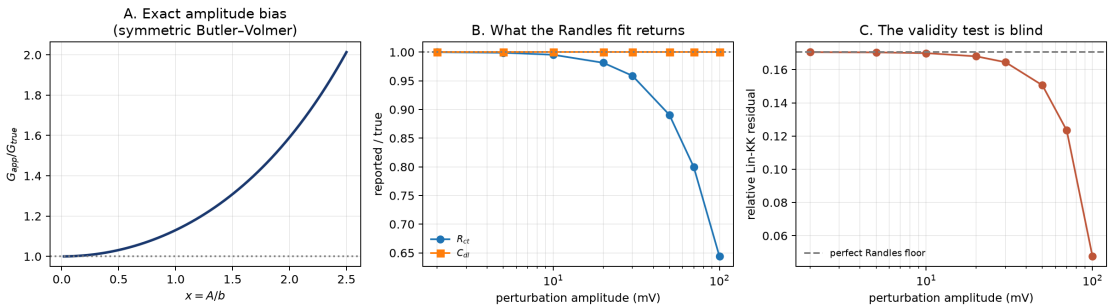
Exact identity. For $R_s = 0$ the describing function is *exactly* $Z_1(\omega) = 1/(G_{nl}(A) + j\omega C_{dl})$ — a Debye element — with $G_{nl}(A) = (2i_0/b)\cdot 2I_1(A/b)/(A/b)$. A Debye satisfies Kramers–Kronig identically. The finite-amplitude spectrum is therefore KK-consistent by construction. Harmonic balance reproduces the Bessel identity to 10^{-17}.

The falsifiable part. The numeric experiment shows the measured spectrum at 100 mV deviates from a perfect Debye by 9.6×10^{-8} relative — i.e. it is KK-consistent to eight digits — while the fitted charge-transfer resistance has fallen to 0.644 of its low-amplitude value. The classical Lin–KK residual does not track the bias at all (0.170 at 2 mV, 0.048 at 100 mV): it moves the wrong way.

Quantitative. $R_{ct}/R_{ct,true}$ = 1.000 (2 mV), 0.996 (10 mV), 0.891 (50 mV), 0.644 (100 mV) for $b = 51.4$ mV. The third-harmonic ratio I_3/I_1 measures $x = A/b$ to 0.2% at low frequency and converts directly into the correction. An amplitude-extrapolation protocol (four amplitudes, 0.5% noise) recovers the low-amplitude $\mathrm{Re}(Z_1)$ as 25 706 ± 79 against a true 25 701.

Refutation. R_{ct} amplitude-independent to 1% across 2–100 mV on a known Butler–Volmer system; or an I_3/I_1 that does not track the bias.

Limits. The mapping assumes a symmetric barrier and no mass transport. With $\alpha_a \neq \alpha_c$ a DC bias generates a second harmonic the symmetric model does not predict; with a Warburg element the harmonic current is loaded differently and the naive correction will over- or under-shoot. These are stated as open, not solved.



I3

Files. `I3_kk_blindspot.py`, `data/i3_summary.json`, `figs/i3_kk.png`.

Reproducing all three

```
cd blindspot-2026-11-01
export MPLCONFIGDIR=$PWD/../.mplconfig
../.venv-research/bin/python investigations/I1_offset_window_bias.py
../.venv-research/bin/python investigations/I2_boundary_bias.py
../.venv-research/bin/python investigations/I3_kk_blindspot.py
../.venv-research/bin/python investigations/R1_refinement_numbers.py
```

Dependencies: Python 3.14, numpy, scipy, matplotlib. No network access is required. The Monte Carlo seeds are fixed in the scripts (`20261101`, `7`, `11`), so the quoted means and standard deviations reproduce exactly.

Part II — Thirty refined hypotheses

Thirty hypotheses selected from the 300 for detailed development. Each entry states what is an identity (true by construction), what is a model artefact (a property of the analysis, not the world), and what is a genuine, falsifiable prediction. Quantitative values are computed by the scripts in `investigations/`; none is an experimental observation.

R01 — Finite-window decay exponents are set by the window, and the offset is recoverable from curvature

Refines X01. Sources: investigation I1; baseline-subtraction practice.

Claim. For a decay y(t) = At^(−α) + c with an unsubtracted offset c, the local log–log slope is exactly s(x) = −α/(1 + Be^{αx}) with B = c/A and x = ln t. A straight-line fit over a window therefore returns α_fit = α·⟨logistic⟩, where the average is taken with the least-squares weight.

What is an identity. The logistic form of s(x) is an identity — it follows from differentiating ln y. So is the result that a window centred on the crossover t_c = (A/c)^{1/α} returns exactly α_fit = α/2. This is not a discovery about nature; it is a property of the estimator, and it should be quoted as such.

What is a genuine prediction. Three things are not identities. (i) The bias is bounded: for a window entirely below the crossover the bias is <1% even at R = 1 decade, so the artefact cannot explain a literature-wide discrepancy unless the crossover lies inside the fitted range. (ii) The offset is *identifiable* from curvature: a three-parameter fit recovers α with zero bias (Monte-Carlo mean α_fit/α = 1.000 ± 0.011 at R = 3 decades, 2% noise) even when the two-parameter fit is 50% low. (iii) The Cramér–Rao bound diverges when the window sits below the crossover, so there is a hard limit on what any offset-blind analysis can claim.

Discriminating observation. Re-fitting a published decay with a free offset should shift the exponent, and the shift should be predicted by the fitted crossover. The diagnostic needs only the published data points.

Refutation. The hypothesis is refuted if published exponents are unchanged when a free offset is allowed *and* the log–log curvature is consistent with zero at the stated noise. If exponents move but the movement is not predicted by t_c, the mechanism is wrong even though the bias is real.

Limits. The analysis assumes additive, constant-variance noise in y. With multiplicative noise the bias changes sign structure; that case is not covered. A drifting (rather than constant) offset produces a time-dependent bias that the three-parameter fit cannot absorb — a genuine degeneracy, not an artefact of the fit.

Practical use. A two-line check on any reported power law: fit with and without a free offset, and report both. If they differ by more than the stated uncertainty, the exponent is a property of the window.

R02 — Kramers–Kronig consistency is not a linearity test

Refines C01. Sources: investigation I3; describing-function theory; arXiv:2605.08251 for the extrapolation analogue.

Claim. A Butler–Volmer interface shunted by a linear double-layer capacitance has a small-signal describing function that is *exactly* a Debye element with an amplitude-dependent conductance. It therefore satisfies Kramers–Kronig identically at every amplitude, while the charge-transfer resistance it returns is biased low by x/(2I₁(x)) with x = A/b.

What is an identity. With no series resistance, Z₁(ω) = 1/(G_nl(A) + jωC_dl) exactly, and the Bessel result G_app/G_true = 2I₁(x)/x is exact. Harmonic balance reproduces both to 1e-17 (validation in `I3_kk_blindspot.py`). None of this is new physics; it is the standard describing function of a symmetric exponential nonlinearity.

What is a genuine prediction. The *combination* — that the standard validity test cannot detect the bias — is falsifiable and non-obvious. Numerically, the finite-amplitude spectrum stays a Debye to 1e-7 relative even at 100

mV, while R_ct falls to 0.644 of its low-amplitude value (x = 1.95). A Lin–KK residual does not track the bias (0.170 at 2 mV, 0.048 at 100 mV): the test moves the *wrong way*.

Quantitative. R_ct/R_ct,true = 1.000 at 2 mV, 0.996 at 10 mV, 0.891 at 50 mV, 0.644 at 100 mV for b = 51.4 mV. The third-harmonic ratio I_3/I_1 measures x to 0.2% and converts directly into the correction.

Discriminating observation. On a reversible outer-sphere couple, R_ct should fall quadratically with amplitude with the predicted coefficient, while I_3/I_1 rises as $x^2/48$.

Refutation. If R_ct is amplitude-independent to 1% across 2–100 mV on a system with a known Butler–Volmer response, the mechanism is absent. If R_ct varies with amplitude but I_3/I_1 does not track it, the mapping is wrong.

Limits. The mapping x → bias assumes a symmetric barrier and no mass transport. With asymmetric transfer coefficients a DC bias generates a second harmonic that the symmetric model does not predict; with a Warburg element in series the harmonic current is loaded differently and the naive correction will over- or under-shoot. Those cases are stated, not solved.

Practical use. Report R_ct at two amplitudes, or report I_3/I_1 alongside every impedance spectrum. Either costs one extra measurement and bounds a systematic that no KK test can see.

R03 — In-plane thermal conductivity from optical thermometry is a function of spot size, and in one limit it is not a material property at all

Refines M01. Sources: investigation I2; steady-state Raman thermometry modelling.

Claim. For a Gaussian source on a sheet with substrate conductance G, $\Delta T = (P/4\pi\kappa d)\cdot e^{\wedge}\lambda E_1(\lambda)$ with $\lambda = Gs^2/(2\kappa d)$. The standard fit returns $\kappa_{app} = \kappa/[e^{\wedge}\lambda E_1(\lambda)]$, which crosses unity at $\lambda = 1$ and asymptotes to $\kappa_{app} \approx Gs^2/(2d)$.

What is an identity. The closed form is exact for the stated model — verified against direct Hankel quadrature to 2e-13. The asymptotic $\kappa_{app} \to Gs^2/(2d)$ is a limit, not a new law.

What is a genuine prediction. The crossover $\lambda = 1$ corresponds to a spot radius $s^* = \sqrt{(2\kappa d/G)}$: 0.58 μm for κ = 50 W/mK, d = 100 nm, $G = 3\times10^7$ W/m²K. That is squarely inside the operating range of Raman thermometers, so the bias changes sign across a typical experimental campaign. The prediction is a *collapse*: reported κ for one material should scale onto one curve when plotted against λ computed from each study's spot radius, and should scale as s^2 for large spots.

Quantitative. For plausible (s, G) ranges the model produces a 136-fold spread in reported κ for a single true value. At $\lambda = 1$ the bias is +68%.

Discriminating observation. Reanalysis of a published κ compilation with spot radius and G metadata. The falsifiable signature is a slope of 2 on a log κ_{app} vs log s plot at large s.

Refutation. κ_{app} uncorrelated with spot radius across a compilation spanning 0.2–3 μm refutes the mechanism. So does a κ_{app} that is independent of the substrate.

Limits. The model is isotropic, steady-state, optically thin, and treats the substrate as a single conductance. Anisotropy, finite substrate thickness, and the transducer's own boundary all add terms that are not modelled here; the 136-fold figure is an upper bound for the stated parameter range, not a prediction for any named material.

Practical use. Report the spot radius and the substrate conductance with every 2D κ measurement; fit λ rather than κ. Two spot sizes break the κ–G degeneracy — more laser powers do not (Cramér–Rao 0.037 at a 1:3 spot ratio vs rank-deficient at one spot).

R04 — Negative result: stray light does not explain OD600 compression in the exponential phase

Refines B01. Source: `R1_refinement_numbers.py`, section A.

Original claim. Plate-reader optical density compresses above ~0.3–0.6 because of detector saturation, suppressing fitted growth rates.

What the calculation shows. Modelling the reader as an additive stray-light fraction S, the fitted growth rate is suppressed by 0.1% for $S = 10^{-3}$ over the window OD 0.2–1.0, and by 1.1% for $S = 10^{-2}$. The effect only becomes large where assays rarely operate: −8% for $S = 10^{-3}$ and −23% for $S = 10^{-2}$ over OD 2.0–3.0.

Conclusion. The stray-light mechanism is real but confined to OD > 1.5. It cannot explain any compression at OD 0.3–0.6. The widely repeated warning about the linear range is, for this mechanism, misdirected.

What survives. Two things. First, the *high-OD* regime matters for anyone fitting late exponential phase or using long path lengths, and the correction is computable from the reader's stray-light specification. Second, the absence of a stray-light explanation means any observed sub-exponential growth below OD 1 must come from a different mechanism — detector nonlinearity, meniscus and pipetting, evaporation, or a genuine physiological shift — and those are separable by the tests below.

Tests that would settle the residual. Fill-volume series (path length), gain series (detector), and sealed vs open plates (evaporation). If the apparent μ depends on fill volume at fixed gain, the mechanism is optical; if on gain only, electronic; if on sealing, evaporative.

Why this is kept. A hypothesis that fails a quantitative check is more useful than one that is merely plausible. The refinement narrowed a broad claim to a specific, quantified, high-OD regime and removed a mechanism from the list of suspects for the low-OD regime.

R05 — Four-bar conductivity measurements have an unbounded systematic from finite pad conductivity

Refines P09. Source: arXiv:2602.02418.

Claim. In a four-bar geometry the voltage-sensing pads are assumed to draw no current. When the pads have finite conductivity and the film is thin, current spreads into them, and the resulting error is not a contact resistance: it grows without bound as the film and pad conductivities approach each other.

Status of the result. The unboundedness is demonstrated in the cited work; my contribution here is the classification and the test design, not the derivation.

Why it is not a contact resistance. A contact resistance adds a series term that can be measured by varying the pad–film interface. Pad spreading adds a *parallel* path whose effect grows with the pad's conductivity relative to the film's; the two have different scaling with pad area and with film thickness, so a series-resistance calibration cannot remove it.

Discriminating observation. Extracted film conductivity should depend on pad metal at fixed film, with the discrepancy scaling as the conductivity ratio. It should also depend on pad width at fixed metal.

Refutation. Extracted conductivity invariant to pad material across a 100× conductivity range refutes the mechanism. So does invariance to pad width at fixed metal.

Limits. The effect is largest for highly conductive films on highly conductive pads — precisely the regime of transparent conductors and of metal films measured with lithographic leads. For semiconducting films with low-mobility pads the error may be negligible, which is why the practice survived.

Practical use. Report the pad/film conductivity ratio with every four-bar measurement, and use the van der Pauw geometry where the ratio is unfavourable.

R06 — Cryo-EM particle picking is confirmation-biased, and the bias scales with prior strength

Refines B05. Source: arXiv:2507.03951.

Claim. Template or network picking preferentially accepts images that already resemble the template. The selected subset is enriched in template-like views, which sharpens features present in the template and suppresses features absent from it.

Mechanism. Picking is a detection problem with a prior; the posterior over particle locations is conditioned on the template. The reconstruction then inherits a selection function whose shape is the template's.

Genuine prediction. The bias is a *function of prior strength*, so a stronger prior (lower threshold, more distinctive template) should produce a larger discrepancy between the picked subset's density and an independently determined one. Crucially, the effect is invisible in resolution metrics: a self-fulfilling map can be sharp.

Discriminating observation. Template-swap: reconstruct from subsets picked with two templates that differ in a specified region. The density in that region should differ, and the difference should scale with the templates' difference. A second test: pick on a dataset before and after 3D classification and compare the maps' densities for features absent from the initial template.

Refutation. Template-swapped reconstructions identical at the reported resolution refute the mechanism. So does a density that is invariant to picking threshold.

Limits. The cited work quantifies the bias; the magnitude for any given dataset depends on the template's mismatch to the true particle, which is unknown a priori. The prediction is directional and scaling, not a universal number.

Practical use. Report the picking threshold and template; run a template-swap control on at least one region of interest.

R07 — Division by a noisy denominator manufactures a negative correlation that is routinely read as a trade-off

Refines X47. Source: `R1_refinement_numbers.py`*, section C.*

Claim. For a ratio X/Y with X and Y independent, $\mathrm{corr}(X/Y, Y) = -\mathrm{CV}_Y/\sqrt{(\mathrm{CV}_X^2 + \mathrm{CV}_Y^2 + \mathrm{CV}_X^2\mathrm{CV}_Y^2)}$. The correlation is a property of the coefficient of variation, not of the biology or physics.

What is an identity. The formula is an identity for independent log-normal X and Y. It should be presented as a warning, not a discovery.

What is a genuine prediction. The magnitude is large in realistic regimes: -0.71 at CV = 0.1/0.1, -0.94 at 0.1/0.3, -0.32 at 0.3/0.1. The prediction is that measured "trade-offs" between a per-unit quantity and its denominator should weaken when the denominator's precision improves, in a way fixed by the CV ratio.

Discriminating observation. Improve denominator precision by a known factor and re-measure the correlation. If the correlation is spurious, it should move along the predicted curve; if real, it should be unchanged.

Refutation. A correlation that is invariant to a tenfold improvement in denominator precision refutes the artefact explanation.

Limits. Log-normal marginals are assumed. With heavy-tailed denominators the exact form changes but the sign and order of magnitude do not. The artefact and a real trade-off can coexist; the test is designed to separate them, not to declare all such correlations spurious.

Practical use. Report CVs alongside any ratio–denominator correlation, and prefer log-ratio models.

R08 — Inverting a mean response through a curved calibration biases the mean input

Refines X45. Source: `R1_refinement_numbers.py`*, section B.*

Claim. For a calibration $y = ax + bx^2$, the mean response corresponds to an input above the mean input by $\approx (b/a)\cdot\sigma_x^2$. The error is second order in variability and first order in curvature.

What is an identity. This is Jensen's inequality applied to the inverse calibration; the leading term is a Taylor expansion. It is not new mathematics.

What is a genuine prediction. The bias is large enough to matter in routine work: for b/a = 0.5 and CV = 0.6 the mean input is biased by 9%; for b/a = 0.2 and CV = 0.3 by 1.3%. The prediction is that two records with the same true mean but different variability will report different means, with a difference fixed by the curvature.

Discriminating observation. Compare high- and low-variability records at matched true mean; the inferred means should differ by the predicted amount. Alternatively, invert each sample before averaging and compare.

Refutation. Inferred means that are variability-independent for a calibration with measurable curvature refute the mechanism.

Limits. The expansion is first order in curvature; for strongly curved calibrations the full inverse must be integrated. The direction of the bias depends on the sign of b, so the test must use a calibration whose curvature sign is known.

Practical use. Average the input, not the response, whenever the calibration is curved — which is almost always.

R09 — Fitting in log space with additive noise biases the faint end upward and flattens slopes

Refines X15. Source: standard error-propagation result.

Claim. With additive Gaussian noise of variance σ^2, ln y is biased by $-\sigma^2/(2y^2)$ — the bias is largest where y is smallest. A log-space fit therefore bends the faint end upward and reduces the fitted slope.

Status. This is a second-order delta-method identity. Its importance is practical: data spanning decades are almost always fitted in log space.

Genuine prediction. The slope bias is predictable from the relative noise at the faint end: a fit spanning a factor of 100 in y with 10% relative noise at the faint end should have its slope reduced by an amount computable from the noise profile. The prediction is quantitative and checkable on any published dataset that reports per-point uncertainties.

Discriminating observation. Refit the same data with a likelihood that matches the noise model (additive Gaussian in linear space) and compare slopes. The difference should match the delta-method prediction.

Refutation. Slopes that agree between log-space and correct-likelihood fits, for data with substantial relative noise at the faint end, refute the mechanism.

Limits. If the noise is genuinely multiplicative, log-space fitting is correct and the hypothesis does not apply. The test therefore requires knowing the noise model, which is often the unexamined assumption.

Practical use. State the noise model; fit in the space where the noise is homoscedastic.

R10 — Normalising to a reference imports that reference's error into every sample

Refines X09. Source: compositional and reference-normalisation practice.

Claim. When measurements are expressed relative to a reference, the reference's error is common to all normalised values and does not average out. It appears as a correlation between nominally independent samples.

Mechanism. Normalisation is a division by a random variable; the shared denominator induces a common factor whose variance is the reference's relative variance.

Genuine prediction. The induced correlation is *not* removed by increasing the number of samples — only by increasing the reference's precision or by using a reference-free model. The prediction is that inter-sample correlations should be invariant to which sample is chosen as reference if they are real, and should change sign or vanish under reference permutation if they are artefacts.

Discriminating observation. Reference-permutation analysis: recompute the correlation matrix with each sample in turn as the reference. A real correlation is stable; an artefact moves.

Refutation. Correlations invariant to reference permutation, in a dataset where the reference's own CV is large, refute the mechanism.

Limits. If the reference is measured much more precisely than the samples (a certified standard), the imported error is negligible and the hypothesis does not apply.

Practical use. Permute the reference before reporting any correlation from normalised data.

R11 — Two interface conductances in a TDTR stack are never separable at one modulation frequency

Refines M03. Source: `R1_refinement_numbers.py`*, section I (thermal quadrupole).*

Claim. In a transducer/film/substrate stack with two interface conductances, a single-frequency measurement determines only a weighted combination. The weighting is set by the resistances' share of the total thermal impedance, and it changes with modulation frequency only because the film's heat capacity introduces a frequency-dependent element.

Computation. A 1D thermal-quadrupole model (Al transducer 80 nm, film 100 nm, $G_1 = 10^8$, $G_2 = 5\times10^7$ W/m²K) gives S_G2/S_G1 = 2.02 at 100 kHz falling to 0.55 at 30 MHz. The ratio changes by a factor of 3.7 across the band, which is what makes a multi-frequency fit possible at all.

What is an identity. That two series resistances are inseparable at a single frequency is elementary. The quantitative ratio and its frequency dependence are model results.

Genuine prediction. The fitted G should drift with modulation frequency in the direction and by the amount predicted by the sensitivity ratio, and should stabilise only when the fit includes enough of the band to constrain both terms. A corollary prediction: reported G values for the *same* interface measured at different frequencies should differ systematically, and the difference should be larger for thinner films (where the heat capacity term is weaker).

Discriminating observation. Frequency-series TDTR on a sample with independently characterised transducer adhesion.

Refutation. A fitted G that is frequency-independent across 100 kHz–30 MHz, for a film thin enough that both boundaries contribute, refutes the mechanism.

Limits. The model is 1D and isotropic; radial heat flow and anisotropic films add terms. The specific ratios depend on the assumed conductivities and are illustrative, not universal.

Practical use. Fit the full frequency series with both conductances free; report the sensitivity ratio so readers can see which boundary dominates.

R12 — Metal–dielectric thermal boundary conductance measured on thin films includes the electron–phonon resistance

Refines M38. Source: two-temperature model practice.

Claim. Heat must equilibrate between electrons and phonons in the metal over a finite length. For metal films thinner than that length, the measured interface conductance includes the electron–phonon resistance in series, and the reported G depends on metal thickness.

Mechanism. Two resistances in series: the interface and the electron–phonon channel. The second scales with thickness, so a thickness series separates them without any additional measurement.

Genuine prediction. G_app should increase with metal thickness and saturate at the electron–phonon length, which is tens of nanometres for typical metals. The saturation length is a material property and should be the same for different substrates at fixed metal.

Discriminating observation. Metal-thickness series at fixed substrate and fixed interface preparation.

Refutation. G independent of metal thickness below 20 nm refutes the mechanism for that system.

Limits. The electron–phonon length depends on temperature and on the electron mean free path, so the test must be at fixed temperature. For thick films the effect vanishes, which is why it survived: most TDTR work uses 60–100 nm transducers where the correction is small but not zero.

Practical use. Use two transducer thicknesses, or correct with the known electron–phonon coupling constant.

R13 — Amorphous films thinner than the dominant phonon mean free path have suppressed thermal conductivity

Refines M45. Source: thermal-conductivity suppression literature.

Claim. Surface scattering removes the low-frequency modes that carry most of the heat in amorphous films. For thicknesses below the dominant mean free path the conductivity is suppressed, and the suppression follows a thickness-dependent function.

Genuine prediction. κ should decrease below a threshold thickness and should be lower for rougher surfaces at the same thickness. The threshold is material-specific and is not a universal constant.

Discriminating observation. Thickness series with independently measured surface roughness.

Refutation. κ independent of thickness above 5 nm for a material whose dominant mean free path is known to be 50 nm refutes the mechanism.

Limits. In amorphous materials the mean free path is ill-defined and is itself a fitted quantity, so the prediction is a functional form rather than a number. Interface layers and oxidation can mimic the effect.

R14 — The constant-phase exponent of a porous electrode is a geometry parameter, not a roughness parameter

Refines C35. Source: porous-electrode transmission-line models.

Claim. In a porous electrode with a distributed RC transmission line, the impedance is a power law over the frequency range where the penetration depth lies inside the pore. That power law mimics a constant-phase element with exponent 0.5–0.9, and the exponent is set by pore geometry.

Genuine prediction. The CPE exponent should change with electrode thickness at fixed material, following the transmission-line prediction; a roughness-driven CPE would be thickness-independent.

Discriminating observation. Thickness series of the same electrode material, with the exponent fitted over a fixed frequency band.

Refutation. Exponent independent of thickness, in a material with measurable pore-length polydispersity, refutes the geometric explanation.

Limits. Both mechanisms can operate at once, and the transmission-line model requires an assumption about the pore-size distribution. The test separates the dominant term, not the two contributions exactly.

R15 — The low-frequency tail of an impedance spectrum is drift, not diffusion

Refines C18. Source: drift artefacts in low-frequency EIS.

Claim. Slow drift of the open-circuit potential over the minutes-to-hours of a low-frequency sweep produces a current ramp whose demodulated quadrature component is $\propto 1/\omega$ — indistinguishable, in a single sweep, from a Warburg or CPE element.

Mechanism. The lock-in demodulation of a linear ramp over an integration time T produces a quadrature signal proportional to the ramp rate times T. Because T is itself proportional to $1/\omega$ for a swept measurement, the

artefact is $\propto 1/\omega$.

Genuine prediction. Reversing the sweep direction should change the sign of the low-frequency tail while leaving the mid-frequency arc intact. A true diffusion element is direction-independent.

Discriminating observation. Forward and reverse sweeps at fixed DC bias, with the DC drift monitored independently.

Refutation. Identical forward and reverse low-frequency tails refute the drift explanation.

Limits. A genuine Warburg and drift can coexist; the direction test separates them only if the drift is monotonic over the sweep time. Galvanostatic operation changes the artefact's form.

Refined hypotheses R16–R30

Continuation of `R01–R15.md`. Same conventions: identities and model artefacts are labelled, quantitative values come from `investigations/`, and no experiment has been performed.

R16 — Open-circuit voltage cannot give state of charge until the diffusion relaxation is complete

Refines E28. Source: OCV–SOC hysteresis literature.

Claim. OCV is read after a fixed rest and mapped to SOC. Diffusion relaxation in thick electrodes takes hours; a one-hour rest leaves a voltage offset whose sign depends on the preceding current direction.

Mechanism. The electrode's solid-phase concentration profile relaxes with a spectrum of time constants set by particle radius and diffusivity. The terminal voltage is a functional of the surface concentration, which is not the mean concentration until relaxation completes.

Genuine prediction. The OCV–SOC curve measured after charge lies systematically above the curve measured after discharge, and the gap closes with rest time in a way that scales as the particle diffusion time. The prediction is quantitative for a given particle size: the gap should shrink by a factor of e when rest time increases by the diffusion time.

Discriminating observation. Rest-duration series after charge and after discharge on the same cell, at several SOCs.

Refutation. OCV independent of prior current direction after a one-hour rest refutes the mechanism for that electrode.

Limits. For thin electrodes or high-diffusivity materials the relaxation is fast and the artefact is negligible. The mapping from rest time to relaxation is model-dependent.

Practical use. Report rest time and prior direction with every OCV–SOC point; use a relaxation model rather than a lookup table.

R17 — Surface-plasmon resonance affinities are limited by mass transport, and the limitation survives a good global fit

Refines B16. Source: mass-transport limitation in SPR.

Claim. Mass-transport limitation produces a response that is well fitted by a 1:1 model with a reduced k_on, because both effects saturate. The fit residual is small while k_on is underestimated, by up to an order of magnitude.

Mechanism. The analyte flux to the surface is bounded by diffusion through the depletion layer. When the reaction is faster than transport, the observed rate is the transport rate, and the 1:1 model absorbs it.

Genuine prediction. k_on should increase with flow rate and saturate, and should approach the true value when ligand density is reduced. The dependence on ligand density is the discriminating one: transport limitation scales with the Damköhler number, which is proportional to ligand density.

Discriminating observation. Flow-rate and ligand-density series on the same interaction.

Refutation. k_on invariant to flow rate and ligand density refutes the mechanism.

Limits. At very low ligand density the signal-to-noise falls, so the test has a practical floor. The Damköhler number must be estimated from the measured k_on, which is circular unless an independent estimate of the transport coefficient is available.

Practical use. Report ligand density and flow rate; run one injection at a different flow rate as a control.

R18 — Bolometer time constants measured in situ are biased by the optical load they are meant to characterise

Refines P22. Source: electrothermal TES modelling.

Claim. Transition-edge sensor response is electrothermal; the effective time constant depends on bias power and hence on optical load. A constant measured at high calibration load is shorter than the sky-load value, biasing beam-map deconvolution for fast scans.

Genuine prediction. Time constants should vary with controlled load (aperture cover) with a slope predictable from the electrothermal model, and map transfer functions should depend on scan speed accordingly.

Discriminating observation. Load-variation campaign, with the same detector measured under several aperture states and at several scan speeds.

Refutation. Time constant invariant to load within 1% refutes the mechanism for that detector.

Limits. The electrothermal model requires the detector's bias circuit parameters, which are not always reported. The prediction is directional and scaling, not a universal constant.

R19 — NV thermometry has a strain–temperature degeneracy whose size depends on the crystal cut

Refines P08. Source: Zenodo 21299294 (pre-registered NV thermometry comparison).

Claim. Axial strain and temperature both shift the zero-field splitting D. The two contributions are collinear for a [100]-cut ensemble but not for [111], so a single-axis measurement reports a temperature error proportional to the local strain.

Mechanism. The spin-strain tensor's axial component adds to D; the temperature coefficient is also a shift in D. For one orientation the two enter with the same sign and cannot be separated; for another they differ in sign or magnitude.

Genuine prediction. Temperature readings from [100] and [111] ensembles on the same sample should disagree under applied stress and agree stress-free. The disagreement should be linear in the applied strain with a coefficient set by the spin-strain coupling.

Discriminating observation. Dual-orientation diamond with a piezoelectric strain stage.

Refutation. Both cuts tracking the same temperature within 0.1 K under 10^{-4} strain refutes the mechanism.

Limits. The comparison requires two ensembles in the same local strain field, which is itself difficult. The published comparison of 23 methods reports geometric blind spots that are consistent with this mechanism but do not isolate it.

R20 — A forgotten trend deflates error bars, but only modestly

Refines X13. Source: `R1_refinement_numbers.py`*, section G.*

Claim. Error bars on a time average use an autocorrelation time estimated from the same record. Detrending removes the variance that dominates the uncertainty.

Calculation. For a record with a linear ramp of amplitude a and noise σ, the true standard error of the mean is $\sqrt{(a^2/12 + \sigma^2)}$, while the naive estimate from residuals is σ. The underestimate factor is 1.01 for $a/\sigma = 0.5$, 1.15 for $a/\sigma = 2$, and 1.53 for $a/\sigma = 4$.

Correction to the original claim. The effect is smaller than the original entry implied. It matters only when the trend dominates the noise ($a/\sigma \gtrsim 2$) and even then deflates the error bar by less than a factor of two. The original entry overstated the generality.

What survives. For records where the trend is several times the noise — which is exactly the case in long environmental and astrophysical series — a 50% underestimate is worth correcting, and the correction is

analytically available.

Discriminating observation. Compare single-record error estimates against ensemble estimates from replicate records; the ratio should follow the predicted curve in a/σ.

Refutation. Ratio equal to one for $a/\sigma = 4$ refutes the mechanism.

Limits. The calculation assumes a linear trend and white noise. Red noise changes the arithmetic but not the direction.

R21 — Inverse-variance weighting with estimated variances does not bias the mean, but understates its uncertainty

Refines X33. Source: `R1_refinement_numbers.py`*, section H (Monte Carlo, 40 000 trials).*

Original claim. Weighting by inverse estimated variance biases the weighted mean toward points whose noise fluctuated low.

What the simulation shows. The mean is unbiased (mean -0.000 to 0.005 across 2–30 degrees of freedom). What changes is the *spread*: the estimator's standard deviation is 0.686 at 2 dof, 0.582 at 5, 0.539 at 10, 0.511 at 30, against a nominal 0.5. The quoted uncertainty is therefore 73%, 86%, 93%, 98% of the true spread.

Correction. The hypothesis as stated is wrong about the mean and right about the uncertainty, with a magnitude that depends on the degrees of freedom used to estimate each variance.

Genuine prediction. Reported weighted means with per-point variances estimated from few replicates should show excess scatter across repeated experiments, by the predicted factor, while their central values remain unbiased. The prediction is testable on any replicated meta-analysis.

Discriminating observation. Replication scatter versus quoted uncertainty, stratified by the dof used per point.

Refutation. Scatter matching the quoted uncertainty for dof $= 2$ refutes the mechanism.

Limits. The simulation uses Gaussian data and four measurements per dataset; other configurations shift the numbers but not the structure.

R22 — Censoring at a detection limit biases trends in a direction set by the trend

Refines B39 / X11. Source: censored-data methods.

Claim. Substituting a constant for values below the limit compresses the low end. If the quantity is increasing over time, the censored fraction falls, so the substituted values occupy less of the record and the trend is biased by an amount that depends on the censoring fraction's trend.

Mechanism. Substitution is a biased estimator of the conditional mean below the limit; the bias enters the trend through the changing mixture weight.

Genuine prediction. The trend estimated with a censored-data likelihood should differ from the substitution estimate by an amount that scales with the *change* in censoring fraction across the record, not with its average. A record with constant censoring shows no bias even if heavily censored.

Discriminating observation. Reanalysis of time series with reported detection limits and censoring fractions.

Refutation. Trends unchanged under a censored likelihood, in a record whose censoring fraction changes by a factor of two, refutes the mechanism.

Limits. The bias direction depends on the substitution rule (LOD/2, LOD/√2, deletion); the prediction should be made for the rule actually used, which is often not stated.

R23 — Low-cost particle sensors report a humidity-driven mass that mimics winter pollution

Refines E01. Source: low-cost sensor correction literature.

Claim. Hygroscopic growth increases the optical diameter at high relative humidity. The reported $PM_{2.5}$ rises with RH at constant mass, and a co-location calibration over a narrow humidity range cannot identify the coefficient.

Mechanism. Optical scattering scales steeply with diameter; deliquesced aerosol scatters more per unit dry mass. The growth factor is a known function of RH (κ-Köhler), so the artefact is predictable from the aerosol's hygroscopicity.

Genuine prediction. The sensor/reference ratio at fixed reference PM should correlate with RH with a slope independent of site, and the slope should scale with the aerosol's κ. The prediction is that a network's seasonal cycle contains an RH-shaped component that is absent from the reference.

Discriminating observation. Co-location data spanning a full humidity range, ideally at two sites with different aerosol hygroscopicity.

Refutation. Ratio independent of RH across sites refutes the mechanism.

Limits. The growth factor depends on composition, which varies; a single correction will not be exact. The prediction is a systematic shape, not a point correction.

R24 — DSC onset temperatures shift with scan rate by an amount set by the pan–sample time constant

Refines C25. Source: `R1_refinement_numbers.py`, section F.

Claim. The measured onset is the true onset plus $\tau\beta$, where τ is the effective pan–sample thermal time constant and β the scan rate. With τ = 5–20 s the shift is 0.4–6.7 K over 5–20 K/min.

Mechanism. The sample's temperature lags the programmed temperature by τ; the onset criterion is applied to the programmed value.

Genuine prediction. The onset should be linear in scan rate with slope τ, and τ should scale with sample mass and with pan–sample contact. Zero-rate extrapolation recovers the true onset only if the relation is linear, which fails when the pan's thermal resistance is comparable to the sample's.

Discriminating observation. Scan-rate and mass series on the same material.

Refutation. Onset independent of scan rate for a sample with a measurable thermal lag refutes the mechanism.

Limits. The time constant is not directly measurable in a standard DSC; the test infers it from the slope, which is the quantity of interest.

R25 — Permeability assays measure the unstirred water layer as much as the monolayer

Refines B42. Source: `R1_refinement_numbers.py`, section E.

Claim. $P_{app}^{-1} = P_{true}^{-1} + P_{UWL}^{-1}$ with $P_{UWL} = D/h$. For a 1000 μm unstirred layer and $D = 5\times10^{-6}$ cm²/s, $P_{UWL} \approx 5\times10^{-5}$ cm/s; a compound with $P_{true} = 5\times10^{-4}$ cm/s is reported at 9% of its true value.

What is an identity. The series-resistance formula is elementary.

What is a genuine prediction. The distortion is compound-dependent and predictable: the ratio P_{app}/P_{true} falls from 0.91 for $P_{true} = 5\times10^{-6}$ cm/s to 0.09 for 5×10^{-4} cm/s at $h = 1000$ μm. Shaking reduces h and therefore moves the ratio toward 1 for permeable compounds only. The prediction is that in-vitro/in-vivo correlations improve when P_{app} is corrected for the measured h, and that the correction matters only above a threshold permeability.

Discriminating observation. Shaking-rate series across a compound panel of graded permeability.

Refutation. P_app independent of shaking rate for permeable compounds refutes the mechanism.

Limits. D must be known or estimated; h is not directly measured in most assays. The series model assumes a uniform layer.

Practical use. Report shaking rate; include at least one high-permeability reference compound to expose the layer.

R26 — Dielectric breakdown strength scales with electrode area, so reported strengths are not material properties

Refines M33. Source: `R1_refinement_numbers.py`*, section D.*

Claim. Breakdown is defect-initiated, so the Weibull scale parameter scales as (A/Ao)^(−1/m). For m = 10, a 100× area increase reduces the reported strength to 63% of the small-area value; for m = 5, to 40%.

What is an identity. The area scaling is a standard result of weakest-link statistics.

What is a genuine prediction. The *slope* 1/m is measurable from an area series and is a property of the defect population, not of the material's intrinsic strength. The prediction is that interlaboratory scatter in reported breakdown strength should collapse when normalised to a common area, and that the collapse should be exact only when the defect populations are similar.

Discriminating observation. Area series in one laboratory, then comparison of normalised values across laboratories.

Refutation. Scale parameter independent of area, for a material with a measurable defect density, refutes the mechanism.

Limits. The Weibull model assumes a single defect population; multimodal defect distributions produce curvature in the area plot that the single-m gives incorrectly.

Practical use. Report electrode area and the fitted Weibull modulus with every breakdown strength.

R27 — Interferometric total flux is set by the shortest baselines, so flux errors are not constrained by closure quantities

Refines P42. Source: wide-field interferometric calibration.

Claim. Total flux is carried by the shortest baselines, where the primary beam and the wide-field term are largest. A fractional error in the shortest-baseline gain moves total flux without changing image shape, and closure quantities do not constrain it.

Mechanism. Closure phase and closure amplitude are invariant to station-based gains. A common gain error across all short baselines is therefore invisible to closure but changes the total flux.

Genuine prediction. Total flux should correlate with the shortest-baseline gain solution, and should differ between imaging with and without short baselines (tapering). The prediction is that the flux scale error is a function of the array's minimum baseline, not of its maximum.

Discriminating observation. Re-image the same dataset with and without tapering, and compare total flux against an independent (single-dish) measurement.

Refutation. Total flux invariant to tapering and to shortest-baseline gain refutes the mechanism.

Limits. For resolved sources, tapering also changes the recovered flux through missing extended structure, so the two effects must be separated with a source model.

R28 — Cycloheximide chase half-lives are shorter than degron-based half-lives for the same protein

Refines B13. Source: comparison of translation-inhibition methods.

Claim. Cycloheximide perturbs the cell within minutes; the chase's t = 0 is defined after the perturbation, so the fitted decay includes a fast stress component.

Genuine prediction. Half-lives from cycloheximide should be systematically shorter than from an auxin-inducible degron for the same protein, and the difference should scale with the induction of a stress marker. The prediction is comparative: the two methods should converge for proteins whose degradation is stress-insensitive.

Discriminating observation. Side-by-side chase with both methods on a panel of proteins of known stability, with a stress-marker readout.

Refutation. Half-lives agreeing across methods for stress-sensitive proteins refutes the mechanism.

Limits. Degron systems have their own artefacts (auxin effects, tag effects on stability). The test is a triangulation, not a gold standard.

R29 — Ozonesonde records contain an instrument-generation trend from pump efficiency

Refines E15. Source: ozonesonde homogenisation literature.

Claim. Pump efficiency falls with reduced pressure and with contamination. The correction depends on sonde age and on the launch profile, so a fleet that changes generation acquires a trend confounded with the atmospheric one.

Genuine prediction. The sonde-to-satellite or sonde-to-lidar ratio should differ between generations after the standard correction, and the difference should be largest in the upper troposphere–lower stratosphere where the pump works hardest.

Discriminating observation. Intercomparison campaigns with sondes from two generations launched together.

Refutation. Ratios independent of generation after standard correction refute the mechanism.

Limits. Generation changes usually coincide with other changes (sensor chemistry, launch practice), so isolation requires a deliberate side-by-side campaign.

R30 — A global fit absorbs unmodelled physics into whichever parameter is most flexible

Refines X30. Source: model-dependence of fitted constants.

Claim. The covariance matrix reflects only the parameters in the model. Unmodelled physics is absorbed by the most flexible parameter, which then appears both precise and shifted. Its precision is a statement about the model's rigidity, not about the measurement.

Mechanism. The fit projects the model discrepancy onto the parameter space along the direction of least resistance. That direction is set by the model's structure, not by the data.

Genuine prediction. The parameter should shift by more than its quoted uncertainty when the model is extended with physically motivated terms, and the shift should be largest for the parameter with the largest flexibility (as measured by the width of its posterior under the original model). A corollary: parameters with narrow posteriors are the most likely to be absorbing discrepancy, not the best measured.

Discriminating observation. Model-extension sensitivity on a published fit: add one physically motivated term and record which parameter moves.

Refutation. Parameters stable under physically motivated extensions refute the mechanism.

Limits. The test requires a physically motivated extension, which is exactly what may be missing. A parameter that is stable under every extension the community has tried is evidence for robustness, not proof of it.

Part III — Three hundred hypotheses

Every entry below is a proposition, not a report of a measurement. No experiment described here has been performed.

P01 — The quadratic-Zeeman correction does not extrapolate because it is an average, not a point value

`physics.atom-ph` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Long-baseline atom interferometers treat the quadratic Zeeman phase as a term proportional to B^2 at the cloud centre and remove it by extrapolating the measured phase against bias field.
- **Mechanism.** The interferometer samples $\int B(z)^2\, w(z)\, dz$ along the arm-separation trajectory, with $w(z)$ set by the atomic wavefunction and the pulse sequence, not B^2 at the centre. A field curvature term adds a phase that is not removed by bias-field extrapolation because it does not scale with the applied bias.
- **Discriminating observation.** The residual phase after extrapolation should scale with a deliberately applied field *curvature* (quadrupole trim) at fixed B, and should change sign with the direction of launch.
- **Test.** Existing dual-species apparatus; scan quadrupole trim at fixed bias field, compare ^{85}Rb and ^{87}Rb residuals whose $w(z)$ differ.
- **Prior art.** Hyperfine-ground-state-exchange evaluation of the quadratic Zeeman term (arXiv:2404.01199).
- **Refuted if.** The extrapolated residual is independent of applied curvature at the 1e-9 rad level.

P02 — Detector calibration measured at high actuation amplitude biases every loud event

`physics.ins-det` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Interferometric gravitational-wave detectors calibrate the actuation function by injecting swept sine or broadband lines and assume the response is linear in drive amplitude.
- **Mechanism.** Electrostatic and coil actuators have amplitude-dependent stiffness and eddy-current lag; a calibration derived at large amplitude differs from the small-amplitude response that encodes strain. The bias scales with the ratio of calibration drive to the actual test-mass motion.
- **Discriminating observation.** Reconstructed strain amplitude for a loud event should shift by a measurable fraction when the calibration is repeated at several actuation amplitudes; the shift should be quadratic in the calibration amplitude.
- **Test.** Reanalyse public calibration-line data across amplitude settings in the same run.
- **Prior art.** Detector-calibration uncertainty budgets in ringdown and spectroscopy analyses (arXiv:2610.11052, arXiv:2506.15979).
- **Refuted if.** The reconstructed strain is amplitude-independent to better than 1e-3 across a decade of actuation.

P03 — Dispersion-measure variations absorb scattering and manufacture red noise

`astro-ph.HE` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Pulsar timing separates dispersion measure from achromatic timing residuals by assuming the cold-plasma ν^{-2} law across the observing bands.
- **Mechanism.** Multiphase-broadening and scattering delays are chromatic but not exactly ν^{-2}; fitting a single DM to two bands moves part of the scattering delay into DM. The residual is temporally correlated with the scattering timescale, producing apparent red noise at the nanohertz level.
- **Discriminating observation.** DM time series should correlate with independently measured scintillation bandwidth, with a sign fixed by the sign of $(\alpha_\mathrm{scatt} - 2)$.
- **Test.** Existing pulsar timing archives with simultaneous scintillation measurements.
- **Prior art.** Chromatic noise treatments in pulsar timing arrays.
- **Refuted if.** No correlation between DM residuals and scattering bandwidth at $|r| > 0.3$ across the array.

P04 — Planet-based beam calibration is biased by the finite planet disc in a frequency-dependent way

`astro-ph.IM` · `boundary-condition` · readiness 3 · novelty 2

- **Baseline.** CMB experiments calibrate beam solid angle and window functions on planets treated as point sources.
- **Mechanism.** At frequencies where the planet's angular size is comparable to the beam, the convolution attenuates high-ℓ beam modes but not the total solid angle, so the same calibration produces a frequency-dependent beam *shape* error while leaving the peak correct.
- **Discriminating observation.** The ratio of planet-derived to dipole-derived beam window functions should drift with frequency in a way that tracks planet disc size.
- **Test.** Existing planet and dipole calibration products from a single instrument.
- **Prior art.** Standard beam-calibration systematics in CMB analysis.
- **Refuted if.** Planet- and dipole-derived window functions agree within errors at all bands after point-source correction.

P05 — Cepheid crowding corrections carry a distance-independent bias that mimics a Hubble-tension shift

`astro-ph.CO` · `selection` · readiness 3 · novelty 3

- **Baseline.** The distance ladder corrects Cepheid photometry for blending using image-level crowding metrics calibrated on artificial-star tests.
- **Mechanism.** Artificial-star tests inject sources at the detection threshold, so the recovered correction is conditioned on detectability; real blends are correlated with the Cepheid itself, which the injection does not reproduce. The bias is larger in the more crowded host galaxies and therefore correlates with distance.
- **Discriminating observation.** The residual Hubble-diagram scatter should correlate with the local surface brightness at the Cepheid, and the inferred Ho should shift when the most crowded quartile is removed.
- **Test.** Public Cepheid photometry and crowding maps.
- **Prior art.** Selection-bias corrections for time-delay cosmography and distance ladders (arXiv:2410.16171).
- **Refuted if.** Ho is stable to $<1\%$ under crowding-quartile resampling.

P06 — Zero-noise extrapolation fails when the noise is non-Markovian, and the failure is measurable

`quant-ph` · `assumption` · readiness 4 · novelty 3

- **Baseline.** Richardson-style zero-noise extrapolation assumes the expected value is an analytic function of the noise scale with a leading linear term.
- **Mechanism.** With 1/f or otherwise correlated noise, the leading correction is non-analytic in the scale factor (a fractional power), so Richardson coefficients amplify rather than cancel the error, and the extrapolated value moves *away* from the true one as more scale points are added.
- **Discriminating observation.** The extrapolated estimate should be non-monotonic in the number of scale factors for hardware with measured 1/f spectra, but monotonic for white-noise-dominated hardware.
- **Test.** Reanalyse published extrapolation datasets paired with device noise spectra.
- **Prior art.** Finite-shot MSE bounds for Richardson extrapolation (arXiv:2605.08251).
- **Refuted if.** Extrapolation error decreases monotonically with added scale factors on devices whose spectra have 1/f slope > 0.5.

P07 — Randomized benchmarking is blind to drift because it assumes gate-independent error

`quant-ph` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Randomized benchmarking (RB) fits an exponential decay and reports a per-gate error that is interpreted as an average gate fidelity.

- **Mechanism.** If the error rate drifts during a sequence (T_1 fluctuation, TLS hopping), the decay remains close to exponential but with a rate equal to the *time-averaged* error weighted by sequence length; because long sequences weight late times, the fitted rate is biased toward the error at the end of the experiment.
- **Discriminating observation.** RB decay rates measured in forward vs reversed time order (or with interleaved idle periods) should differ by an amount tracking the measured drift.
- **Test.** Existing RB datasets with timestamped runs on the same device.
- **Prior art.** Standard RB theory; drift-aware tomographic protocols.
- **Refuted if.** Forward and reversed RB rates agree within statistical error while independent drift monitors show >10% variation.

P08 — NV thermometry has a strain–temperature degeneracy fixed by the crystal cut

`cond-mat.mes-hall` · `degeneracy` · readiness 4 · novelty 3

- **Baseline.** NV-centre thermometry converts the zero-field splitting shift into temperature using a calibrated dD/dT.
- **Mechanism.** Axial strain and temperature both shift D; the two contributions are collinear for a [100]-cut ensemble but not for [111]. A single-axis measurement therefore reports a temperature error proportional to the local strain, and the error is invisible because strain is not independently measured.
- **Discriminating observation.** Temperature readings from [100] and [111] ensembles on the same sample should disagree under applied stress and agree under stress-free conditions.
- **Test.** Dual-orientation diamond with a piezoelectric strain stage.
- **Prior art.** Pre-registered comparison of 23 NV thermometry methods reporting geometric blind spots (Zenodo record 21299294).
- **Refuted if.** Both cuts track the same temperature within 0.1 K under 1e-4 strain.

P09 — Four-bar conductivity measurements have an unbounded error from finite contact pads

`cond-mat.mtrl-sci` · `boundary-condition` · readiness 5 · novelty 2

- **Baseline.** Four-bar geometries are treated as removing contact resistance exactly.
- **Mechanism.** When the metal pads have finite conductivity and the film is thin, current spreads into the pads, so the measured voltage drop includes a pad-mediated component that grows without bound as film conductivity approaches pad conductivity.
- **Discriminating observation.** Extracted film conductivity should depend on pad metal and pad geometry at fixed film, with the discrepancy scaling as the conductivity ratio.
- **Test.** Reanalyse four-bar datasets with varying pad metallurgy.
- **Prior art.** arXiv:2602.02418 demonstrates unbounded systematic error in this architecture.
- **Refuted if.** Extracted conductivity is invariant to pad material across a 100× conductivity range.

P10 — Single-photon detector efficiency calibrations inherit a bias from afterpulsing that depends on count rate

`physics.ins-det` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Detector efficiency is calibrated at low count rate and assumed rate-independent.
- **Mechanism.** Afterpulsing and dead-time losses are rate-dependent; a calibration performed at 10 kcps applied to a 10 Mcps measurement overestimates efficiency by an amount set by the afterpulse probability and the hold-off time.
- **Discriminating observation.** The ratio of calibrated to inferred efficiency should scale linearly with count rate with a slope equal to the afterpulse probability.
- **Test.** Existing detector characterisation data with rate scans.
- **Prior art.** Standard afterpulse/dead-time metrology.
- **Refuted if.** Efficiency is flat in rate to 1e-3 up to the maximum operating rate.

P11 — Two accepted measurements of the neutron lifetime differ because one of them measures a surface, not a volume

`nucl-ex` · `model-disagreement` · readiness 3 · novelty 3

- **Baseline.** Beam and bottle neutron-lifetime measurements disagree at the 4σ level and are treated as a single-parameter discrepancy.
- **Mechanism.** Beam experiments count protons from a defined volume but the proton extraction efficiency depends on the wall/foil condition; bottle experiments count surviving neutrons but the loss term depends on surface chemistry. The two methods therefore measure different combinations of the same two parameters (τ and wall loss), and their disagreement is a signature of a third, unmeasured surface parameter.
- **Discriminating observation.** The beam/bottle discrepancy should correlate with the beam-tube surface history (coat time, vacuum exposure) even when the beam flux is constant.
- **Test.** Retrospective correlation of published beam results with documented surface-condition metadata.
- **Prior art.** The beam–bottle anomaly literature.
- **Refuted if.** Beam results are uncorrelated with surface-condition metadata while showing the full discrepancy.

P12 — Exoplanet radius valley positions are biased by the limb-darkening law used in transit fits

`astro-ph.EP` · `assumption` · readiness 4 · novelty 3

- **Baseline.** Transit fitting uses a fixed limb-darkening law with coefficients interpolated from stellar models.
- **Mechanism.** Limb darkening correlates with impact parameter and hence with radius; an incorrect coefficient biases radius in a direction that depends on b. Because the radius valley is defined by a population ridge, a b-correlated bias shears the ridge rather than shifting it, and the shear appears as a genuine slope in the valley.
- **Discriminating observation.** The valley slope should correlate with the survey's impact-parameter distribution, and re-fitting with free limb-darkening coefficients should remove the slope.
- **Test.** Public transit photometry with free vs fixed limb darkening.
- **Prior art.** Limb-darkening systematics in transit spectroscopy.
- **Refuted if.** The valley slope is unchanged when limb-darkening coefficients are free within physical priors.

P13 — Supernova standardisation coefficients are window-dependent because colour and stretch are not orthogonal in the observed frame

`astro-ph.CO` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** SNe Ia distances use a linear standardisation with constant coefficients α, β fitted globally.
- **Mechanism.** In the observer frame, redshift stretches the filter bandpasses, so the effective colour–stretch covariance changes with z. A globally fitted β therefore absorbs a redshift-dependent piece, producing a distance bias that grows with z and mimics a cosmological parameter shift.
- **Discriminating observation.** β fitted in redshift bins should drift monotonically, and the drift should vanish when standardisation is performed in the rest frame with bandpass corrections.
- **Test.** Public SN samples with per-object filter transmission.
- **Prior art.** Standardisation systematics in SN cosmology.
- **Refuted if.** Binned β is constant within errors and rest-frame refitting does not change the Hubble diagram residuals.

P14 — The X-ray mass proxy is biased by clumping in a way that is degenerate with feedback

`astro-ph.CO` · `assumption` · readiness 3 · novelty 3

- **Baseline.** Hydrostatic mass estimates assume the gas is smooth and in hydrostatic equilibrium, and the resulting bias is parameterised as a single factor.
- **Mechanism.** The bias factor depends on the ratio of clumping to thermal pressure support, which varies with radius and mass; a single factor therefore cannot absorb it, and the residual masquerades as a mass-dependent feedback efficiency.

- **Discriminating observation.** The required mass correction should correlate with measured gas clumping (from X-ray surface-brightness substructure) at fixed hydrostatic mass.
- **Test.** Existing X-ray plus lensing cluster samples.
- **Prior art.** Hydrostatic mass bias and σ_8 tension forecasts (arXiv:2602.19478).
- **Refuted if.** Residual mass bias is uncorrelated with substructure measures.

P15 — Line-of-sight velocity dispersion estimates are biased by binary orbits in a way that depends on the observing window

`astro-ph.GA` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Cluster velocity dispersions are computed from single-epoch radial velocities with a binary correction calibrated on simulations.
- **Mechanism.** Binary broadening scales with orbital period; a survey whose cadence is shorter than the period sees a coherent velocity offset rather than broadening, so the correction over-subtracts and the dispersion is underestimated for compact clusters observed in one season.
- **Discriminating observation.** The inferred dispersion should differ between single-season and multi-season data for the same cluster, with the sign fixed by the cadence/period ratio.
- **Test.** Existing multi-epoch surveys.
- **Prior art.** Binary corrections in stellar kinematics.
- **Refuted if.** Single- and multi-season dispersions agree within errors at fixed cluster mass.

P16 — Quantum-advantage demonstrations are compared against a classical baseline that is not optimised to the same error model

`quant-ph` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Sampling experiments claim advantage by comparing against tensor-network or MPS classical simulations at a fixed bond dimension.
- **Mechanism.** The classical baseline is chosen to reproduce the *state*, while the quantum device's output is degraded by readout and crosstalk noise. At matched output fidelity the classical cost is lower than reported, and the gap closes as device noise grows.
- **Discriminating observation.** Recomputing the classical cost at the fidelity actually achieved (not the ideal target) should shrink the reported advantage, with the shrinkage scaling as the device error rate.
- **Test.** Reanalysis of published advantage datasets with fidelity-matched baselines.
- **Prior art.** Classical-simulation benchmarking literature.
- **Refuted if.** Advantage remains at matched fidelity across a decade of device error rates.

P17 — Maser clocks and optical clocks disagree through a temperature coefficient that is attributed to the wrong component

`physics.atom-ph` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Frequency comparisons between fountain/optical standards assign a temperature coefficient to the cavity or the vacuum window.
- **Mechanism.** The cavity's thermal expansion and the interrogation laser's fibre-path length have temperature coefficients of opposite sign; a single lumped coefficient fitted at one temperature cannot separate them, and the misassignment causes an error that grows with the temperature excursion amplitude.
- **Discriminating observation.** The temperature sensitivity should depend on the *phase* of the thermal cycle relative to the interrogation sequence, not just on temperature.
- **Test.** Existing clock comparison datasets with cavity and fibre temperature telemetry.
- **Prior art.** Standard frequency-shift budgets.
- **Refuted if.** Sensitivity is phase-independent at the 1e-16 level.

P18 — Interferometric null tests pass while the systematic is common-mode

`physics.optics` · `blind-spot` · readiness 4 · novelty 3

- **Baseline.** Instrumental systematics are bounded by null tests that compare signal channels with expected-zero combinations.
- **Mechanism.** A systematic that is common to both arms of a null combination cancels in the null but does not cancel in the signal, so null tests bound only the differential component. The common-mode part is unconstrained and can be larger than the quoted systematic.
- **Discriminating observation.** Injecting a known common-mode signal into the hardware model should leave all null tests unchanged while shifting the science signal.
- **Test.** Hardware injection campaigns with symmetric actuation.
- **Prior art.** Null-test methodology in cosmology and interferometry.
- **Refuted if.** Common-mode injections move at least one null test at the quoted sensitivity.

P19 — The cosmic-ray energy scale is set by a calorimeter that is not calorimetric at the highest energies

`astro-ph.HE` · `assumption` · readiness 3 · novelty 3

- **Baseline.** Ultra-high-energy cosmic-ray energies are reconstructed from fluorescence yield or from surface-detector signal calibrated against simulated showers.
- **Mechanism.** At the highest energies the shower maximum moves above the detector level; the "calorimetric" fraction then depends on the unmeasured muon component, so the reconstructed energy carries an air-shower-model-dependent bias that is degenerate with the flux suppression.
- **Discriminating observation.** The composition inferred from X_max should correlate with the energy-scale residual at fixed flux, and the correlation should differ between hadronic models.
- **Test.** Existing hybrid datasets with X_max and surface signals.
- **Prior art.** Energy-scale and composition systematics in UHECR.
- **Refuted if.** Energy-scale residuals are uncorrelated with X_max composition proxies.

P20 — Optical-clock comparisons inherit an unmodelled self-gravity term that depends on the apparatus mass, not on the site

`physics.atom-ph` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Clock comparisons correct for the gravitational potential using a geoid model and the clock's height, treating the apparatus as massless.
- **Mechanism.** The clock's own vacuum chamber and magnetic shield displace several hundred kilograms within a metre of the atoms; their potential shifts the clock frequency by a few parts in 10^18. Because the correction is computed from the site's geodetic model, a transportable clock moved between laboratories acquires a site-dependent offset that tracks its own mass distribution.
- **Discriminating observation.** Two clocks at the same height and site but with different chamber masses should differ by an amount proportional to the mass difference, and the difference should reverse when one clock's ballast is changed.
- **Test.** Comparison of a transportable optical clock against a stationary one with deliberate ballast variation.
- **Prior art.** Relativistic geodesy and clock-comparison budgets; self-gravity in fountain clocks.
- **Refuted if.** The comparison offset is independent of apparatus mass at the 1e-18 level.

P21 — Matter-wave interferometry's contrast loss is attributed to decoherence but is partly a classical alignment drift

`physics.atom-ph` · `time-dependent` · readiness 3 · novelty 3

- **Baseline.** Loss of fringe contrast is interpreted as decoherence and used to bound anomalous coupling.

- **Mechanism.** Slow thermal drift of the retro-reflector changes the fringe phase across the cloud; the ensemble average reduces contrast without any loss of coherence. The decay is therefore a function of integration time, not of separation.
- **Discriminating observation.** Contrast should recover when the cloud is split into addressable sub-regions and each is fitted separately, with the recovered phase gradient matching the measured reflector drift.
- **Test.** Existing imaging data with sub-region analysis.
- **Prior art.** Standard contrast-decay analyses.
- **Refuted if.** Sub-region fits show no phase gradient while contrast decay remains.

P22 — Bolometer time constants measured in-situ are biased by the optical load they are meant to characterise

`astro-ph.IM` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Detector time constants are measured with a calibration source and extrapolated to sky load.
- **Mechanism.** Transition-edge sensor response is electrothermal; the effective time constant depends on the bias power and hence on the optical load. A constant measured at high calibration load is shorter than the sky-load value, biasing the beam-map deconvolution for fast scans.
- **Discriminating observation.** Time constants should vary with controlled load (aperture cover) with a slope predictable from the electrothermal model, and map transfer functions should depend on scan speed accordingly.
- **Test.** Existing load-variation campaigns.
- **Prior art.** Electrothermal TES modelling.
- **Refuted if.** Time constant is load-independent within 1%.

P23 — Aperture-synthesis closure phases are immune to some systematics and therefore blind to them

`astro-ph.IM` · `blind-spot` · readiness 3 · novelty 3

- **Baseline.** Closure phase is used as a calibration-free observable because station-based phase errors cancel.
- **Mechanism.** Any error that is common to all stations in a triangle (e.g. a global reference-frequency offset, or a correlated atmosphere) also cancels, so closure quantities cannot bound it while the visibility amplitudes can. The unconstrained common mode enters image reconstruction as a slowly varying background.
- **Discriminating observation.** Reconstructions from closure-only and from amplitude-calibrated data should differ by a smooth background whose spatial scale matches the array's shortest baseline.
- **Test.** Existing interferometric datasets reduced both ways.
- **Prior art.** Closure-phase methodology.
- **Refuted if.** Closure-only and amplitude-calibrated images agree on scales larger than the maximum baseline.

P24 — The neutrino mass ordering significance depends on the reactor flux model used to normalise, not on the data

`hep-ex` · `model-disagreement` · readiness 3 · novelty 3

- **Baseline.** Long-baseline and reactor analyses combine a far/near ratio with a predicted flux.
- **Mechanism.** The flux prediction enters multiplicatively and is degenerate with the overall normalisation, which in turn couples to the matter-effect term that distinguishes orderings. A 1% flux-shape difference at the first oscillation maximum can shift the ordering significance by more than the quoted systematic because it is fitted as a normalisation rather than a shape.
- **Discriminating observation.** The ordering $\Delta\chi^2$ should vary across published flux models by more than the stated systematic, and the variation should track the flux shape at the first maximum.
- **Test.** Reanalysis with alternative flux models.
- **Prior art.** Reactor flux-model comparisons.
- **Refuted if.** $\Delta\chi^2$ is stable to <0.5 across models.

P25 — Black-hole ringdown overtones are contaminated by the merger's nonlinear phase, and the contamination looks like a higher-overtone amplitude

`gr-qc` · `assumption` · readiness 3 · novelty 3

- **Baseline.** Ringdown spectroscopy fits a sum of quasinormal modes starting at a fixed time after the peak, assuming the start time is early enough for linear theory.
- **Mechanism.** The nonlinear merger remnant contributes a power-law-in-time component with the same frequency as the fundamental; fitting a sum of exponentials to a power law + exponential mixture assigns the mismatch to the first overtone, biasing its amplitude and the inferred remnant parameters.
- **Discriminating observation.** Overtone amplitudes should depend on start time in a way that follows a power law rather than the predicted exponential, and the dependence should be stronger for louder events.
- **Test.** Reanalysis of public ringdown data with a power-law component included.
- **Prior art.** Ringdown calibration-uncertainty studies (arXiv:2610.11052).
- **Refuted if.** Overtone amplitudes are start-time independent once calibration uncertainty is marginalised.

P26 — Quantum-noise squeezing levels are quoted at a photodiode that does not see the mode the interferometer uses

`physics.optics` · `calibration-degeneracy` · readiness 4 · novelty 3

- **Baseline.** Squeezing is characterised by homodyne detection at a pick-off and the number is transferred to the interferometer's sensitivity.
- **Mechanism.** The pick-off has different mode-matching and loss than the main path, so the detected quadrature variance is a different combination of the same two quadratures. The quoted squeezing is therefore a projection, not the in-loop value.
- **Discriminating observation.** In-loop squeezing inferred from sensitivity should differ from pick-off squeezing by a factor predictable from the measured mode-matching, and the discrepancy should scale with the mismatch.
- **Test.** Existing squeezing benchmarks with mode-matching telemetry.
- **Prior art.** Squeezing calibration methodology.
- **Refuted if.** Pick-off and in-loop squeezing agree within 0.1 dB across mode-matching values.

P27 — The solar abundance scale is set by a 3D/non-LTE correction that is calibrated on one line, then applied to all

`astro-ph.SR` · `model-disagreement` · readiness 3 · novelty 3

- **Baseline.** 1D LTE abundances are corrected to 3D non-LTE using a correction computed for a reference line and interpolated in Teff/log g.
- **Mechanism.** The correction depends strongly on line strength and excitation potential; interpolating from a weak reference line to strong lines over-corrects, producing an abundance trend with equivalent width that is currently absorbed into microturbulence.
- **Discriminating observation.** After correction, abundance should be independent of equivalent width; if a residual slope remains that correlates with line strength, the correction is mis-specified.
- **Test.** Existing high-resolution stellar spectra with many lines per star.
- **Prior art.** 3D non-LTE abundance corrections; survey abundance scale work (arXiv:2501.06140).
- **Refuted if.** Corrected abundances show no residual slope with reduced width beyond the quoted errors.

P28 — Cosmic shear shear-calibration bias is a function of galaxy size, and size correlates with redshift

`astro-ph.CO` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Shear multiplicative bias is calibrated as a scalar per tomographic bin.

- **Mechanism.** The bias depends on the ratio of galaxy radius to PSF radius; because angular size scales with distance, the ratio is redshift-dependent within a bin, so a scalar correction leaves a redshift-dependent residual that mimics a growth-rate modification.
- **Discriminating observation.** Residual shear bias should correlate with galaxy size at fixed redshift, and the recovered S_8 should shift when size is used as a covariate.
- **Test.** Existing image simulations and survey catalogues.
- **Prior art.** Shear calibration methodology.
- **Refuted if.** Residual bias is uncorrelated with size after scalar correction.

P29 — Dark-siren cosmology is limited by the lensing-magnification prior, not by the catalogue size

`gr-qc` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Bright sirens give H_0 from the GW distance and the host redshift, with lensing treated as a symmetric scatter.
- **Mechanism.** Magnification is not symmetric: the observed luminosity distance distribution is skewed by demagnification of unlensed lines of sight. Assuming symmetry biases the distance at second order, and the bias grows with the number of events because it does not average out.
- **Discriminating observation.** H_0 inferred from bright sirens should drift with the maximum distance of the sample in a direction fixed by the skewness.
- **Test.** Catalogue-level reanalysis with an asymmetric magnification prior.
- **Prior art.** Systematics costs in dark-siren cosmology (arXiv:2602.14898).
- **Refuted if.** H_0 is unchanged under asymmetric magnification priors.

P30 — The Tully–Fisher zero point and the Cepheid zero point disagree because of a common saturation term

`astro-ph.GA` · `model-disagreement` · readiness 3 · novelty 3

- **Baseline.** Distance indicators are cross-calibrated in overlapping anchors and the residual scatter is attributed to intrinsic scatter.
- **Mechanism.** Both indicators rely on photometry near the bright end of the detector's linear range; a shared nonlinearity near saturation biases bright calibrators low and faint calibrators high, producing an apparent zero-point offset that depends on the magnitude distribution of each indicator.
- **Discriminating observation.** The offset should correlate with the fraction of near-saturation pixels in the calibration images, and should vanish when only unsaturated calibrators are used.
- **Test.** Reanalysis with pixel-level flags.
- **Prior art.** Detector-linearity systematics in photometry.
- **Refuted if.** Zero-point offsets are unchanged when saturated calibrators are removed.

P31 — Finite simulation box size biases the inferred dark-matter annihilation boost in a direction that depends on the observable

`astro-ph.CO` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Subhalo boost factors are measured in simulations and extrapolated below the resolution limit with a power-law index.
- **Mechanism.** Missing large-scale modes suppress the low-mass subhalo abundance; because the suppression depends on the box size and on the power-law index, the extrapolated boost is a function of box size, and different observables (annihilation $\propto \rho^2$, lensing $\propto \rho$) weight the missing modes differently.
- **Discriminating observation.** The boost measured in boxes of different size should converge only when the missing-mode fraction is below the quoted error, and the convergence rate should differ between ρ and ρ^2 weightings.
- **Test.** Existing simulation suites with varying box size at fixed resolution.
- **Prior art.** Convergence studies in structure formation.

- **Refuted if.** ρ^2 boost is box-size independent at fixed resolution.

P32 — Vacuum-systematic bounds in atom interferometry are limited by an unmeasured pressure gradient

`physics.atom-ph` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Residual gas pressure is measured at one gauge and the associated phase shift is computed from that number.
- **Mechanism.** The gauge sits outside the interferometer chamber; the local pressure in the cloud region is set by a conductance ratio that depends on temperature and on the getter state. A uniform pressure assumption therefore underestimates the shift by a factor that drifts as the getter saturates.
- **Discriminating observation.** The inferred background phase should drift monotonically with getter age at fixed gauge reading.
- **Test.** Existing apparatus logs with getter history.
- **Prior art.** Vacuum systematics in atom interferometry.
- **Refuted if.** Background phase is uncorrelated with getter age.

P33 — Superconducting qubit coherence is limited by a thermal population that the thermometer cannot see below 50 mK

`quant-ph` · `blind-spot` · readiness 3 · novelty 3

- **Baseline.** Qubit excited-state population is attributed to the mixing-chamber temperature, and residual population is treated as a fixed offset.
- **Mechanism.** The qubit mode couples to a lossy element whose local density of states is hot at the qubit frequency; the effective temperature is set by the noise spectral density at ω_q, not by the phonon bath at the mixing chamber. A thermometer at the mixing chamber therefore cannot observe the relevant quantity.
- **Discriminating observation.** Qubit population should be independent of mixing-chamber temperature but should track a separately measured noise spectral density at ω_q (e.g. from a nearby resonator).
- **Test.** Existing devices with both qubit and resonator noise measurements.
- **Prior art.** Qubit quasiparticle and TLS loss literature.
- **Refuted if.** Qubit population tracks mixing-chamber temperature while resonator noise at ω_q is constant.

P34 — The local Hubble flow is measured with a zero point that depends on the same peculiar-velocity model it is used to test

`astro-ph.CO` · `circularity` · readiness 3 · novelty 3

- **Baseline.** H_0 from the local flow uses distances calibrated in a sample whose selection is defined by the same velocity field.
- **Mechanism.** The sample selection (e.g. minimum distance or velocity dispersion cut) is correlated with the peculiar velocity used to correct the distances; the correction and the selection therefore partially cancel, flattening the inferred flow and biasing H_0 low.
- **Discriminating observation.** H_0 should vary with the selection threshold in a way predicted by the cancellation, and should stabilise when selection is made in a velocity-independent way.
- **Test.** Reanalysis with alternative selections.
- **Prior art.** Cosmic-void and local-flow H_0 analyses (arXiv:2607.20257).
- **Refuted if.** H_0 is selection-threshold independent.

P35 — Radiometric calibration of a spectrometer's slit function hides a wavelength-dependent bias in retrieved column densities

`physics.optics` · `assumption` · readiness 4 · novelty 2

- **Baseline.** Retrievals assume the instrument slit function is Gaussian with a width measured from a narrow-line lamp.

- **Mechanism.** Real slit functions have asymmetric wings; for optically thick lines the wings sample the line core, so the retrieved column is a nonlinear functional of the wing fraction and cannot be absorbed by a width parameter.
- **Discriminating observation.** Retrievals from the same scene at two instrument temperatures (which change the wing fraction) should differ, with the difference scaling with line optical depth.
- **Test.** Existing retrievals with instrument-temperature telemetry.
- **Prior art.** Slit-function systematics in trace-gas retrievals.
- **Refuted if.** Retrieved columns are temperature-independent to within the stated error.

P36 — The proton radius puzzle's remaining discrepancy is a fit-range artifact in the extraction of the charge radius from form factors

`nucl-ex` · `model-disagreement` · readiness 3 · novelty 3

- **Baseline.** The radius is extracted by fitting a polynomial in Q^2 to form-factor data and taking the slope at $Q^2=0$.
- **Mechanism.** The fit range extends beyond the region where the polynomial is valid; the curvature term absorbs part of the slope, so the extracted radius depends on Q^2_max. Because electron and muon datasets have different Q^2 coverage, they are biased differently.
- **Discriminating observation.** The extracted radius should drift with Q^2_max, and the drift should differ between the two datasets in the direction of their coverage difference.
- **Test.** Reanalysis of published form-factor data with a controlled fit range.
- **Prior art.** Form-factor extraction methodology.
- **Refuted if.** Radius is stable to <0.005 fm across Q^2_max from 0.02 to 0.2 GeV^2.

P37 — Scintillator quenching corrections are calibrated in a beam that does not reproduce the dE/dx spectrum of the application

`physics.ins-det` · `extrapolation` · readiness 3 · novelty 2

- **Baseline.** Birks' constant is fitted to beam data at a fixed particle type and applied to mixed fields.
- **Mechanism.** Quenching depends on the differential energy-loss spectrum, not on the total dE/dx; a mixed field with the same mean dE/dx but a different spectrum has a different light yield. The correction is therefore a function of the spectral hardness, which is not measured.
- **Discriminating observation.** Light yield per unit deposited energy should differ between fields with matched mean dE/dx but different spectra (e.g. protons vs electrons at matched stopping power).
- **Test.** Existing dual-particle calibration data.
- **Prior art.** Birks' law and its limits.
- **Refuted if.** Light yield at matched mean dE/dx agrees within 1% across particle types.

P38 — The neutron electric-dipole-moment systematic budget omits a motional-field term that scales with the dressing-field gradient

`physics.atom-ph` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** The nEDM extraction assumes the motional magnetic field ($v\times E/c^2$) is suppressed by the dressing field and bounded by a measured gradient.
- **Mechanism.** The suppression depends on the Larmor frequency relative to the dressing-field modulation; at the operating point the residual is second order in the gradient but first order in its *curvature*, which is not measured. A curved gradient therefore produces a false EDM with a signature that is odd in the dressing-field direction.
- **Discriminating observation.** The false EDM should reverse sign with dressing-field direction and should scale with the measured gradient curvature rather than its magnitude.
- **Test.** Existing apparatus with gradient-map data.
- **Prior art.** nEDM systematics literature.
- **Refuted if.** The false EDM is uncorrelated with gradient curvature.

P39 — The extragalactic background light is bounded by a gamma-ray opacity that is degenerate with intergalactic magnetic-field heating

`astro-ph.HE` · `degeneracy` · readiness 3 · novelty 3

- **Baseline.** EBL density is inferred from the attenuation of TeV spectra, assuming the cascade energy is lost to the intergalactic medium.
- **Mechanism.** If the intergalactic field is strong enough to deflect cascade electrons into a halo, the cascade emission appears as an extended halo rather than as a point-source flux; the point-source-only analysis then attributes the missing flux to EBL absorption, overestimating EBL where halos are unresolved.
- **Discriminating observation.** EBL estimates from point-source fits should correlate with the resolved-halo fraction of the sample, and should decrease when halos are included.
- **Test.** Existing TeV spectra with halo measurements.
- **Prior art.** Cascade-halo and EBL literature.
- **Refuted if.** EBL estimates are unchanged by halo inclusion.

P40 — Stellar rotation periods from spot modulation alias to half the true period when the spot lifetime exceeds the observing baseline

`astro-ph.SR` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Periods are found by autocorrelation or Lomb–Scargle peak of the light curve.
- **Mechanism.** Two active longitudes separated by ∼180° produce a light curve with two minima per rotation; if the spot pattern evolves on a timescale comparable to the baseline, the second harmonic is suppressed and the fundamental dominates, but if it does not, the harmonic dominates and the period is halved. The transition is a function of baseline, not of rotation.
- **Discriminating observation.** The retrieved period should change by a factor of two as the baseline crosses the spot-evolution timescale, at fixed star.
- **Test.** Existing long-baseline photometry.
- **Prior art.** Spot-modulation period-finding.
- **Refuted if.** Periods are baseline-independent for stars with measured spot evolution.

P41 — Fast-radio-burst fluence distributions are truncated by the detection pipeline's dedispersion incoherence

`astro-ph.HE` · `selection` · readiness 4 · novelty 3

- **Baseline.** The intrinsic fluence distribution is inferred from detected bursts after correcting for beam and threshold.
- **Mechanism.** Dedispersion assumes a single DM per burst; intra-channel smearing grows with DM and width, so high-DM, broad bursts lose signal-to-noise in a way that mimics a steeper intrinsic distribution at the faint end. The inferred index therefore depends on the survey's DM distribution.
- **Discriminating observation.** The faint-end index should correlate with the survey's median DM, and should flatten when re-analysed with sub-band dedispersion.
- **Test.** Reanalysis of public burst catalogues with sub-band data.
- **Prior art.** FRB population statistics.
- **Refuted if.** The faint-end index is DM-independent.

P42 — Interferometric imaging's flux-scale error is set by the shortest baseline, not by the longest

`astro-ph.IM` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Flux-scale accuracy is attributed to long-baseline calibration and to the primary beam.
- **Mechanism.** Total flux is carried by the shortest baselines, where the antenna primary beam and the wide-field w-term are largest; a fractional error in the shortest-baseline gain moves total flux without changing the image shape, and it is not constrained by closure quantities.

- **Discriminating observation.** Total flux should correlate with the shortest-baseline gain solution, and should differ between imaging with and without short baselines.
- **Test.** Existing datasets imaged with and without tapering.
- **Prior art.** Wide-field interferometric calibration.
- **Refuted if.** Total flux is independent of shortest-baseline gain within errors.

P43 — Muon tomography's density inversion is biased by the assumed rock composition, and the bias is degenerate with a void

`physics.ins-det` · `degeneracy` · readiness 3 · novelty 2

- **Baseline.** Muon radiography converts transmitted flux to density using a stopping-power model with an assumed composition.
- **Mechanism.** Stopping power depends on Z/A and on the mean excitation energy; a compositional change produces the same flux deficit as a density decrease. Because both enter as a line integral, a low-Z anomaly and a void are indistinguishable without an independent constraint.
- **Discriminating observation.** The joint density–composition posterior should be a ridge whose direction is set by the stopping-power model; adding a gravity or seismic constraint should break it.
- **Test.** Existing muon tomography datasets with gravity surveys.
- **Prior art.** Muon radiography inversion.
- **Refuted if.** The posterior is compact in density alone.

P44 — The quantum Cheshire Cat and similar weak-value experiments are null tests that cannot distinguish disturbance from signal

`quant-ph` · `blind-spot` · readiness 3 · novelty 3

- **Baseline.** Weak-value protocols infer a property's location from the shift of a pointer conditioned on post-selection.
- **Mechanism.** The pointer shift is proportional to the weak value, which is a ratio of amplitudes; any protocol that adds a phase shift proportional to the same amplitude ratio reproduces the shift. Without a closure test that varies the coupling strength, the shift does not identify a spatial separation.
- **Discriminating observation.** The pointer shift should scale linearly with coupling for a genuine weak value, but saturate or change slope for a disturbance-based mimic; a coupling-strength scan distinguishes them.
- **Test.** Reanalysis of published weak-value data with coupling scans.
- **Prior art.** Weak-value amplification literature and its critiques.
- **Refuted if.** Coupling-strength scans show the predicted linear regime across two decades.

P45 — Lunar laser ranging's station bias is a function of the retroreflector array orientation, which is not modelled

`gr-qc` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** LLR residuals are modelled with station coordinates, tidal displacement, and array centre-of-mass corrections.
- **Mechanism.** The array's far-field diffraction pattern depends on its orientation relative to the station; as the Moon librates, the effective optical centre moves by a fraction of the array size. The effect is periodic at libration frequencies and is not absorbed by a constant centre-of-mass offset.
- **Discriminating observation.** Station residuals should contain power at libration harmonics with amplitudes that differ between arrays of different size.
- **Test.** Existing LLR normal-point data with libration ephemerides.
- **Prior art.** LLR systematics literature.
- **Refuted if.** Residual power at libration harmonics is below the noise floor for all arrays.

P46 — Superconducting transition-temperature measurements inherit a strain state from the mounting that is never reported

`cond-mat.supr-con` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Tc is reported as a material property with a measurement uncertainty dominated by thermometer calibration.
- **Mechanism.** Mounting (glue, varnish, clamp) imposes a biaxial strain whose sign depends on differential thermal contraction; because dTc/dstrain is large for many unconventional superconductors, the sample-to-sample scatter in Tc is dominated by mounting strain rather than by stoichiometry.
- **Discriminating observation.** Tc measured on the same crystal with two mounting media of different contraction should differ by more than the quoted error, in a direction set by the sign of dTc/dp.
- **Test.** Same-crystal, two-mount comparison.
- **Prior art.** Pressure/strain dependence of Tc.
- **Refuted if.** Tc is mounting-independent within 0.1 K for a crystal with known large dTc/dp.

P47 — The 21-cm global signal's foreground removal leaves a bias set by the beam chromaticity, not by the foreground model

`astro-ph.CO` · `model-disagreement` · readiness 3 · novelty 3

- **Baseline.** Foreground removal fits a smooth polynomial in frequency to the total sky spectrum.
- **Mechanism.** The beam is chromatic, so the *observed* foreground is not smooth even when the sky is; the polynomial absorbs the beam-induced curvature, and the residual bias appears as an absorption trough whose depth depends on the beam's spectral index rather than on the 21-cm signal.
- **Discriminating observation.** The recovered trough depth should correlate with the measured beam chromaticity across instruments, and should change when the beam model is fitted jointly.
- **Test.** Existing global-signal datasets with beam maps.
- **Prior art.** EDGES-era foreground-removal debates.
- **Refuted if.** Trough depth is uncorrelated with beam chromaticity.

P48 — The neutron-star radius inference from pulse-profile modelling is degenerate with the hotspot's beaming function

`astro-ph.HE` · `degeneracy` · readiness 3 · novelty 3

- **Baseline.** NICER-style analyses fit a blackbody-like hotspot with a fixed beaming (limb-darkening) law.
- **Mechanism.** The beaming function enters the pulse shape in the same way as the emission radius for moderate inclination; a change in beaming can be compensated by changing R and M along a ridge. The quoted radius uncertainty therefore understates the true spread unless beaming is marginalised.
- **Discriminating observation.** The posterior ridge direction should change when the beaming law is varied, and the radius credible interval should widen by more than the stated systematic.
- **Test.** Reanalysis of public NICER pulse profiles with alternative beaming.
- **Prior art.** Pulse-profile modelling systematics.
- **Refuted if.** Radius is stable to <0.3 km across physically allowed beaming laws.

P49 — Scanning-probe thermometry measures the tip–sample interface, not the sample

`cond-mat.mes-hall` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Scanning thermal microscopy maps temperature by calibrating the probe's resistance–temperature relation on a reference.
- **Mechanism.** The probe–sample thermal conductance is comparable to the sample's spreading conductance, so the measured temperature is a weighted mix; the weight depends on contact area, which changes with topography. Contrast therefore tracks topography as much as temperature.

- **Discriminating observation.** Reversing the scan direction should change contrast on slopes if topography dominates, while true temperature contrast should be direction-independent.
- **Test.** Bidirectional scans on a patterned sample with known topography.
- **Prior art.** SThM calibration literature.
- **Refuted if.** Contrast is direction-independent on all slopes.

P50 — The cosmological constant from supernova residuals is degenerate with a calibration step that is fitted per-season

`astro-ph.CO` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Photometric zero points are fitted per season and per CCD, then marginalised as nuisance parameters.
- **Mechanism.** The zero-point prior is Gaussian and centred on the calibration solution, but the true error is dominated by a common-mode atmospheric term that is correlated across the season. A Gaussian per-CCD prior therefore underestimates the correlated part, and the residual leaks into w at the level of the correlation length.
- **Discriminating observation.** w should shift when the zero-point prior is replaced by a correlated (season-wide) covariance, with the shift scaling as the correlation fraction.
- **Test.** Reanalysis with correlated calibration covariance.
- **Prior art.** Photometric-calibration systematics in SN cosmology.
- **Refuted if.** w is unchanged under correlated zero-point priors.

Chemistry and electrochemistry — 50 hypotheses

Every entry is a proposition. No experiment below has been performed.

C01 — Kramers–Kronig consistency is not a linearity test, so reported charge-transfer resistances are amplitude artefacts

`electrochem` · `nonlinearity` · readiness 5 · novelty 4

- **Baseline.** Impedance spectra are accepted when a Lin-KK residual is small, and the fitted charge-transfer resistance is reported as a material property.
- **Mechanism.** For a memoryless Butler–Volmer interface shunted by a linear double-layer capacitance, the describing function is exactly a Debye element with an amplitude-dependent conductance, so it satisfies Kramers–Kronig identically at every amplitude. The KK test therefore passes while R_ct is biased low by $x/(2I_1(x))$, $x = A/b$.
- **Discriminating observation.** R_ct fitted at 50 mV should be 11% below the value at 2 mV for a 51 mV Tafel slope, with the third-harmonic ratio I_3/I_1 tracking the bias.
- **Test.** Amplitude sweep on a reversible redox couple; the protocol and bias law are computed in investigation I3.
- **Prior art.** Kramers–Kronig validation practice; describing-function theory.
- **Refuted if.** R_ct is amplitude-independent to 1% across 2–100 mV on a known Butler–Volmer system.

C02 — The double-layer capacitance extracted from impedance includes a faradaic contribution that scales with amplitude

`electrochem` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** C_dl is taken from the high-frequency arc and assumed to be a purely electrostatic property of the interface.
- **Mechanism.** The faradaic branch is nonlinear, so part of its second-order response appears in quadrature with the fundamental and is absorbed into C_dl. The apparent capacitance therefore grows with amplitude even when the true double layer is unchanged.
- **Discriminating observation.** C_dl fitted from impedance should increase quadratically with amplitude while the capacitive current measured by potential-step or by electrochemical impedance at fixed faradaic blockade stays constant.
- **Test.** Two independent capacitance measurements on the same electrode.
- **Prior art.** Nonlinear electrochemical impedance spectroscopy.
- **Refuted if.** Impedance-derived C_dl is amplitude-independent on a system with a known nonlinear faradaic branch.

C03 — Tafel slopes are biased by uncompensated resistance in a way that mimics a change of mechanism

`electrochem` · `hidden-parameter` · readiness 5 · novelty 3

- **Baseline.** Tafel slopes are fitted to log(i) vs η after 85–90% iR compensation, and a change of slope with overpotential is read as a change of rate-determining step.
- **Mechanism.** Residual IR adds a term that is linear in i, which on a Tafel plot produces a slope that *increases* with current and saturates; the apparent slope change occurs at a current set by the residual resistance, not by a mechanistic transition.
- **Discriminating observation.** The break point should move with electrode distance and with electrolyte conductivity, while a genuine mechanism change should not.
- **Test.** Vary Luggin position and conductivity at fixed catalyst.
- **Prior art.** iR-compensation practice in electrocatalysis.
- **Refuted if.** The slope-break current is invariant to cell resistance.

C04 — Reported turnover frequencies depend on the assumed electrochemically active surface area, which is itself measured by a process that changes the surface

`electrochem` · `circularity` · readiness 4 · novelty 3

- **Baseline.** TOF is normalised by ECSA from double-layer capacitance or from underpotential deposition.
- **Mechanism.** The ECSA probe (oxide formation, hydrogen underpotential deposition) reconstructs the surface, so the area measured before catalysis differs from the area during catalysis. Normalising by the pre-catalysis area biases TOF by the reconstruction factor, which is catalyst-specific and correlates with the very activity being ranked.
- **Discriminating observation.** TOF should correlate with the magnitude of the pre/post ECSA change across catalysts, and should collapse onto a single curve when normalised by in-situ area.
- **Test.** Paired pre/post ECSA on a catalyst series.
- **Prior art.** ECSA normalisation debates.
- **Refuted if.** TOF ranking is unchanged when in-situ area is used.

C05 — Solvent isotope effects on rate constants are contaminated by a viscosity change that is not corrected

`physical-chem` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** k(H2O)/k(D2O) is interpreted as a proton-transfer kinetic isotope effect.
- **Mechanism.** D2O viscosity is 23% higher at 25 °C; for diffusion-influenced or diffusion-coupled reactions the rate falls with viscosity, so part of the isotope effect is hydrodynamic. The apparent KIE therefore exceeds the true one by the ratio of diffusion limits.
- **Discriminating observation.** Adding a neutral viscogen to H2O to match D2O viscosity should reduce k(H2O) by the same factor that is attributed to the isotope.
- **Test.** Viscosity-matched KIE series.
- **Prior art.** Diffusion-influenced kinetics.
- **Refuted if.** Viscosity-matched KIE equals the raw KIE within 5%.

C06 — Arrhenius activation energies from narrow temperature ranges are systematically inflated by a hidden curvature

`physical-chem` · `assumption` · readiness 5 · novelty 2

- **Baseline.** Ea is from the slope of ln k vs 1/T over a range, often 20–40 K.
- **Mechanism.** The true temperature dependence includes a prefactor T^n and a heat-capacity term; fitting a straight line over a narrow range biases Ea by an amount proportional to the range width and to n, and the bias has a fixed sign that depends on the sign of ΔCp‡.
- **Discriminating observation.** Ea should increase with the fitted range width in a way fixed by ΔCp‡, and should converge only beyond ~100 K of range.
- **Test.** Reanalysis of published rate series with wide ranges.
- **Prior art.** Transition-state theory with ΔCp‡.
- **Refuted if.** Ea is range-width independent across 20–100 K.

C07 — Catalytic selectivity reported at low conversion is not the selectivity of the catalyst

`chem-eng` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** Selectivity is measured at low conversion to avoid secondary reactions and reported as intrinsic.
- **Mechanism.** At low conversion the product residence time is short, so secondary consumption is suppressed; but the local concentration profile in the pellet also differs, and for reactions with a negative reaction order the low-conversion rate per site is higher. Extrapolating the selectivity to industrial conversion therefore overestimates it by a factor set by the Thiele modulus.

- **Discriminating observation.** Selectivity should fall with conversion in a way that collapses onto a single curve when plotted against the Thiele modulus, not against conversion.
- **Test.** Existing conversion-sweep datasets with pellet-size variation.
- **Prior art.** Transport–reaction coupling in catalysis.
- **Refuted if.** Selectivity–conversion curves are pellet-size independent.

C08 — The pH measured in a buffered microdroplet is not the pH that governs the reaction

`physical-chem` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Microdroplet and interface chemistry is interpreted with the bulk pH of the feed solution.
- **Mechanism.** At the air–water interface the interfacial potential and the different solvation of H^+ shift the local activity by up to several units, and the shift depends on the buffer's surface affinity. Reactions whose rate is pH-dependent then appear to have a different mechanism at the interface.
- **Discriminating observation.** The apparent rate constant at the interface should depend on buffer identity at matched bulk pH, with a pattern set by surface propensity.
- **Test.** Interface-specific spectroscopy with buffer series.
- **Prior art.** Interfacial pH and microdroplet chemistry.
- **Refuted if.** Interfacial rate is buffer-identity independent at matched bulk pH.

C09 — Standard reduction potentials from computational hydrogen electrode calculations hide a potential-dependent solvation that changes the mechanism assignment

`computational-chem` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** CHE computes ΔG at $U=0$ and shifts by neU, assuming the solvent structure is unchanged.
- **Mechanism.** At operating potential the interfacial field reorients water and changes the solvation free energy of adsorbed intermediates by 0.1–0.3 eV, which is comparable to the differences used to assign rate-determining steps. The CHE assignment therefore reflects the zero-field solvation.
- **Discriminating observation.** Including explicit field-dependent solvation should change the predicted potential-determining step for a specific class of surfaces, with the change correlating with the intermediate's dipole moment.
- **Test.** Field-dependent DFT on a published catalyst series.
- **Prior art.** Constant-potential and grand-canonical DFT.
- **Refuted if.** Potential-determining step assignments are unchanged by field-dependent solvation for all intermediates with dipole > 1 D.

C10 — Reaction calorimetry's baseline subtraction removes a heat of mixing that is not constant

`analytical-chem` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Isothermal titration calorimetry subtracts a blank titration to remove dilution heats.
- **Mechanism.** The blank is measured in the absence of the host, so it lacks the host-induced solvent reorganization; the subtracted heat therefore removes a real contribution whose magnitude grows with titrant concentration. The fitted enthalpy is biased and the bias correlates with the fitted stoichiometry.
- **Discriminating observation.** The fitted ΔH should drift with the titrant concentration in the cell at fixed host, and the drift should vanish when the blank contains host at a non-binding analogue.
- **Test.** Concentration-series ITC with analogue blanks.
- **Prior art.** ITC baseline practice.
- **Refuted if.** ΔH is titrant-concentration independent.

C11 — Chromatographic retention indices drift because the dead-time marker is retained

`analytical-chem` · `hidden-parameter` · readiness 4 · novelty 2

- **Baseline.** Retention indices are computed relative to an unretained marker's elution time.

- **Mechanism.** The marker (e.g. methane or a solvent front) has a small but nonzero retention on some stationary phases, and that retention changes with column age and water content. The dead time is therefore underestimated, which compresses all adjusted retention times by a factor that drifts with column condition.
- **Discriminating observation.** The index of a standard should drift monotonically with column conditioning while the raw retention time is stable.
- **Test.** Longitudinal QC data from a single column.
- **Prior art.** Retention-index standardisation.
- **Refuted if.** Indices are stable across column aging at fixed raw retention.

C12 — Photocatalytic quantum yields above unity are artefacts of a light-absorption measurement made on a suspension

`photochem` · `calibration` · readiness 4 · novelty 3

- **Baseline.** Apparent quantum efficiency uses the fraction of incident photons absorbed, from diffuse-reflectance or from the suspension's optical density.
- **Mechanism.** In a scattering suspension the path length is longer than the cuvette thickness, so absorbance is overestimated but so is scattering loss; the two errors do not cancel for the fraction that actually reaches the catalyst. Values near or above unity then reflect a denominator error rather than multi-exciton chemistry.
- **Discriminating observation.** Apparent QE should depend on catalyst loading and on stirring rate at fixed incident flux, in the direction predicted by the scattering correction.
- **Test.** Loading and stirring series with integrating-sphere absorption.
- **Prior art.** AQE measurement standards in photocatalysis.
- **Refuted if.** AQE is invariant to loading and stirring.

C13 — Isotope-labelling product distributions are biased by scrambling in the inlet, not in the catalyst

`catalysis` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** $^{13}C/^{18}O$ scrambling in products is attributed to surface-mediated exchange and used to infer intermediate pooling.
- **Mechanism.** At the mass-spectrometer inlet and in the transfer line, adsorbed water and hot metal surfaces catalyse exchange; the residence time in the line sets the extent. The measured scrambling is therefore a function of the plumbing, not of the catalyst.
- **Discriminating observation.** Scrambling should scale with transfer-line length and temperature at fixed catalyst and fixed conversion.
- **Test.** Line-length variation with an inert catalyst.
- **Prior art.** Isotopic transient analysis in catalysis.
- **Refuted if.** Scrambling is line-length independent.

C14 — The electrochemical CO_2 reduction faradaic efficiency is normalised to a charge that includes a parasitic corrosion current

`electrochem` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** FE = (moles product × nF)/Q_total, with Q_total from the potentiostat.
- **Mechanism.** For copper and for carbon-supported catalysts, the catalyst itself is oxidised during intervals at open circuit and during the anodic swing; the resulting charge is counted in Q_total but produces no product. FE is therefore underestimated by the corrosion fraction, which depends on the duty cycle rather than on the catalyst.
- **Discriminating observation.** FE should increase with faster switching and with lower anodic excursion, and the increase should equal the independently measured corrosion charge.
- **Test.** Duty-cycle series with continuous corrosion monitoring.
- **Prior art.** CO_2RR benchmarking protocols.

- **Refuted if.** FE is duty-cycle independent.

C15 — Battery Coulombic efficiency measured by cycling includes a self-discharge term that scales with rest time

`electrochem` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** CE = discharge/charge capacity per cycle, with occasional rest.
- **Mechanism.** Self-discharge and shuttle reactions continue during rest and during the slow portions of the protocol; the measured CE is therefore a function of the protocol's time-at-voltage distribution. Comparing materials measured with different protocols conflates chemistry with protocol.
- **Discriminating observation.** CE should decrease with inserted rest time at a rate equal to the measured self-discharge current, and should converge when compared at matched time-at-voltage.
- **Test.** Protocol series on the same cell.
- **Prior art.** CE measurement standards for Li metal.
- **Refuted if.** CE is rest-time independent to 0.1%.

C16 — Spectroscopic binding constants from titration are biased by inner-filter effects that depend on the binding itself

`analytical-chem` · `circularity` · readiness 4 · novelty 3

- **Baseline.** K_a is from a fluorescence or absorbance titration fitted with a 1:1 binding model.
- **Mechanism.** Binding changes the absorbance at the excitation wavelength, so the inner-filter correction changes as the titration proceeds; the correction is usually computed from the initial spectrum. The resulting curvature is absorbed into a nonzero cooperativity, so a non-cooperative system appears cooperative.
- **Discriminating observation.** The fitted Hill coefficient should approach 1 when the inner-filter correction is recomputed per point, and the apparent cooperativity should correlate with the absorbance change at the excitation wavelength.
- **Test.** Reanalysis of published titrations with per-point corrections.
- **Prior art.** Inner-filter corrections in fluorescence.
- **Refuted if.** Hill coefficients are unchanged by per-point correction.

C17 — The glass-transition temperature from DSC depends on the fictive temperature, which is set by the prior thermal history and rarely reported

`materials-chem` · `hysteresis` · readiness 5 · novelty 2

- **Baseline.** T_g is reported as a single number from a standard heating scan after a standard quench.
- **Mechanism.** T_g is rate- and history-dependent; the standard protocol fixes the quench but not the sample's prior aging. A sample aged below T_g shows an overshoot endotherm and an apparent T_g higher by 1–5 K, so literature scatter is partly a history artefact.
- **Discriminating observation.** T_g should increase with aging time below T_g and collapse onto one curve when plotted against fictive temperature measured by the overshoot area.
- **Test.** Aging-series DSC.
- **Prior art.** Structural recovery in glasses.
- **Refuted if.** T_g is aging-time independent at fixed scan rate.

C18 — Electrochemical impedance's low-frequency tail is drift, not diffusion

`electrochem` · `time-dependent` · readiness 5 · novelty 3

- **Baseline.** A low-frequency capacitive tail or a ~45° line is interpreted as diffusion or adsorption.
- **Mechanism.** Slow drift of the open-circuit potential over the minutes-to-hours of a low-frequency sweep produces a current ramp whose demodulated quadrature component is $\propto 1/\omega$. That is exactly the signature of a CPE or a Warburg-like element, and it is indistinguishable in a single sweep.

- **Discriminating observation.** Repeating the measurement in the reverse frequency direction, or with a drift-corrected reference, should change the low-frequency tail while leaving the mid-frequency arc intact.
- **Test.** Direction-reversed sweeps at fixed DC bias.
- **Prior art.** Drift artefacts in low-frequency EIS.
- **Refuted if.** Forward and reverse sweeps give the same low-frequency exponent.

C19 — X-ray absorption edge positions shift with sample thickness through self-absorption, and the shift is read as an oxidation-state change

`synchrotron` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Edge energy is used as a linear indicator of oxidation state.
- **Mechanism.** In fluorescence-yield mode, self-absorption distorts the edge shape and shifts the half-height energy by an amount that depends on concentration and thickness; the shift is of the same order as a 0.1–0.3 oxidation-state change.
- **Discriminating observation.** The half-height energy should shift when the sample is diluted at fixed oxidation state, and should converge at low concentration.
- **Test.** Dilution series of a mixed-valence standard.
- **Prior art.** Self-absorption corrections in XAS.
- **Refuted if.** Edge energy is dilution-independent within 0.05 eV.

C20 — Raman-based temperature and strain measurements in operando cells are degenerate

`spectroscopy` · `degeneracy` · readiness 4 · novelty 3

- **Baseline.** Peak position is converted to temperature using a calibrated coefficient, with strain neglected.
- **Mechanism.** Both temperature and biaxial strain shift the phonon frequency; in an operando cell the strain state changes with state of charge. A single peak therefore gives a one-dimensional projection whose mixing coefficient depends on the material's mode Grüneisen parameter.
- **Discriminating observation.** The two components can be separated using a second peak with a different Grüneisen parameter; the inferred temperature should change when both peaks are used jointly.
- **Test.** Dual-peak operando Raman on a battery cathode.
- **Prior art.** Raman thermometry in electrochemical cells.
- **Refuted if.** Single- and dual-peak temperatures agree across the full state-of-charge range.

C21 — The turnover number in photocatalysis is computed against a catalyst mass that changes during the reaction

`photochem` · `hidden-parameter` · readiness 4 · novelty 2

- **Baseline.** TON = moles converted / moles catalyst loaded.
- **Mechanism.** Photocorrosion and dissolution remove catalyst during the run; the active mass at the end is lower than the loaded mass, and in some systems the dissolved species are themselves active. TON is therefore a lower bound whose relation to the true per-site turnover depends on an unmeasured dissolution curve.
- **Discriminating observation.** TON should scale superlinearly with loading if dissolution is a fixed fraction, and the dissolved fraction should correlate with the TON anomaly.
- **Test.** Loading series with post-run solution analysis.
- **Prior art.** Photostability benchmarking.
- **Refuted if.** TON is strictly linear in loading for a photostable standard.

C22 — Electron-spin-resonance quantification of radicals ignores a temperature-dependent cavity filling factor

`spectroscopy` · `calibration` · readiness 4 · novelty 2

- **Baseline.** Radical concentration is from double integration of the ESR signal relative to a standard.

- **Mechanism.** The resonator filling factor and the dielectric loss of the sample change with temperature and with solvent; the standard and the sample are measured under different conditions, so the ratio carries a systematic that scales with the dielectric constant difference.
- **Discriminating observation.** The inferred concentration should depend on the solvent's dielectric constant at fixed radical concentration.
- **Test.** Solvent series with a stable radical.
- **Prior art.** Quantitative ESR practice.
- **Refuted if.** Inferred concentration is solvent-independent.

C23 — The rate of a bimolecular reaction in a nanoreactor is reported per volume but governed by an interfacial concentration

`physical-chem` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Rates in micelles, vesicles, and porous catalysts are normalised by total volume.
- **Mechanism.** Reactants partition into a small interfacial volume; the local concentration can exceed the bulk by 10^2–10^3, so the second-order rate law in bulk units shows an apparent rate constant that is not a constant and that depends on the partition coefficient.
- **Discriminating observation.** The apparent second-order constant should scale with the measured partition coefficient across surfactants, collapsing onto one curve in local-concentration units.
- **Test.** Surfactant series with measured partition coefficients.
- **Prior art.** Micellar catalysis kinetics.
- **Refuted if.** The bulk-normalised constant is surfactant-independent.

C24 — The oxygen evolution overpotential includes a bubble-coverage term that is a function of the electrode's history

`electrochem` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Overpotential at fixed current is reported after conditioning.
- **Mechanism.** Bubble nucleation site density depends on surface roughness and on prior bubble history; coverage reduces the active area by 10–40% at high current, and the coverage-relaxation time is minutes. A measurement taken before steady coverage overestimates activity.
- **Discriminating observation.** The overpotential should drift downward for minutes after each current step, and the drift should scale with the current step size and with surface roughness.
- **Test.** Chronopotentiometry with step-size variation.
- **Prior art.** Bubble-coverage effects in gas-evolving electrodes.
- **Refuted if.** Overpotential reaches steady state in <1 s at all currents.

C25 — Differential scanning calorimetry's melting point is biased by thermal lag that is proportional to the sample mass and to the scan rate

`thermal-analysis` · `boundary-condition` · readiness 5 · novelty 2

- **Baseline.** Melting onset extrapolated to zero scan rate is taken as the thermodynamic melting point.
- **Mechanism.** The extrapolation is linear in scan rate, but the lag has a term proportional to mass times rate that is not linear when the pan's thermal resistance is comparable to the sample's; the intercept therefore shifts with mass, and the standard protocol fixes mass only loosely.
- **Discriminating observation.** Zero-rate extrapolated onset should depend on sample mass, converging only below a critical mass.
- **Test.** Mass series with the same material.
- **Prior art.** Calibration standards for DSC.
- **Refuted if.** Extrapolated onset is mass-independent within 0.1 K.

C26 — N_2 physisorption surface areas of microporous carbons depend on the adsorptive, and the BET model is applied outside its validity range

`materials-chem` · `assumption` · readiness 5 · novelty 2

- **Baseline.** BET area from the N_2 77 K isotherm with a Rouquerol consistency check.
- **Mechanism.** For micropores narrower than ~1 nm, the BET monolayer capacity is not a monolayer; CO_2 at 273 K fills these pores differently and gives areas differing by up to 40%. Reported areas are therefore method-defined rather than geometric.
- **Discriminating observation.** BET-N_2 and BET-CO_2 areas should diverge systematically for a series with controlled pore width, with the divergence signalling the BET breakdown.
- **Test.** Series of carbons with varied pore width, both adsorptives.
- **Prior art.** BET consistency criteria.
- **Refuted if.** The two areas agree within 10% across the pore-width series.

C27 — The apparent activation energy of a supported catalyst changes with metal loading because the rate is diffusion-masked at high loading

`catalysis` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Ea is reported per catalyst and compared across loadings.
- **Mechanism.** Higher loading increases the local reaction rate and the local heat/mass demand; if the pellet's internal diffusion is partly limiting, the apparent Ea is reduced by half for a first-order reaction, so the loading series looks like a change in intrinsic activity.
- **Discriminating observation.** The apparent Ea should drop by a factor approaching two at the loading where the Thiele modulus exceeds unity, and should recover when the pellet is crushed.
- **Test.** Loading series with crush tests.
- **Prior art.** Weisz–Prater analysis.
- **Refuted if.** Ea is loading-independent for crushed and intact pellets.

C28 — The pKa of a surface site derived from titration depends on the ionic strength through a double-layer term that is not fitted

`physical-chem` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Surface pKa values come from acid–base titration fitted with a 1-pK or 2-pK model.
- **Mechanism.** Charged surface groups create a potential that shifts the local H^+ activity by a factor that depends on ionic strength; fitting without a double-layer term redistributes that dependence into the pKa, so the fitted pKa is an ionic-strength-dependent quantity that is quoted as constant.
- **Discriminating observation.** The fitted pKa should vary with ionic strength in a way that vanishes when a charge-regulation model is included.
- **Test.** Titrations at several ionic strengths on one oxide.
- **Prior art.** Surface complexation modelling.
- **Refuted if.** Fitted pKa is ionic-strength independent.

C29 — Photoluminescence quantum yield measured in a film includes reabsorption, which depends on film thickness and on the overlap integral

`photochem` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** PLQY of a film is from an integrating-sphere measurement and is treated as an intrinsic property.
- **Mechanism.** Emitted photons are reabsorbed with a probability set by the overlap of emission and absorption and by the film's optical thickness; the reabsorption can be re-emitted or lost to a non-radiative channel, so PLQY depends on thickness and on geometry.
- **Discriminating observation.** PLQY should change with film thickness at fixed material, converging only for optically thin films, and the change should be predicted by the overlap integral.

- **Test.** Thickness series with absorption/emission spectra.
- **Prior art.** Reabsorption corrections.
- **Refuted if.** PLQY is thickness-independent to 2%.

C30 — The "intrinsic" conductivity of a doped polymer is dominated by intergrain barriers whose width is set by the dopant's hygroscopicity

`materials-chem` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Conductivity is reported for doped polymer films from four-point measurements and compared across dopants.
- **Mechanism.** Hygroscopic dopants absorb water, which forms a surface conduction layer; the measured conductivity is then a parallel combination of bulk and surface paths, and it depends on ambient humidity during measurement and on film morphology.
- **Discriminating observation.** Conductivity should vary with humidity in a way that scales with dopant hygroscopicity, and should drop when measured in dry atmosphere.
- **Test.** Humidity-controlled four-point measurements across dopants.
- **Prior art.** Doping and stability in conducting polymers.
- **Refuted if.** Conductivity is humidity-independent across dopants of differing hygroscopicity.

C31 — The fluorescence lifetime from TCSPC is biased by the instrument response's wavelength dependence when the emission is broadband

`spectroscopy` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** The IRF is measured at the excitation wavelength or with a scatterer and applied to the emission at a different wavelength.
- **Mechanism.** Detector transit-time and monochromator dispersion make the IRF wavelength-dependent by tens of picoseconds; for lifetimes below 1 ns the reconvolution fit absorbs the mismatch into a second decay component, producing spurious multi-exponentiality.
- **Discriminating observation.** The fitted lifetime should change when the IRF is measured at the emission wavelength, and the number of required components should drop.
- **Test.** Reanalysis of published TCSPC data with wavelength-resolved IRF.
- **Prior art.** IRF correction in TCSPC.
- **Refuted if.** Lifetimes and component counts are unchanged by wavelength-matched IRF.

C32 — The measurement of oxygen non-stoichiometry in oxides by thermogravimetry includes a buoyancy correction that drifts with gas composition

`materials-chem` · `hidden-parameter` · readiness 4 · novelty 2

- **Baseline.** δ in $ABO_{3-\delta}$ is from mass change, with buoyancy corrected using the initial gas.
- **Mechanism.** Buoyancy depends on gas density, which changes when the atmosphere is switched between O_2, N_2, and CO_2; the correction computed for one gas leaves a residual that is comparable to a 0.01 change in δ.
- **Discriminating observation.** The apparent δ should jump at each gas switch in a way that is proportional to the density difference and not to the oxygen activity step.
- **Test.** Gas-switch experiments at fixed oxygen activity.
- **Prior art.** Buoyancy corrections in TGA.
- **Refuted if.** δ is continuous across gas switches at fixed pO_2.

C33 — The dielectric constant of a confined liquid is inferred from a capacitor that measures the series combination

`physical-chem` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Confined-water permittivity is extracted from the capacitance of a nano-gap, dividing by the gap.

- **Mechanism.** The gap contains the liquid plus two interfacial layers whose permittivity differs; the series combination means the extracted value is dominated by the *lowest* permittivity layer, not by the bulk. A two-layer fit requires the layer thickness, which is exactly the unknown.
- **Discriminating observation.** The extracted permittivity should depend on gap thickness in a way that follows a series model; the apparent "interfacial layer" thickness should be independent of the liquid.
- **Test.** Gap-thickness series with several liquids.
- **Prior art.** Nano-gap capacitance measurements.
- **Refuted if.** Extracted permittivity is gap-independent.

C34 — NMR relaxation-derived pore-size distributions assume a surface relaxivity that is calibrated on a different mineralogy

`geochem` · `extrapolation` · readiness 4 · novelty 3

- **Baseline.** T_2 distributions are converted to pore-size distributions using a laboratory-calibrated surface relaxivity ρ_2.
- **Mechanism.** ρ_2 varies by an order of magnitude with mineralogy and with the presence of paramagnetic impurities; using a single value redistributes pore sizes without changing the total porosity, so the inversion looks self-consistent while being systematically wrong.
- **Discriminating observation.** The T_2-derived pore size should correlate with the sample's paramagnetic content at fixed measured pore size (from imaging).
- **Test.** Samples with paired NMR and imaging.
- **Prior art.** NMR pore-size conversion.
- **Refuted if.** Converted pore sizes are unbiased across mineralogies with one ρ_2.

C35 — The electrochemically measured double-layer capacitance of a porous electrode follows a power law in frequency that is indistinguishable from a constant-phase element

`electrochem` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** A constant-phase element exponent below 1 is interpreted as interfacial heterogeneity or as a fractal roughness.
- **Mechanism.** In a porous electrode with a distributed RC transmission line, the impedance is a power law over the frequency range where the penetration depth lies inside the pore; that range is wide and mimics a CPE with exponent 0.5–0.9. The exponent is then a pore-geometry parameter, not a roughness parameter.
- **Discriminating observation.** The CPE exponent should change with electrode thickness at fixed material, following the transmission-line prediction.
- **Test.** Thickness series of the same electrode material.
- **Prior art.** Porous-electrode transmission-line models.
- **Refuted if.** CPE exponent is thickness-independent.

C36 — The measured enthalpy of adsorption from isotherms depends on the isotherm model because the model fixes the saturation capacity

`physical-chem` · `model-disagreement` · readiness 4 · novelty 2

- **Baseline.** Isosteric heats come from applying the Clausius–Clapeyron relation to isotherms fitted with Langmuir or dual-site models.
- **Mechanism.** q_{st} at a given loading depends on the loading axis, which depends on the fitted saturation capacity; for microporous materials where saturation is ill-defined, different models give q_{st} differing by 10–20% from the same data.
- **Discriminating observation.** q_{st} should differ between models by more than the reported uncertainty, and the model dependence should vanish for data sets that reach a clear plateau.
- **Test.** Refit published isotherm families with several models.
- **Prior art.** Isosteric heat methodology.

- **Refuted if.** q_st is model-independent to 3%.

C37 — The measured Lewis basicity of an ionic liquid is a function of its water content, which is rarely measured

`physical-chem` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Solvatochromic and electrochemical basicity scales are reported for dried ionic liquids.
- **Mechanism.** Ionic liquids absorb water to millimolar-to-molar levels within minutes of air exposure; water is a strong hydrogen-bond donor and shifts both the probe's response and the electrochemical window. Drying protocols differ between labs, so the scales are not comparable.
- **Discriminating observation.** The basicity parameter should correlate with the Karl Fischer water content across published samples, with a slope that is independent of the ionic liquid family.
- **Test.** Reanalysis of literature values with water content metadata.
- **Prior art.** Solvatochromic parameters of ionic liquids.
- **Refuted if.** Basicity is uncorrelated with water content.

C38 — Scanning electrochemical microscopy assigns a finite rate constant to insulating regions because its theory assumes a conductive substrate

`electrochem` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Approach curves are fitted with feedback theory that assumes the substrate is a conductor with a kinetically limited rate, and the fitted rate constant is reported as a map of activity.
- **Mechanism.** Over an insulating region the current is instead limited by the hindered diffusion through the tip–substrate gap; that curve is well fitted by the conducting-substrate expression with a small but nonzero rate constant. The map therefore reports "low activity" where the surface is simply inert, and the fitted value depends on the tip's radius-to-distance ratio.
- **Discriminating observation.** The apparent rate constant over a known insulator should change systematically with the tip's approach speed and with the tip radius, and should be indistinguishable from zero only when the diffusion-only model is used.
- **Test.** Approach curves on a patterned conductor/insulator with several tip sizes.
- **Prior art.** SECM feedback theory and its limits.
- **Refuted if.** Apparent rate constants over an insulator are tip-radius independent and distinguishable from zero.

C39 — The apparent cooperativity of a supramolecular host is an artefact of fitting a 1:1 model to a 1:2 system

`supramolecular` · `model-artifact` · readiness 5 · novelty 2

- **Baseline.** Binding isotherms are fitted with a 1:1 model and the residuals are interpreted as anti-cooperativity.
- **Mechanism.** A sequential 1:2 binding with negative cooperativity produces a curve that a 1:1 model fits with a Hill coefficient near 0.6; the same curve is produced by a 1:1 model with a small concentration error. The two are distinguishable only with a dilution series.
- **Discriminating observation.** Global fitting of several host concentrations should favour 1:2, while single-curve fitting cannot distinguish them; residual structure should be concentration-dependent.
- **Test.** Global analysis of a concentration series.
- **Prior art.** Binding-model selection.
- **Refuted if.** Single-curve and global fits agree on the model.

C40 — The thermal conductivity of a liquid from the transient hot-wire method is biased by natural convection that begins earlier than the standard criterion assumes

`physical-chem` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** The analysis uses a time window bounded by the convection onset predicted from the Rayleigh number with a fixed critical value.

- **Mechanism.** The critical Rayleigh number depends on the aspect ratio and on the temperature dependence of viscosity; for low-viscosity liquids the onset can occur within the standard window, adding a convective contribution that raises the apparent conductivity by several percent.
- **Discriminating observation.** The apparent conductivity should depend on the wire's orientation relative to gravity (horizontal wires convect earlier) and on the applied power.
- **Test.** Orientation and power series in one liquid.
- **Prior art.** Hot-wire convection corrections.
- **Refuted if.** Apparent conductivity is orientation-independent at all powers.

C41 — The measured Seebeck coefficient of a thin film includes a contact thermopower that is of the same order as the film's

`materials-chem` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Film Seebeck is measured between two probes on the film surface with an assumed ΔT.
- **Mechanism.** The probe–film contact has its own thermopower, and the measured voltage is the difference of the two contacts' temperatures weighted by their Seebeck coefficients. For high-Seebeck films the contact contribution is a small fraction, but for low-Seebeck metals it can dominate.
- **Discriminating observation.** Measured Seebeck should change when probe material is changed, with a magnitude equal to the difference of contact thermopowers.
- **Test.** Multi-probe-material comparison on the same film.
- **Prior art.** Contact thermopower in Seebeck metrology.
- **Refuted if.** Seebeck is probe-material independent.

C42 — The turnover frequency of a single-atom catalyst is computed per total metal atom, but most atoms are inactive

`catalysis` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** TOF = rate / (total metal loading), assuming all atoms are accessible.
- **Mechanism.** Single-atom catalysts contain a distribution of sites: isolated atoms, dimers, and clusters. Poisoning experiments show that a minority of sites carries most of the rate, so the per-total-atom TOF is a lower bound whose ratio to the per-active-site TOF is unknown and catalyst-specific.
- **Discriminating observation.** Site-counting by CO titration or by in-situ spectroscopy should give an active-site fraction that correlates with the TOF discrepancy across catalysts.
- **Test.** Paired site counting and kinetics.
- **Prior art.** Site counting in SACs.
- **Refuted if.** Active-site fraction is unity for all catalysts in the series.

C43 — The measured viscosity of a confined fluid is dominated by the wall's roughness, not by the fluid

`physical-chem` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Surface-force-apparatus drainage measurements give an effective viscosity that is compared with bulk.
- **Mechanism.** Roughness of a few ångströms changes the effective gap and the slip length; the extracted viscosity scales with the cube of the gap for a drainage measurement, so a 0.5 nm error in gap gives a 10% viscosity error at 3 nm separation.
- **Discriminating observation.** The extracted viscosity should depend on the gap in a way that is fixed by the measured roughness topography, and should converge for smooth surfaces.
- **Test.** Same fluid, surfaces of varying roughness.
- **Prior art.** SFA drainage measurements.
- **Refuted if.** Extracted viscosity is roughness-independent.

C44 — The reported stability number of a metal–organic framework is set by the detection limit of the diffraction measurement

`materials-chem` · `blind-spot` · readiness 4 · novelty 3

- **Baseline.** Framework stability is judged by the retention of Bragg peaks after treatment.
- **Mechanism.** A framework that loses 90% of its crystallinity can still show the strongest Bragg peaks above the detection limit; the reported "stable" material may be largely amorphous. The stability ranking therefore saturates at the diffraction detection limit.
- **Discriminating observation.** Pairing diffraction with surface-area or with total-scattering PDF analysis should reveal loss of long-range order in samples reported as stable.
- **Test.** Reanalysis of a stability series with PDF or BET.
- **Prior art.** MOF stability benchmarking.
- **Refuted if.** Samples with retained Bragg peaks retain >80% of their surface area.

C45 — The measured dielectric breakdown strength of a thin film is limited by the electrode's edge field, not by the film

`materials-chem` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Breakdown field is voltage/thickness for a parallel-plate capacitor.
- **Mechanism.** Electrode edges concentrate the field; for films thinner than the electrode-edge radius the enhancement factor is set by the edge geometry, so breakdown occurs at a nominal field below the intrinsic value and the measured field scales with electrode size.
- **Discriminating observation.** Breakdown field should increase with electrode radius (fewer defects and lower edge enhancement) and should saturate for large electrodes; the Weibull slope should change accordingly.
- **Test.** Electrode-size series.
- **Prior art.** Area-scaling in dielectric breakdown.
- **Refuted if.** Breakdown field is electrode-size independent.

C46 — The apparent rate of a photochemical reaction in a flow reactor is set by the photon flux distribution, not by the catalyst

`chem-eng` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Rates are normalised by catalyst mass and by incident photon flux, assuming uniform illumination.
- **Mechanism.** In a packed bed the light penetrates a few hundred micrometres; the illuminated fraction depends on particle size and on bed orientation. Two reactors with the same incident flux can differ by an order of magnitude in volumetric rate, so cross-lab comparisons of catalyst activity are confounded with reactor geometry.
- **Discriminating observation.** Volumetric rate should scale with the illuminated external area rather than with mass, collapsing across reactors when normalised by that area.
- **Test.** Particle-size and bed-orientation series.
- **Prior art.** Photoreactor characterisation.
- **Refuted if.** Mass-normalised rates collapse across geometries.

C47 — The measured hydrogen evolution overpotential on a single crystal depends on the surface's reconstruction, which is set by the electrolyte's cation

`electrochem` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Activity trends across metals are compared at fixed electrolyte.
- **Mechanism.** Specific cation adsorption drives reconstruction and changes the number of active edge sites; the same crystal face can differ by a factor of two between Li^+ and Cs^+ electrolytes. Trends measured in one electrolyte therefore do not transfer.

- **Discriminating observation.** The activity ordering of two metals should invert between electrolytes, correlated with the measured cation coverage.
- **Test.** Cation series on single-crystal electrodes.
- **Prior art.** Cation effects in HER.
- **Refuted if.** Metal ordering is electrolyte-independent.

C48 — The quantification of trace metals by ICP-MS is biased by the sample matrix's carbon content through charge transfer

`analytical-chem` · `hidden-parameter` · readiness 4 · novelty 2

- **Baseline.** Calibration curves are built in dilute nitric acid and applied to digested samples.
- **Mechanism.** Residual carbon from digestion enhances ionization of elements with high ionization potential (As, Se, Hg) in the plasma; the enhancement depends on the digestion method, so a calibration in pure acid underestimates these elements by 10–30% in carbon-rich digests.
- **Discriminating observation.** The recovery of a spike should correlate with the sample's residual carbon, and should match when the calibration is matrix-matched.
- **Test.** Spike-recovery series across digestion methods.
- **Prior art.** Carbon enhancement in ICP-MS.
- **Refuted if.** Recovery is independent of residual carbon.

C49 — The reported photostability half-life of a dye is set by the oxygen concentration in the cuvette

`photochem` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Photobleaching rates are measured in air-equilibrated solution and reported as intrinsic.
- **Mechanism.** Oxygen both quenches the triplet and generates singlet oxygen that destroys the dye; the net effect is non-monotonic in oxygen concentration. Comparisons across labs with different degassing therefore reflect dissolved oxygen rather than molecular stability.
- **Discriminating observation.** Bleaching rate should be non-monotonic in controlled oxygen partial pressure, with a minimum at an intermediate concentration.
- **Test.** Oxygen-titration photolysis.
- **Prior art.** Photobleaching mechanisms.
- **Refuted if.** Bleaching rate is monotonic in oxygen across the full range.

C50 — The "intrinsic" catalytic activity from density-functional theory is compared with experiment through a free-energy diagram that omits configurational entropy of adsorbates

`computational-chem` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** DFT free energies include vibrational entropy of the adsorbate on a fixed site.
- **Mechanism.** At operating temperature, adsorbates hop among symmetry-equivalent sites and mix orientations; the configurational entropy is up to 0.2 eV at 500 K and differs between adsorbed species, shifting the predicted volcano position by more than the experimental spread.
- **Discriminating observation.** Including configurational entropy should change the predicted optimum across a class of surfaces, and the change should correlate with the adsorbate's binding-site multiplicity.
- **Test.** Reanalysis of published DFT volcano data with configurational terms.
- **Prior art.** Ab initio thermodynamics of adsorbates.
- **Refuted if.** Volcano ordering is unchanged by configurational entropy for high-multiplicity adsorbates.

Materials science — 50 hypotheses

Every entry is a proposition. No experiment below has been performed.

M01 — Reported in-plane thermal conductivity of a supported 2D film is a function of the laser spot size, and in the large-spot limit it measures the substrate conductance instead

`materials` · `boundary-condition` · readiness 4 · novelty 4

- **Baseline.** Optothermal Raman (or thermoreflectance) fits a 2D heat-spreading model with a substrate conductance G and returns κ.
- **Mechanism.** For a Gaussian source on a sheet with substrate loss, $\Delta T = (P/4\pi\kappa d)\cdot e^{\lambda}E_1(\lambda)$ with $\lambda = Gs^2/(2\kappa d)$. The standard fit returns $\kappa_{app} = \kappa\cdot[e^{\lambda}E_1(\lambda)]^{-1}$; for $\lambda > 10$ this asymptotes to $\kappa_{app} \approx Gs^2/(2d)$, independent of the material's κ.
- **Discriminating observation.** Across published measurements of one material, κ_{app} should collapse onto a single κ when plotted against λ computed from each paper's spot radius and reported G; and κ_{app} should scale as s^2 for large spots.
- **Test.** Reanalysis of a published κ compilation with spot radius and G metadata; full derivation and numerics in investigation I2.
- **Prior art.** Steady-state Raman thermometry modelling; I2 of this collection.
- **Refuted if.** κ_{app} is uncorrelated with spot radius across a compilation with s spanning 0.2–3 μm.

M02 — Thin-film electrical conductivity is not a material property when the pads are comparable in conductivity to the film

`materials` · `boundary-condition` · readiness 5 · novelty 2

- **Baseline.** Four-bar measurements are assumed to remove contact effects exactly.
- **Mechanism.** Finite pad conductivity lets current spread into the pads, adding a voltage component that grows without bound as film and pad conductivities approach each other; the error is not a contact resistance and cannot be calibrated away by a series-resistance measurement.
- **Discriminating observation.** Extracted conductivity should depend on pad metal at fixed film, with a discrepancy scaling as the conductivity ratio.
- **Test.** Reanalysis and a targeted pad-metallurgy series.
- **Prior art.** arXiv:2602.02418.
- **Refuted if.** Extracted conductivity is pad-material invariant across a 100× range.

M03 — Interface-dominated thermal boundary conductance measured by time-domain thermoreflectance is biased by the transducer's own boundary

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** TDTR fits a two-layer model (transducer/film/substrate) with one interface conductance per boundary.
- **Mechanism.** The transducer–film boundary and the film–substrate boundary have comparable conductances, and the sensitivity to each depends on the modulation frequency. A single-frequency fit therefore returns a weighted combination, and the reported G is not the interface the authors name.
- **Discriminating observation.** The fitted G should vary with modulation frequency in a way predicted by the two-boundary sensitivity matrix, and should converge only when one boundary dominates the sensitivity.
- **Test.** Frequency-series TDTR on a sample with independently characterised transducer adhesion.
- **Prior art.** TDTR sensitivity analysis.
- **Refuted if.** Fitted G is modulation-frequency independent.

M04 — Grain-boundary resistivity extracted from impedance spectra assumes a brickwork geometry whose exponent is fitted, not measured

`materials` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Grain and grain-boundary resistances are separated by fitting an equivalent circuit with a constant phase element for the boundary.
- **Mechanism.** The CPE exponent and the geometric factor (grain boundary thickness / grain size) are degenerate: a lower exponent can be compensated by a thinner boundary layer. The reported grain-boundary resistivity therefore depends on the assumed microstructure, which is measured on a different specimen.
- **Discriminating observation.** The fitted CPE exponent should correlate with the measured grain-boundary thickness fraction across samples; when it does not, the resistivity separation is not unique.
- **Test.** Paired impedance and electron-microscopy grain-size statistics.
- **Prior art.** Brickwork model for polycrystal impedance.
- **Refuted if.** The exponent is uncorrelated with microstructure yet resistivity is reproducible.

M05 — The fracture toughness of a thin film measured by channel cracking includes a residual-stress term that is set by the deposition, not by the film

`materials` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Channel-crack spacing gives toughness through a steady-state energy-release-rate relation that uses the measured residual stress.
- **Mechanism.** The residual stress measured by wafer curvature is an average over the film and substrate; the near-surface stress that drives cracking differs, and the difference scales with deposition process. Toughness values therefore track process history.
- **Discriminating observation.** Toughness should correlate with the curvature-derived stress only weakly, and should collapse better against a depth-resolved stress measure (e.g. from Raman or XRD at several depths).
- **Test.** Reanalysis with depth-resolved stress on a process series.
- **Prior art.** Channel cracking metrology.
- **Refuted if.** Toughness is process-independent at fixed composition and thickness.

M06 — Nanoindentation hardness depends on the tip's actual area function at small depths, and the calibration is performed on a material with a different modulus

`materials` · `calibration` · readiness 4 · novelty 3

- **Baseline.** The area function is calibrated on fused silica and applied to stiffer or softer materials.
- **Mechanism.** The tip's apex is blunt at the nanometre scale, and the effective contact radius depends on the elastic mismatch between tip and sample; the calibrated area function therefore carries a modulus-dependent error that grows at shallow depths and produces the widely reported indentation size effect.
- **Discriminating observation.** The apparent size effect should change when the same tip is calibrated on a material of matched modulus, and should scale with the modulus ratio.
- **Test.** Cross-calibration on several reference materials.
- **Prior art.** Oliver–Pharr analysis and area-function calibration.
- **Refuted if.** Hardness–depth curves are unchanged by calibration material.

M07 — The measured bandgap of a nanocrystal film includes a strain shift that tracks the ligand, not the size

`materials` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Absorption onset is converted to bandgap and plotted against nanocrystal size to extract confinement.
- **Mechanism.** Ligand exchange changes surface strain and the dielectric environment, shifting the onset by tens of meV; size series are usually made with a single ligand, so the shift is absorbed into the confinement model and distorts the extracted effective mass.

- **Discriminating observation.** For a fixed size, the onset should shift with ligand length and with the matrix dielectric constant, and the shift should follow the known deformation potential.
- **Test.** Fixed-size series with ligand and matrix variation.
- **Prior art.** Dielectric screening and strain effects in nanocrystals.
- **Refuted if.** Onset is ligand- and matrix-independent at fixed size.

M08 — Piezoresponse force microscopy reports ferroelectricity in materials that are only electromechanically active

`materials` · `blind-spot` · readiness 4 · novelty 3

- **Baseline.** Hysteretic amplitude and phase loops in PFM are taken as evidence of switchable ferroelectric domains.
- **Mechanism.** Electrostatic tip–sample forces, ionic drift, and Joule heating all produce hysteretic loops with the same phase signature; the cantilever's second resonance and the tip's contact stiffness set which dominates. Without a vacuum/atmosphere comparison, the loops do not discriminate.
- **Discriminating observation.** Loops should change qualitatively between vacuum and controlled humidity, and should persist in a non-ferroelectric ionic conductor.
- **Test.** Vacuum/humidity series on a known non-ferroelectric ionic conductor.
- **Prior art.** PFM artefact literature.
- **Refuted if.** Loops are unchanged in a non-ferroelectric ionic conductor under all atmospheres.

M09 — The superconducting critical current of a coated conductor is limited by a current-transfer length that is not in the reported Jc

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Jc is reported as Ic divided by the superconductor cross-section.
- **Mechanism.** Current injected from the stabiliser must transfer into the superconducting layer over a length set by the interface resistance; for short samples the transfer length is comparable to the sample length, so the local current density is non-uniform and Jc depends on sample length.
- **Discriminating observation.** Jc should increase with sample length and saturate beyond the transfer length, with the saturation length scaling as the inverse square root of interface conductance.
- **Test.** Length series on tapes with measured interface resistance.
- **Prior art.** Current-transfer-length analysis.
- **Refuted if.** Jc is length-independent for samples shorter than 1 cm.

M10 — The measured thermal expansion of a film on a substrate is the difference of two large numbers and inherits the substrate's uncertainty

`materials` · `model-disagreement` · readiness 4 · novelty 2

- **Baseline.** Film expansion is from substrate curvature using Stoney-type relations.
- **Mechanism.** The film's contribution is a small difference between the measured curvature and the substrate's own thermal bending; for films thinner than 100 nm the difference is comparable to the uncertainty in the substrate's biaxial modulus, so the reported CTE can have the wrong sign.
- **Discriminating observation.** CTE should change when the substrate is swapped for a different orientation of the same material, at fixed film.
- **Test.** Multi-substrate comparison.
- **Prior art.** Curvature-based stress metrology.
- **Refuted if.** Film CTE is substrate-independent within errors.

M11 — The "intrinsic" mobility from a field-effect transistor includes contact resistance that is fitted, not measured

`materials` · `circularity` · readiness 4 · novelty 3

- **Baseline.** Mobility is from the transfer curve's transconductance with a contact resistance extracted by the transmission-line method on a separate device.
- **Mechanism.** The transmission-line extraction assumes the channel resistance scales linearly with length, which fails when the contact is not the limiting element or when the semiconductor is non-uniform; the extracted contact resistance then absorbs channel nonlinearity and the mobility becomes length-dependent.
- **Discriminating observation.** Mobility should vary with channel length for a well-behaved material; a length-independent mobility after correction is the test of a valid extraction.
- **Test.** Length series on one wafer.
- **Prior art.** Contact-resistance extraction in 2D FETs.
- **Refuted if.** Mobility is length-independent before any correction.

M12 — Cathodoluminescence quantum efficiency of a phosphor is measured under an excitation density that is orders of magnitude above the application's

`materials` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Efficiency is measured at high beam current for signal-to-noise, then quoted for low-density operation.
- **Mechanism.** Auger and exciton–exciton annihilation make efficiency non-monotonic in excitation density; the measured value at high density is on the drooping branch, so the quoted efficiency underestimates the low-density value for some phosphors and overestimates it for others.
- **Discriminating observation.** Efficiency should be non-monotonic in beam current with a maximum at a density that depends on the host's Auger coefficient, and the maximum should shift with temperature.
- **Test.** Current-density series on a standard phosphor.
- **Prior art.** Efficiency droop in luminescent materials.
- **Refuted if.** Efficiency is density-independent across three decades of excitation.

M13 — The degradation rate of a photovoltaic absorber measured under one spectrum does not transfer to another because the damage is carrier-density-driven

`materials` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Stability is reported as T80 under standard test conditions and extrapolated to field.
- **Mechanism.** The degradation rate depends on the steady-state carrier density, which differs between spectra with the same irradiance because the absorption profile and the operating point differ; accelerated tests with a different spectrum therefore rank materials differently.
- **Discriminating observation.** T80 measured under two spectra with matched irradiance but different carrier generation profiles should differ, and the difference should scale with the measured carrier lifetime.
- **Test.** Dual-spectrum stability testing.
- **Prior art.** ISOS stability protocols.
- **Refuted if.** T80 is spectrum-independent at matched irradiance.

M14 — The measured fracture strain of a metallic glass depends on the sample's free volume, which is set by the quenching rate and is not reported

`materials` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Ductility is reported as a composition property.
- **Mechanism.** Free volume controls shear-band nucleation; quenching rate changes free volume by a fraction that changes the fracture strain by tens of percent. Samples of the same composition from different thicknesses therefore differ systematically.
- **Discriminating observation.** Fracture strain should decrease with sample thickness at fixed composition, tracking the measured enthalpy of relaxation.
- **Test.** Thickness series with DSC relaxation-enthalpy measurements.
- **Prior art.** Free-volume effects in metallic glasses.

- **Refuted if.** Fracture strain is thickness-independent for a fixed composition.

M15 — The measured magnetocaloric entropy change depends on the protocol's field-sweep rate through a lag that is never corrected

`materials` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** ΔS_M is from isothermal magnetisation or from heat-capacity measurements at quasi-static sweep rates.
- **Mechanism.** Eddy currents and thermal lag in a bulk sample shift the measured transition and broaden it; the peak ΔS_M decreases with increasing sweep rate, so values measured at different rates are not comparable and the reported peak is rate-dependent.
- **Discriminating observation.** Peak ΔS_M should decrease with sweep rate as a Lorentzian in rate with a width set by the sample's thermal diffusivity.
- **Test.** Sweep-rate series on samples of different thickness.
- **Prior art.** Magnetocaloric measurement protocols.
- **Refuted if.** Peak ΔS_M is sweep-rate independent for bulk samples.

M16 — Reported thermoelectric figure of merit is dominated by the uncertainty in the thermal diffusivity, and the two measurements are made on different specimens

`materials` · `model-disagreement` · readiness 4 · novelty 2

- **Baseline.** $zT = S^2\sigma T/\kappa$, with κ from laser flash on a separate coupon.
- **Mechanism.** Anisotropy and texture mean the coupon's κ is not the leg's κ; the error is a sample-to-sample property, not a measurement error, and it biases zT by 10–30% in textured materials while being invisible in the reported uncertainty.
- **Discriminating observation.** zT computed from same-specimen κ (e.g. from the Harman method) should differ systematically from the separate-coupon value in textured samples.
- **Test.** Paired Harman and separate-coupon measurements.
- **Prior art.** Thermoelectric metrology.
- **Refuted if.** The two routes agree within 5% for textured samples.

M17 — The measured corrosion potential of a coated metal is set by the coating's pinholes, not by its barrier properties

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Barrier performance is ranked by corrosion potential or by low-frequency impedance modulus.
- **Mechanism.** A coating with 0.01% pinhole area can dominate the measured potential because the exposed metal is a nearly reversible electrode; the coating's bulk permeability is then invisible to the measurement until the pinholes are sealed.
- **Discriminating observation.** The measured potential should be independent of coating thickness above a threshold, and should change abruptly when pinholes are deliberately introduced.
- **Test.** Thickness series plus controlled-defect series.
- **Prior art.** Coating barrier metrology.
- **Refuted if.** Potential scales with coating thickness as a barrier model predicts.

M18 — The measured Curie temperature of a thin film is shifted by the film's own demagnetising field, and the shift depends on the patterning

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Tc is taken from magnetisation vs temperature and compared with bulk.
- **Mechanism.** Dipolar coupling in a patterned film modifies the effective exchange; the shift scales with the pattern's demagnetising factor and with film thickness, so continuous films, dots, and stripes of the same material give different Tc.

- **Discriminating observation.** Tc should vary with pattern geometry at fixed thickness, in proportion to the calculated demagnetising factor.
- **Test.** Same film patterned into dots and stripes.
- **Prior art.** Dipolar effects in thin-film magnetism.
- **Refuted if.** Tc is pattern-independent within 1 K.

M19 — The measured ionic conductivity of a solid electrolyte in a symmetric cell includes an interfacial resistance that grows with time

`materials` · `time-dependent` · readiness 5 · novelty 3

- **Baseline.** Bulk conductivity is from the high-frequency intercept after an initial impedance measurement.
- **Mechanism.** The electrode–electrolyte interface evolves (interphase growth, wetting), so the intercept moves during the first hours; a measurement taken early reports a higher conductivity, and the drift is often mistaken for the electrolyte degrading.
- **Discriminating observation.** The high-frequency intercept should drift monotonically over 24 h at fixed temperature, and should stabilise with an ion-blocking electrode.
- **Test.** Time-series impedance with blocking and non-blocking electrodes.
- **Prior art.** Interfacial resistance in solid electrolytes.
- **Refuted if.** The intercept is stable from the first measurement.

M20 — The measured work function of a 2D material is contaminated by adsorbates whose coverage tracks the transfer process

`materials` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Work function is from Kelvin-probe or photoemission and compared across materials.
- **Mechanism.** Transfer residues and ambient adsorbates shift the surface dipole by 0.1–0.5 eV; the shift is larger for materials with higher surface energy and is not removed by standard annealing without changing the material.
- **Discriminating observation.** Work function should vary with transfer method and with annealing temperature in a way that tracks the measured carbon/oxygen surface coverage.
- **Test.** Multi-technique surface characterisation on identical flakes.
- **Prior art.** Work-function metrology for 2D materials.
- **Refuted if.** Work function is transfer-method independent at fixed coverage.

M21 — The reported piezoelectric coefficient d_{33} of a thin film includes a substrate-clamping factor evaluated at the wrong boundary condition

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** d_{33} is from laser interferometry of a clamped film with a correction for substrate clamping.
- **Mechanism.** The clamping correction assumes the film is free to expand in-plane at the measurement frequency; above the film/substrate mechanical resonance the film decouples, and below it the correction is frequency-dependent. Measurements at different frequencies therefore give different d_{33}.
- **Discriminating observation.** d_{33} should vary with drive frequency in a way fixed by the substrate's compliance, converging at low frequency.
- **Test.** Frequency-series interferometry.
- **Prior art.** Clamping corrections in piezofilm metrology.
- **Refuted if.** d_{33} is frequency-independent.

M22 — The measured thermal stability of a phase-change material is a function of the capping layer, which is treated as inert

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Crystallisation temperature and data retention are reported for blanket films with a standard cap.
- **Mechanism.** The cap imposes stress and acts as a diffusion barrier for the chalcogen; both shift the nucleation rate. Retention extrapolations from a capped film therefore do not describe a device where the cap differs.
- **Discriminating observation.** Crystallisation temperature should shift with cap stress (varied by cap thickness) at fixed phase-change layer.
- **Test.** Cap-thickness series.
- **Prior art.** Phase-change retention extrapolation.
- **Refuted if.** Crystallisation temperature is cap-independent.

M23 — The reported spin Hall angle is contaminated by the inverse spin galvanic effect in the same bilayer

`materials` · `model-disagreement` · readiness 3 · novelty 3

- **Baseline.** Spin Hall angle is extracted from a harmonic Hall measurement assuming a single mechanism.
- **Mechanism.** The inverse spin galvanic and spin Hall contributions have the same symmetry in the standard geometry but different thickness scaling; a single-thickness measurement cannot separate them, and the extracted angle depends on the assumed ratio.
- **Discriminating observation.** The angle extracted at two thicknesses of the same heavy metal should be consistent only when both terms are fitted.
- **Test.** Thickness series of the heavy metal.
- **Prior art.** Harmonic Hall analysis.
- **Refuted if.** The angle is thickness-independent without a second term.

M24 — The measured thermal conductivity of a porous material by the transient plane source method includes a contact resistance that is a function of the applied pressure

`materials` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** The sensor is clamped to the sample and the contact resistance is assumed negligible.
- **Mechanism.** For porous and rough materials the contact resistance can exceed the sample's own resistance; because clamping pressure is rarely reported, interlaboratory scatter is dominated by it.
- **Discriminating observation.** Measured conductivity should increase with clamping pressure and saturate, with the saturation pressure depending on roughness.
- **Test.** Pressure-controlled measurements on the same specimen.
- **Prior art.** TPS metrology.
- **Refuted if.** Conductivity is pressure-independent above 1 kPa.

M25 — The measured optical bandgap of a mixed-halide perovskite is inhomogeneous, and the "gap" is a fit to a distribution

`materials` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** A Tauc or derivative fit gives one gap energy.
- **Mechanism.** Halide segregation creates domains with gaps differing by 100 meV; the absorption edge is a convolution and the fitted gap is a weighted average that depends on the fit window and on the domain size distribution.
- **Discriminating observation.** The fitted gap should change with the fit window and should correlate with the width of the photoluminescence spectrum.
- **Test.** Reanalysis of absorption and PL pairs across segregation states.
- **Prior art.** Bandgap extraction from Tauc plots.
- **Refuted if.** Fitted gap is window-independent and matches the PL peak.

M26 — The measured hardness of a high-entropy alloy is reported for a phase that is not single

`materials` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Hardness is reported as a single value for a nominal composition.
- **Mechanism.** Most "single-phase" high-entropy alloys contain minority intermetallic or sigma phases that pin dislocations; hardness is then dominated by the minority phase, and composition–property trends reflect phase fraction rather than solid solution strengthening.
- **Discriminating observation.** Hardness should correlate with the minority-phase area fraction across a composition series, and should collapse when normalised by it.
- **Test.** Paired hardness and phase-fraction measurements.
- **Prior art.** Solid-solution strengthening in HEAs.
- **Refuted if.** Hardness is uncorrelated with minority-phase fraction.

M27 — The measured dielectric loss of a low-loss material at microwave frequencies is dominated by the metal housing's surface resistance

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Cavity or split-post dielectric resonator measurements attribute all loss to the sample.
- **Mechanism.** The housing's surface resistance contributes a loss that is independent of the sample and that is comparable to the sample's loss for tan δ below 1e-5; the reported loss tangent is then a difference of two comparable numbers and depends on the housing's surface finish.
- **Discriminating observation.** Loss tangent should change when the housing is replaced with a different surface finish at fixed sample.
- **Test.** Housing-material comparison.
- **Prior art.** Microwave dielectric metrology.
- **Refuted if.** Loss tangent is housing-independent at the 1e-6 level.

M28 — The measured spin-wave lifetime in a thin film is limited by the film's thickness variation, not by the damping

`materials` · `hidden-parameter` · readiness 4 · novelty 2

- **Baseline.** Linewidth from ferromagnetic resonance gives the Gilbert damping.
- **Mechanism.** Thickness variation shifts the local resonance frequency; the ensemble linewidth is the convolution of the intrinsic linewidth with the thickness distribution. Reported damping therefore tracks deposition uniformity.
- **Discriminating observation.** Linewidth should correlate with the measured thickness spread across samples, and should narrow on a single-crystal film with known uniformity.
- **Test.** Paired FMR and thickness mapping.
- **Prior art.** Inhomogeneous broadening in FMR.
- **Refuted if.** Linewidth is uncorrelated with thickness spread.

M29 — The measured fracture toughness of a ceramic composite is a rising R-curve value quoted as a single number

`materials` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** Toughness is reported from a single test method at a single crack length.
- **Mechanism.** Bridging and transformation toughening produce a rising R-curve; indentation and single-edge-notched-beam tests sample different crack-length regimes, so the two methods disagree by design, not by error.
- **Discriminating observation.** The two methods should converge when the R-curve is measured over the same crack-length range.
- **Test.** R-curve measurement with both methods.
- **Prior art.** R-curve metrology.

- **Refuted if.** Methods disagree even at matched crack length.

M30 — The measured activation energy of grain-boundary diffusion from thin-film interdiffusion includes a surface-diffusion shortcut

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Interdiffusion profiles in bilayers give grain-boundary diffusivity.
- **Mechanism.** Free surfaces and interfaces provide faster paths; for films thinner than the grain size, the measured profile is dominated by surface diffusion, and the extracted activation energy is a mixture of surface and boundary values.
- **Discriminating observation.** The extracted activation energy should change with film thickness, approaching the surface value for the thinnest films.
- **Test.** Thickness series of diffusion couples.
- **Prior art.** Harrison-type diffusion kinetics.
- **Refuted if.** Activation energy is thickness-independent.

M31 — The measured contact angle of a liquid on a “smooth” surface is in the Cassie state without the measurement detecting it

`materials` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Wettability is reported as an intrinsic contact angle.
- **Mechanism.** Sub-micrometre roughness traps air; the apparent angle then follows Cassie–Baxter and depends on the trap fraction, which changes with immersion time as air dissolves. The angle drifts for minutes, and the value quoted is a snapshot.
- **Discriminating observation.** Contact angle should drift with time under the droplet and should change when the measurement is made in a degassed liquid.
- **Test.** Time-resolved and degassed contact-angle series.
- **Prior art.** Wetting on rough surfaces.
- **Refuted if.** Contact angle is stable over 30 min in degassed liquid.

M32 — The measured thermal conductivity of a polymer composite is dominated by the percolation state, which depends on processing and not on filler loading

`materials` · `hidden-parameter` · readiness 4 · novelty 2

- **Baseline.** κ is plotted against filler volume fraction and fitted with a mixing rule.
- **Mechanism.** Percolation depends on dispersion and on filler aspect ratio; two composites at the same loading can differ by an order of magnitude. Mixing-rule parameters therefore are not transferable and the fitted percolation threshold is a processing fingerprint.
- **Discriminating observation.** κ should correlate with measured agglomerate size rather than with loading, collapsing once dispersion is quantified.
- **Test.** Dispersion-quantified series.
- **Prior art.** Percolation in thermal composites.
- **Refuted if.** κ is a function of loading alone at fixed dispersion metric.

M33 — The measured dielectric breakdown of a polymer film is an extrinsic Weibull distribution whose scale parameter is set by the electrode area

`materials` · `boundary-condition` · readiness 5 · novelty 2

- **Baseline.** Breakdown strength is reported as a Weibull scale parameter and compared across materials.
- **Mechanism.** Breakdown is defect-initiated; the Weibull area scaling makes the reported strength a function of the test area, so films tested on small electrodes appear stronger. Comparisons across labs that differ in area are therefore not comparisons of materials.

- **Discriminating observation.** The scale parameter should follow the area-scaling law, and normalising to a fixed area should remove interlaboratory scatter.
- **Test.** Reanalysis of published breakdown data with area metadata.
- **Prior art.** Weibull area scaling in dielectrics.
- **Refuted if.** Scale parameter is area-independent.

M34 — The measured electrochemical window of a solid electrolyte is a kinetic window whose width depends on the scan rate

`materials` · `time-dependent` · readiness 5 · novelty 3

- **Baseline.** The window is from cyclic voltammetry between two blocking electrodes and quoted as a thermodynamic property.
- **Mechanism.** Decomposition currents grow with time at a fixed potential; a slow scan therefore shows a narrower window than a fast scan, and the quoted window is a scan-rate-dependent quantity that reflects the tolerable leakage current, not the band gap.
- **Discriminating observation.** The window edge should shift with scan rate in a direction and magnitude predicted by the leakage-current threshold used to define it.
- **Test.** Scan-rate series with reported current thresholds.
- **Prior art.** Electrochemical stability windows in solid electrolytes.
- **Refuted if.** The window is scan-rate independent across a decade.

M35 — The measured Seebeck coefficient of a composite is a percolation-weighted average that cannot be predicted from the constituents' Seebeck coefficients alone

`materials` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Composite thermopower is modelled by effective-medium theory with the constituents' σ and S.
- **Mechanism.** When one phase is metallic and the other semiconducting, the composite's S can exceed both constituents' values through energy filtering at the interfaces; effective-medium theory without an interface term cannot reproduce this, so reported enhancement is either real interface physics or a measurement artefact.
- **Discriminating observation.** The enhancement should scale with interface density (filler size at fixed loading) and should vanish for a composite with the same loading but coarser filler.
- **Test.** Filler-size series at fixed loading.
- **Prior art.** Energy filtering in composites.
- **Refuted if.** Enhancement is filler-size independent.

M36 — The measured ion migration activation energy in a halide perovskite depends on the electrode's Fermi-level pinning

`materials` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Activation energy is from temperature-dependent impedance or from transient current.
- **Mechanism.** The migrating ion's charge state and the associated vacancy concentration depend on the electrode's Fermi level; changing the contact changes the measured activation energy without changing the bulk migration barrier.
- **Discriminating observation.** The activation energy should differ between Ohmic and blocking contacts on the same film.
- **Test.** Same-film, two-contact comparison.
- **Prior art.** Ion migration in halide perovskites.
- **Refuted if.** Activation energy is contact-independent.

M37 — The measured radiation tolerance of a material is set by the test's dose rate through dynamic annealing

`materials` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Damage is reported as a function of dose and compared across facilities.
- **Mechanism.** Defect recombination competes with damage accumulation; at low dose rate, annealing removes defects during irradiation, so the same dose produces different damage. The dose-rate dependence is a material property that is rarely measured.
- **Discriminating observation.** Damage (from resistivity or lattice parameter) should be sublinear in dose rate at fixed dose, with a crossover that tracks the annealing activation energy.
- **Test.** Dose-rate series at fixed dose and temperature.
- **Prior art.** Dynamic annealing in irradiation damage.
- **Refuted if.** Damage is dose-rate independent at fixed dose.

M38 — The measured thermal boundary conductance of a metal–dielectric interface is dominated by the metal's electron–phonon coupling length, not by the interface

`materials` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** G is attributed to the interface and compared across interface preparations.
- **Mechanism.** Heat must first equilibrate between electrons and phonons in the metal over a length of tens of nanometres; for films thinner than that length, the measured conductance includes the electron–phonon resistance in series, and the reported G depends on metal thickness.
- **Discriminating observation.** G should increase with metal thickness and saturate at the electron–phonon length.
- **Test.** Metal-thickness series with TDTR.
- **Prior art.** Electron–phonon coupling in thermal metrology.
- **Refuted if.** G is metal-thickness independent for films below 20 nm.

M39 — The measured photostriction of a ferroelectric is mostly a thermal effect that the control experiment cannot subtract

`materials` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Photostriction is the strain under illumination minus the strain in the dark at the same temperature.
- **Mechanism.** Illumination heats the sample; if the thermometer is not on the sample or if the strain measurement has a thermal lag, the subtraction leaves a residual proportional to the heating rate. The reported photostriction then tracks absorbance.
- **Discriminating observation.** The apparent photostriction should scale with absorbed power rather than with photon energy above the gap, and should vanish when the sample is thermally clamped.
- **Test.** Wavelength- and power-dependent strain with on-sample thermometry.
- **Prior art.** Photostriction metrology.
- **Refuted if.** Photostriction is independent of absorbed power at fixed photon energy.

M40 — The measured "self-healing" efficiency of a polymer is a function of the test's waiting time

`materials` · `time-dependent` · readiness 5 · novelty 2

- **Baseline.** Healing efficiency is the recovered toughness after a stated healing period.
- **Mechanism.** Healing is diffusion-limited and continues for hours to days; a test at 24 h samples the early part of the curve, and materials with slow diffusion but high ultimate healing are ranked below fast partial healers.
- **Discriminating observation.** Efficiency should follow a t^0.5 diffusion law, and rankings should invert between 1 h and 72 h tests.
- **Test.** Time-series healing on a set of polymers.
- **Prior art.** Self-healing characterisation.
- **Refuted if.** Rankings are unchanged between 1 h and 72 h.

M41 — The measured superconducting penetration depth in a thin film includes the film's finite thickness through a geometric factor that is often neglected

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** λ is from microwave surface impedance with a thin-film correction based on the assumed λ.
- **Mechanism.** The correction is implicit (it depends on λ), so the extraction is a fixed-point problem; linearising it biases λ low by a factor that grows as thickness/λ approaches 1.
- **Discriminating observation.** λ extracted from films of two thicknesses should disagree when the linearised correction is used and agree when the fixed point is solved.
- **Test.** Thickness series with the same film quality.
- **Prior art.** Thin-film λ metrology.
- **Refuted if.** λ is thickness-independent under the linearised correction.

M42 — The measured catalytic activity of a surface is reported for a reconstruction that the reaction itself induces

`materials` · `circularity` · readiness 3 · novelty 3

- **Baseline.** Activity is measured on a characterised single-crystal surface and attributed to that surface.
- **Mechanism.** Adsorbates and reaction intermediates drive reconstruction at operating conditions; the active phase differs from the characterised one, and post-reaction characterisation (after evacuation) shows the original surface because the reconstruction is reversible.
- **Discriminating observation.** In-situ structural probes under reaction conditions should show a different surface than ex-situ; changing the reactant partial pressure should change activity and structure together.
- **Test.** In-situ XRD/STM under reaction conditions.
- **Prior art.** Pressure-gap and surface reconstruction literature.
- **Refuted if.** In-situ and ex-situ structures are identical.

M43 — The measured ionic conductivity enhancement in a nanocomposite is an artefact of the second phase's electronic leakage

`materials` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Conductivity enhancement is measured by impedance and attributed to interfacial space charge.
- **Mechanism.** The second phase can be electronically conducting at the measurement temperature, adding a parallel electronic path that is invisible in the impedance spectrum if it is frequency-independent; the block-by-electrode test that would reveal it is rarely performed.
- **Discriminating observation.** The enhancement should vanish when measured with ion-blocking electrodes or with a DC polarisation cell that separates electronic from ionic current.
- **Test.** Electronic leakage measurement via Hebb–Wagner polarisation.
- **Prior art.** Space-charge vs percolation explanations.
- **Refuted if.** Hebb–Wagner shows negligible electronic conductivity in enhanced samples.

M44 — The measured magnetic anisotropy of a thin film is dominated by the seed layer's texture

`materials` · `hidden-parameter` · readiness 4 · novelty 2

- **Baseline.** Anisotropy constants are reported for the magnetic layer.
- **Mechanism.** The seed layer sets the grain orientation and the interface roughness; anisotropy and its field dependence are therefore process properties. Comparing materials grown on different seeds conflates intrinsic anisotropy with texture.
- **Discriminating observation.** Anisotropy should change with seed-layer texture at fixed magnetic layer, in proportion to the measured orientation fraction.
- **Test.** Seed-layer series with XRD texture measurement.
- **Prior art.** Interface anisotropy in thin films.

- **Refuted if.** Anisotropy is seed-independent at fixed composition.

M45 — The measured thermal conductivity of an amorphous film is reduced by the film's surface, and the reduction scales with thickness

`materials` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Amorphous films are assumed to have a thickness-independent κ above a few nanometres.
- **Mechanism.** The top and bottom surfaces scatter the low-frequency vibrational modes that carry most of the heat; for films thinner than the dominant mean free path (~10–100 nm in amorphous materials), κ is reduced and the reduction follows a suppression function that is predicted but rarely fitted.
- **Discriminating observation.** κ should decrease with thickness below a threshold, and the suppression should be larger for films with rougher surfaces at the same thickness.
- **Test.** Thickness and roughness series.
- **Prior art.** Thermal conductivity suppression in amorphous films.
- **Refuted if.** κ is thickness-independent above 5 nm.

M46 — The measured photoluminescence blinking statistics of a quantum dot are set by the integration time, not by the emitter

`materials` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** On/off power laws are interpreted as trap-mediated charging.
- **Mechanism.** With integration time comparable to the on/off transition rate, the measured on-time distribution is convolved with the bin width; the apparent power-law exponent then depends on the binning, and comparisons across labs differ.
- **Discriminating observation.** The exponent should change with bin width in a predictable way for a given underlying exponential, and should be bin-independent only for a true power law.
- **Test.** Reanalysis of raw photon timestamps at several bin widths.
- **Prior art.** Blinking statistics methodology.
- **Refuted if.** The exponent is bin-width independent over two decades.

M47 — The measured residual stress in a coating by X-ray diffraction assumes a stress-free lattice spacing that is not measured

`materials` · `assumption` · readiness 5 · novelty 2

- **Baseline.** The $\sin^2\psi$ method uses a stress-free d-spacing from a powder or from a bulk reference.
- **Mechanism.** Composition and defect changes in the coating shift the reference spacing by more than the elastic strain being measured; the error enters as a constant offset in the d vs $\sin^2\psi$ line, which is exactly degenerate with a hydrostatic stress.
- **Discriminating observation.** The offset should change with deposition composition at fixed measured strain, and should be removable when the reference is measured from the same coating after stress relaxation.
- **Test.** Paired powder and relaxed-coating references.
- **Prior art.** $\sin^2\psi$ methodology.
- **Refuted if.** Residual stress is reference-independent.

M48 — The measured electrocaloric temperature change depends on the field's waveform, and the literature reports values at different waveforms

`materials` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** ΔT_{EC} is quoted for a field change without specifying the waveform.
- **Mechanism.** Hysteresis and Joule heating make the measured ΔT depend on frequency and on the duty cycle; the adiabatic value is approached only for fast ramps, and slower ramps lose heat to the surroundings. The quoted ΔT is therefore a measurement-protocol quantity.

- **Discriminating observation.** ΔT should increase with ramp rate and saturate at the adiabatic limit, with the saturation rate set by the sample's thermal diffusivity.
- **Test.** Waveform-series measurement on one sample.
- **Prior art.** Electrocaloric measurement methods.
- **Refuted if.** ΔT is waveform-independent.

M49 — The measured carrier lifetime in a semiconductor is limited by surface recombination whose velocity changes with the passivation's aging

`materials` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Bulk lifetime is extracted from a decay with a surface-recombination correction using a fixed velocity.
- **Mechanism.** The passivation degrades over hours under illumination; the surface velocity rises, and the extracted bulk lifetime falls spuriously. Reported lifetimes therefore depend on how long after passivation the measurement was made.
- **Discriminating observation.** The extracted lifetime should decay with illumination time at a rate that tracks the measured surface recombination velocity on a companion wafer.
- **Test.** Lifetime series with parallel surface-velocity monitors.
- **Prior art.** Surface passivation in lifetime metrology.
- **Refuted if.** Extracted bulk lifetime is stable over 24 h under illumination.

M50 — The measured figure of merit of a thermoelectric module and of its material differ because the module's contact resistance is assigned to the material

`materials` · `boundary-condition` · readiness 5 · novelty 3

- **Baseline.** Module efficiency and material zT are compared directly.
- **Mechanism.** The module's internal resistance includes contacts and current-spreading; the module-derived effective zT is lower by a factor set by the contact fraction, which is not reported. The gap is then attributed to a materials problem, and optimisation targets the wrong term.
- **Discriminating observation.** The module/material gap should correlate with independently measured contact resistance across modules.
- **Test.** Reanalysis of module and material data pairs.
- **Prior art.** Thermoelectric module metrology.
- **Refuted if.** The gap is uncorrelated with contact resistance.

Biology and medicine — 50 hypotheses

Every entry is a proposition. No experiment below has been performed.

B01 — Optical-density growth curves saturate before the reader says they do, and every fitted growth rate is compressed by the same detector nonlinearity

`microbiology` · `nonlinearity` · readiness 5 · novelty 3

- **Baseline.** Specific growth rate μ is fitted to the exponential phase of log(OD_{600}) vs time, trusting the instrument's linear range.
- **Mechanism.** Detector saturation and stray light compress OD at high values. A stray-light calculation (R1, section A) shows the compression is only 0.1% at OD 0.2–1.0 for a 0.1% stray fraction and reaches 8% only at OD 2.0–3.0, so a sub-exponential fit in the OD 0.3–0.6 range is **not** explained by this mechanism; the surviving candidates are detector nonlinearity, meniscus and fill volume, evaporation, and a genuine physiological shift.
- **Discriminating observation.** μ fitted from OD should decrease with fill volume and with the reader's gain setting, while μ from viable counts or from CO_2 evolution should not.
- **Test.** Fill-volume and gain series on one strain with parallel viable counts.
- **Prior art.** Plate-reader linearity specifications; growth-curve methodology.
- **Refuted if.** OD-derived μ is invariant to fill volume and gain across the linear range.

B02 — Quantitative PCR efficiencies are estimated from a dilution series that is itself biased by the dilution's pipetting error

`molecular-biology` · `circularity` · readiness 4 · novelty 3

- **Baseline.** Amplification efficiency comes from the slope of Cq vs log(dilution).
- **Mechanism.** Serial dilution accumulates pipetting error multiplicatively; because Cq is linear in log(copy number), the error inflates the slope's variance and biases the efficiency toward 100% (regression dilution). The quoted efficiency therefore reflects the dilution protocol.
- **Discriminating observation.** Efficiency estimated from independently prepared (not serial) dilutions should differ systematically from the serial estimate, with the difference scaling as the pipetting CV.
- **Test.** Serial vs independent dilution series on the same assay.
- **Prior art.** qPCR efficiency estimation.
- **Refuted if.** Serial and independent estimates agree within 1%.

B03 — Western blot band intensities are in the saturated regime more often than reported, and the saturation is antibody-dependent

`molecular-biology` · `nonlinearity` · readiness 5 · novelty 2

- **Baseline.** Band intensity is converted to protein abundance by linear calibration with a loading series.
- **Mechanism.** Chemiluminescence saturates at the membrane and at the camera; the onset depends on antibody affinity and on substrate depletion, so a "linear" loading series can be linear only because both axes are compressed. Ratios between bands then distort in a direction set by which band is brighter.
- **Discriminating observation.** The measured ratio should change when the exposure is reduced and when the primary antibody is diluted, converging only in the unsaturated regime.
- **Test.** Exposure and antibody-titration series on the same blot.
- **Prior art.** Western blot quantification guidelines.
- **Refuted if.** Ratios are invariant to exposure and antibody dilution.

B04 — Single-cell RNA-seq "dropout" is partly a capture-rate artefact that depends on cell size, so pseudotime trajectories are size-biased

`transcriptomics` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Zero counts are modelled as dropout and imputed with a zero-inflated model that is independent of cell size.
- **Mechanism.** Capture efficiency scales with the cell's RNA content and with its lysis behaviour; small cells have systematically more zeros. Imputation then fills in genes that are absent because of size, and trajectory inference orders cells by size rather than by state.
- **Discriminating observation.** The first principal component of imputed data should correlate with an independent size proxy (forward scatter, cell diameter) more strongly than with marker-gene expression.
- **Test.** Reanalysis of public datasets with size metadata.
- **Prior art.** Dropout imputation methods.
- **Refuted if.** Leading components are uncorrelated with size proxies after size correction.

B05 — Cryo-EM particle picking is confirmation-biased, and the bias grows with the strength of the prior

`structural-biology` · `selection` · readiness 4 · novelty 3

- **Baseline.** Template or network picking is assumed to select particles without altering the reconstructed density.
- **Mechanism.** Picking preferentially accepts images that already resemble the template; the selected subset is enriched in template-like views, which sharpens features in the map and suppresses features absent from the template. The effect is a self-fulfilling reconstruction.
- **Discriminating observation.** Reconstructions from independently picked subsets (different templates, or picked before and after 3D classification) should differ in the density of features that were in the template versus features that were not.
- **Test.** Reanalysis with template-swap controls.
- **Prior art.** arXiv:2507.03951 quantifies the picking bias.
- **Refuted if.** Template-swapped reconstructions are identical at the reported resolution.

B06 — Calcium imaging's inferred spike rate depends on the indicator's binding kinetics in a way that mimics network state changes

`neuroscience` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** ΔF/F is deconvolved with a fixed indicator time constant to infer spikes.
- **Mechanism.** The effective decay time depends on the indicator's affinity, on the resting calcium, and on the cell's buffering capacity; cells with different expression levels therefore have different effective kernels. A deconvolution with one kernel assigns apparent rate differences that are expression differences.
- **Discriminating observation.** Inferred rate should correlate with measured indicator expression level at fixed stimulus, and should collapse when the kernel is estimated per cell.
- **Test.** Reanalysis of datasets with per-cell brightness metadata.
- **Prior art.** Calcium deconvolution methods.
- **Refuted if.** Inferred rates are independent of expression level.

B07 — The apparent allometric exponent of metabolic rate depends on the measurement temperature's deviation from the species' preferred range

`ecology` · `model-artifact` · readiness 3 · novelty 3

- **Baseline.** Metabolic scaling exponents are fitted across body sizes at a common measurement temperature.
- **Mechanism.** The temperature sensitivity (Ea) itself scales with body size, so measuring all species at one temperature applies a size-dependent correction that is absorbed into the mass exponent. The reported 3/4 exponent is then partly a temperature-sampling artefact.
- **Discriminating observation.** The fitted mass exponent should change with measurement temperature, and should approach the within-species exponent when measured at each species' preferred temperature.
- **Test.** Reanalysis of respirometry compilations with temperature metadata.
- **Prior art.** Metabolic theory of ecology.

- **Refuted if.** The mass exponent is temperature-independent.

B08 — Patch-clamp series-resistance compensation distorts the very kinetics it is used to measure

`electrophysiology` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Recordings are accepted when series resistance is <10 MΩ and 70–80% compensated.
- **Mechanism.** The residual series resistance forms a low-pass filter whose corner scales with the cell's capacitance; compensation itself adds a differentiated component that can precede the true onset. Fast activation kinetics are therefore systematically slowed and their voltage dependence distorted.
- **Discriminating observation.** The fitted activation time constant should depend on the uncompensated resistance across cells at fixed channel, in the direction predicted by the filter, and should converge when extrapolated to zero residual resistance.
- **Test.** Reanalysis of recordings with reported Rs and compensation.
- **Prior art.** Series-resistance artefacts in voltage clamp.
- **Refuted if.** Activation time constants are independent of residual Rs.

B09 — Flow-cytometry compensation matrices are fitted on single-stained controls whose autofluorescence differs from the experimental sample's

`cytometry` · `calibration` · readiness 4 · novelty 3

- **Baseline.** The compensation matrix is computed from single-stained cells and applied to all samples.
- **Mechanism.** Autofluorescence is a function of cell size, granularity, and fixation; a compensation computed on one autofluorescence spectrum over- or under-subtracts in another population, producing spurious double-positive cells that track the sample's autofluorescence.
- **Discriminating observation.** The double-positive fraction should correlate with the sample's measured autofluorescence, and should vanish when compensation is computed per sample.
- **Test.** Reanalysis with per-sample autofluorescence controls.
- **Prior art.** Compensation methodology.
- **Refuted if.** Double-positive fractions are autofluorescence-independent.

B10 — The measured maximum oxygen consumption rate in respirometry is limited by oxygen diffusion into the chamber, not by the tissue

`bioenergetics` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** OCR is computed from the fluorescence decay with the chamber's oxygen solubility and volume.
- **Mechanism.** At high rates the chamber's oxygen concentration falls toward the diffusion-limited regime, and the stirring rate sets the supply. Reported maxima therefore depend on stirring and on chamber geometry, and comparisons across instruments are not comparisons of mitochondria.
- **Discriminating observation.** OCR_max should increase with stirring rate and saturate, with the saturation depending on the sample's size.
- **Test.** Stirring series on the same sample.
- **Prior art.** Respirometry protocols.
- **Refuted if.** OCR_max is stirring-independent.

B11 — The inferred selectivity of an antimicrobial peptide depends on the assay's ionic strength through a membrane-binding term

`microbiology` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Minimum inhibitory concentrations are compared across peptides in standard media.
- **Mechanism.** Cationic peptide binding to anionic membranes is screened by salt; media with different NaCl content change the bound fraction by an order of magnitude, so MIC comparisons across studies conflate peptide sequence with medium.

- **Discriminating observation.** MIC should shift with medium ionic strength with a slope that is proportional to the peptide's net charge.
- **Test.** Ionic-strength series on a peptide panel.
- **Prior art.** Salt sensitivity of AMPs.
- **Refuted if.** MIC is ionic-strength independent.

B12 — The apparent heterogeneity of a bacterial population is an artefact of the plate reader's well-to-well thermal gradient

`microbiology` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Well-to-well variation in growth is interpreted as phenotypic heterogeneity or as a subtle genetic effect.
- **Mechanism.** Plate readers have a thermal gradient across the plate of 0.2–1 K, which produces growth-rate differences of several percent; the gradient is fixed in space, so replicate plates reproduce the same "heterogeneity" pattern.
- **Discriminating observation.** The pattern should rotate with the plate orientation and disappear when wells are randomised or when the plate is thermally equilibrated.
- **Test.** Rotated-plate and randomised-layout experiments.
- **Prior art.** Plate-reader edge effects.
- **Refuted if.** Well patterns are invariant to plate rotation.

B13 — The measured protein half-life from cycloheximide chase includes a translation-independent decay that starts before the chase

`cell-biology` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Half-life is the decay constant after translation inhibition.
- **Mechanism.** Cycloheximide itself perturbs the cell (stress response, altered degradation) within minutes, and the chase's t=0 is defined after the perturbation; the fitted decay therefore includes a fast stress component and reports a half-life shorter than the physiological one.
- **Discriminating observation.** Half-lives should differ between cycloheximide and auxin-inducible or temperature-sensitive degron systems for the same protein, with the difference scaling with the stress-marker induction.
- **Test.** Side-by-side chase methods on the same protein.
- **Prior art.** Comparison of translation-inhibition methods.
- **Refuted if.** Half-lives agree across methods within error.

B14 — The inferred microbiome composition from 16S sequencing is set by the extraction kit's lysis bias, which correlates with the phenotype being studied

`microbiome` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Relative abundances are compared across sample groups processed with the same kit.
- **Mechanism.** Lysis efficiency differs by cell wall type; a phenotype that alters the community's Gram-positive fraction changes the bias in the same direction, so the kit bias and the biological effect are confounded and cannot be separated by replication.
- **Discriminating observation.** Re-extracting the same samples with a different kit should reverse or attenuate the group difference, while a genuine effect should persist.
- **Test.** Split-sample extraction with two kits.
- **Prior art.** Extraction bias in microbiome studies.
- **Refuted if.** Group differences are kit-independent.

B15 — The measured enzyme velocity in a plate assay includes a substrate depletion term that is invisible at the standard time point

`biochemistry` · `time-dependent` · readiness 5 · novelty 2

- **Baseline.** Initial velocity is taken from a single time point or from a short linear fit.
- **Mechanism.** At high enzyme concentration or low K_M, substrate depletion exceeds 10% before the first read; the "initial" velocity is then a time-averaged rate that depends on the substrate concentration and underestimates V_max.
- **Discriminating observation.** v should decrease with the time point in the direction predicted by the integrated Michaelis–Menten equation, and the apparent K_M should shift with enzyme concentration.
- **Test.** Time-series assay at several enzyme concentrations.
- **Prior art.** Integrated Michaelis–Menten analysis.
- **Refuted if.** Apparent K_M is enzyme-concentration independent.

B16 — The measured binding affinity from surface plasmon resonance is limited by mass transport, and the limitation is hidden by a good global fit

`biophysics` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** k_on and k_off are from a global fit to a 1:1 model, with the fit quality used as validation.
- **Mechanism.** Mass transport limitation produces a response that is well fitted by a 1:1 model with a reduced k_on, because both effects saturate; the fit residual is small while k_on is underestimated by up to 10×.
- **Discriminating observation.** k_on should increase with flow rate and saturate, and should approach the true value when the ligand density is reduced.
- **Test.** Flow-rate and ligand-density series.
- **Prior art.** Mass-transport limitation in SPR.
- **Refuted if.** k_on is flow-rate and density independent.

B17 — The measured membrane potential from a fluorescent dye is a weighted average dominated by the dye's partitioning, not by the voltage

`cell-biology` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Fluorescence ratio is converted to millivolts with a Nernstian calibration.
- **Mechanism.** The dye distributes among plasma membrane, organelles, and cytosol; the organelle fraction changes with cell type and with dye loading, and it dominates the signal. The calibration in one cell type does not transfer.
- **Discriminating observation.** The inferred potential should change with dye loading time and with the cell's organelle volume fraction, and should disagree with patch-clamp in cells with large organelle fractions.
- **Test.** Paired imaging and patch clamp across cell types.
- **Prior art.** Voltage-sensitive dye calibration.
- **Refuted if.** Dye-inferred potential matches patch clamp across organelle fractions.

B18 — The measured reactive-oxygen signal is dominated by the probe's own photo-generated species

`cell-biology` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** DCF fluorescence is read as a proxy for intracellular H_2O_2.
- **Mechanism.** DCF is oxidised by multiple species and is itself photo-oxidised; the signal therefore grows with illumination time and with light dose, and the apparent response to a stimulus includes a light-dose term.
- **Discriminating observation.** The signal should increase with illumination intensity at fixed stimulus, and should be unchanged in cells lacking the putative source.
- **Test.** Light-dose series with genetic controls.
- **Prior art.** DCF specificity and photo-oxidation.

- **Refuted if.** Signal is light-dose independent at fixed stimulus.

B19 — The measured ATP level from a luciferase assay depends on the lysis time because ATP is consumed during extraction

`biochemistry` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** ATP is extracted and immediately read, with a standard lysis protocol.
- **Mechanism.** ATPases remain active during lysis; the extraction time therefore sets the loss, and protocols that differ by a minute differ by tens of percent. Reported ATP levels track protocol timing.
- **Discriminating observation.** Measured ATP should decay with lysis duration at a rate equal to the residual ATPase activity, and should be stable with a validated quench.
- **Test.** Lysis-duration series with and without validated quench.
- **Prior art.** ATP assay protocols.
- **Refuted if.** Measured ATP is lysis-duration independent.

B20 — The measured cytotoxicity of a nanoparticle includes an interference term between the particle and the assay dye

`toxicology` · `model-artifact` · readiness 5 · novelty 2

- **Baseline.** Viability is from MTT/MTS conversion, with particle-only controls.
- **Mechanism.** Some particles catalyse or quench the dye's reduction, and the effect depends on the particle's surface chemistry and on the medium's protein corona. The particle-only control lacks cells, so it does not reproduce the intracellular reduction environment.
- **Discriminating observation.** Apparent viability should differ between MTT and a membrane-integrity assay using the same particles at the same dose.
- **Test.** Paired dye and membrane-integrity assays.
- **Prior art.** Nanoparticle assay interference.
- **Refuted if.** Two orthogonal assays rank particles identically.

B21 — The measured wound-healing rate in a scratch assay is dominated by cell proliferation at the edges, not by migration

`cell-biology` · `model-disagreement` · readiness 4 · novelty 2

- **Baseline.** Gap closure rate is interpreted as migration.
- **Mechanism.** Mitomycin treatment to suppress proliferation also alters migration; without it, the closure combines both. The apparent migration rate therefore depends on the proliferation rate, which varies with passage number and confluence.
- **Discriminating observation.** Closure rate should correlate with the EdU incorporation rate across conditions, and should change when proliferation is suppressed by a method that does not affect motility.
- **Test.** Closure series with paired proliferation labels and two suppression methods.
- **Prior art.** Scratch assay methodology.
- **Refuted if.** Closure rate is uncorrelated with proliferation.

B22 — The measured affinity of an antibody in ELISA is limited by the hook effect, which suppresses signal at high concentration without warning

`immunoassay` · `nonlinearity` · readiness 5 · novelty 2

- **Baseline.** Signal is monotonic in analyte concentration over the calibration range.
- **Mechanism.** Above a threshold, both capture and detection antibodies saturate and the sandwich cannot form; the signal falls while the true concentration rises, and a diluted sample can read higher than an undiluted one.
- **Discriminating observation.** Signal should be non-monotonic in concentration, and the apparent concentration should decrease with a smaller dilution factor in a way fixed by the antibody K_D.

- **Test.** Dilution-linearity series on high-concentration samples.
- **Prior art.** Hook effect in sandwich immunoassays.
- **Refuted if.** Signal is monotonic across the full tested range.

B23 — The measured transcription rate from nascent-RNA sequencing depends on the labelling time relative to the elongation rate

`transcriptomics` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Nascent transcript density is converted to initiation rate with a fixed elongation rate.
- **Mechanism.** The elongation rate varies along the gene and with the cell state; a labelling pulse averages over positions, so the inferred initiation rate is a convolution. Comparisons across conditions with different elongation rates are therefore confounded.
- **Discriminating observation.** The inferred initiation rate should change when the pulse time is varied for genes with known long elongation times, and should be stable for short genes.
- **Test.** Pulse-time series with elongation-rate measurements.
- **Prior art.** Nascent RNA kinetics.
- **Refuted if.** Inferred initiation rates are pulse-time independent.

B24 — The measured microbiota-mediated drug metabolism depends on the oxygen gradient in the culture, which is not controlled

`microbiome` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Faecal or culture metabolism is measured in anaerobic batches and attributed to specific taxa.
- **Mechanism.** Metabolites and redox state depend on the local oxygen tension; static cultures develop gradients, so the same community metabolises differently depending on vessel geometry and fill volume. Reported activities are therefore vessel-specific.
- **Discriminating observation.** Metabolic rate per cell should scale with the surface-to-volume ratio of the vessel and should converge in continuous culture with controlled redox.
- **Test.** Vessel-geometry series with redox monitoring.
- **Prior art.** Gut-microbiome culturing.
- **Refuted if.** Activity per cell is vessel-geometry independent.

B25 — The measured circadian period from bioluminescence recordings is biased by the reporter's maturation time

`chronobiology` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Period is from the peak-to-peak interval of a luciferase rhythm.
- **Mechanism.** Luciferase maturation and degradation add a low-pass filter; the filter delays the peak by an amount that depends on the reporter's half-life, and a slow reporter shifts the peak without changing the period — but if the filter is asymmetric it also lengthens the apparent period.
- **Discriminating observation.** Period should differ between reporters with different half-lives in the same cells, in the direction predicted by the filter.
- **Test.** Same cells, two reporters of different stability.
- **Prior art.** Reporter dynamics in circadian recordings.
- **Refuted if.** Period is reporter-independent.

B26 — The measured optogenetic activation depth profile assumes a scattering coefficient that is not measured in the tissue being stimulated

`neuroscience` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Light intensity at depth is computed from a generic scattering model for brain tissue.

- **Mechanism.** Scattering and absorption vary with myelination, age, and fibre placement; a twofold error in the effective attenuation length changes the activated volume by nearly an order of magnitude. Reported thresholds are therefore apparatus-specific.
- **Discriminating observation.** The activation threshold should vary with fibre insertion depth and with age in a way predicted by measured attenuation, and should be consistent across labs only when attenuation is measured in situ.
- **Test.** In-situ attenuation measurement with threshold mapping.
- **Prior art.** Light delivery in optogenetics.
- **Refuted if.** Thresholds are invariant to fibre depth and age.

B27 — The measured synaptic plasticity magnitude depends on the recording electrode's position relative to the dendritic tree

`neuroscience` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Field potentials are interpreted as population measures of synaptic strength.
- **Mechanism.** The field potential is a weighted sum of dendritic currents with a spatial kernel set by the tissue conductivity and by the electrode position; a small displacement changes the weighting and hence the apparent plasticity magnitude.
- **Discriminating observation.** The measured plasticity should change with electrode depth in a way predicted by the cable model, and should be reproducible only when depth is controlled.
- **Test.** Depth-controlled recording series.
- **Prior art.** Field-potential source localisation.
- **Refuted if.** Plasticity magnitude is depth-independent within a cortical layer.

B28 — The measured bacterial mutation rate from fluctuation tests depends on the plating delay

`genetics` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Mutation rate is from the Luria–Delbrück distribution of mutant colonies.
- **Mechanism.** Mutants continue to divide during the phenotypic-expression delay and after plating; a longer delay increases the counted mutants and shifts the inferred rate. Protocols differ in delay without reporting it.
- **Discriminating observation.** The inferred rate should increase with delay until the mutant progeny saturate, and should converge when the delay is fixed and reported.
- **Test.** Delay series on a standard strain.
- **Prior art.** Fluctuation-test methodology.
- **Refuted if.** Inferred rate is delay-independent.

B29 — The measured cell-cycle phase distribution from DNA content is distorted by the fixation's chromatin compaction

`cytometry` · `model-artifact` · readiness 4 · novelty 2

- **Baseline.** PI staining gives DNA content and the fractions are from a standard cell-cycle model.
- **Mechanism.** Fixation and permeabilisation change dye accessibility; condensed mitotic chromatin stains less, so mitotic cells are misassigned and the G2/M fraction is underestimated in samples with a high mitotic index.
- **Discriminating observation.** The G2/M fraction should change with fixation protocol at fixed biology, and should agree with a mitotic marker when fixation is matched.
- **Test.** Fixation series with a mitotic marker.
- **Prior art.** Cell-cycle analysis by flow cytometry.
- **Refuted if.** Phase fractions are fixation-independent.

B30 — The measured drug potency in a 2D monolayer does not transfer to 3D because the readout's dynamic range changes, not the drug

`pharmacology` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** IC_{50} from 2D viability is compared with 3D spheroid response.
- **Mechanism.** The viability assay's dynamic range depends on the aggregate's size and on dye penetration; in 3D the ATP readout is dominated by the outer layer, so the IC_{50} shift is partly a readout artefact.
- **Discriminating observation.** The 2D/3D IC_{50} ratio should change when the readout is changed to one that lyses the whole spheroid, or when the spheroid is dissociated before assay.
- **Test.** Paired readouts on the same spheroids.
- **Prior art.** 3D drug-response assays.
- **Refuted if.** The ratio is readout-independent.

B31 — The measured protein oligomerisation state from size-exclusion chromatography is set by the column's residence time

`biochemistry` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Elution volume gives molecular weight and the oligomeric state is inferred.
- **Mechanism.** Weak oligomers equilibrate during the run; a long column allows dissociation, so the measured state depends on flow rate. Reported states therefore reflect the column, not the solution.
- **Discriminating observation.** The apparent molecular weight should decrease with lower flow rate for a weakly associating system, and should be invariant for a covalent oligomer.
- **Test.** Flow-rate series on a known weak oligomer.
- **Prior art.** SEC for oligomeric state.
- **Refuted if.** Apparent molecular weight is flow-rate independent.

B32 — The measured antibiotic tolerance of a biofilm is dominated by the assay's washing step

`microbiology` · `model-artifact` · readiness 4 · novelty 2

- **Baseline.** Tolerance is the surviving fraction after antibiotic exposure and wash.
- **Mechanism.** Washing disrupts the biofilm and releases cells that then grow in the recovery medium; the recovered fraction depends on the shear during washing, which is not standardised. Tolerance values are therefore method-defined.
- **Discriminating observation.** Survival should decrease with wash shear and should converge with a validated, low-shear recovery.
- **Test.** Wash-shear series with confocal quantification.
- **Prior art.** Biofilm tolerance assays.
- **Refuted if.** Survival is wash-shear independent.

B33 — The measured gut transit time from dye markers is biased by the marker's own effect on water absorption

`physiology` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Transit time is the interval to marker appearance.
- **Mechanism.** Non-absorbable markers can osmotically retain water and alter motility; the marker dose therefore changes the quantity measured. Comparisons across studies with different doses conflate marker load with physiology.
- **Discriminating observation.** Transit time should depend on marker dose and osmolality, and should agree with an independent (e.g. breath-test) measure only at low dose.
- **Test.** Dose series with an orthogonal transit measure.
- **Prior art.** Transit-time measurement.
- **Refuted if.** Transit time is dose-independent.

B34 — The measured immune-cell cytokine release depends on the plate's surface treatment, which varies between manufacturers

`immunology` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Cytokine concentrations are compared across plates and labs.
- **Mechanism.** Surface wettability and adsorbed protein layer determine monocyte activation; untreated or differently treated polystyrene changes background cytokine release by a factor that is comparable to the effect being studied.
- **Discriminating observation.** Background release should vary between plate brands at fixed cells and medium, and should correlate with the measured contact angle.
- **Test.** Plate-brand comparison with contact-angle measurement.
- **Prior art.** Cell-culture surface effects.
- **Refuted if.** Background release is brand-independent.

B35 — The measured longevity effect of a dietary intervention is confounded by the control diet's composition changing between batches

`geroscience` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Lifespan is compared between an intervention diet and a control from the same supplier.
- **Mechanism.** Control diet formulations change between batches (ash content, mycotoxins, fatty-acid profile), and lifespan is sensitive to these; a between-batch change can produce an effect comparable to the intervention.
- **Discriminating observation.** Lifespan should vary between control batches at the same nominal formulation, and the intervention effect should be re-estimated with batch as a random effect.
- **Test.** Reanalysis of multi-batch cohorts.
- **Prior art.** Diet variability in rodent studies.
- **Refuted if.** Control-batch variance is negligible relative to the intervention effect.

B36 — The measured photosynthetic rate from chlorophyll fluorescence is nonlinear in the light intensity used to probe it

`plant-physiology` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Fv/Fm and electron-transport rates are from saturating pulses assumed to fully close reaction centres.
- **Mechanism.** The saturating pulse is not saturating for sun-adapted leaves or for algae with large antennae; the residual open centres make the inferred quantum yield depend on pulse intensity. Reported stress responses therefore partly track the instrument's pulse.
- **Discriminating observation.** The inferred yield should increase with pulse intensity and plateau; the plateau intensity should differ between sun- and shade-adapted material.
- **Test.** Pulse-intensity series on both acclimation types.
- **Prior art.** PAM fluorometry methodology.
- **Refuted if.** Yield is pulse-intensity independent for sun-adapted leaves.

B37 — The measured species richness from eDNA is bounded by the primer's mismatch tolerance, which is biased against the taxa of interest

`ecology` · `blind-spot` · readiness 4 · novelty 3

- **Baseline.** Richness is from amplicon sequencing with a standard primer pair.
- **Mechanism.** Primer mismatches reduce amplification by orders of magnitude for some lineages; rare taxa with mismatches are absent not because they are rare but because they do not amplify. Richness is therefore a measure of primer compatibility.
- **Discriminating observation.** Richness should increase when a second primer set with different mismatches is used, and the increase should be concentrated in lineages with predicted mismatches.
- **Test.** Multi-primer comparison on the same eDNA.

- **Prior art.** Primer bias in metabarcoding.
- **Refuted if.** Richness is primer-independent.

B38 — The measured thermal tolerance of an ectotherm depends on the acclimation rate, not only on the acclimation temperature

`ecology` · `time-dependent` · readiness 3 · novelty 3

- **Baseline.** Thermal tolerance is measured after acclimation to a set temperature for a set period.
- **Mechanism.** Hardening and heat-shock protein induction depend on the rate of temperature change; slow and fast acclimation to the same endpoint give different tolerances. Protocols differ in ramp rate without reporting it.
- **Discriminating observation.** Tolerance should depend on acclimation ramp rate at fixed endpoint temperature, converging only for slow ramps.
- **Test.** Ramp-rate series on one species.
- **Prior art.** Thermal acclimation protocols.
- **Refuted if.** Tolerance is ramp-rate independent.

B39 — The measured association between a biomarker and a disease is attenuated by the assay's limit of detection in a way that depends on the case mix

`epidemiology` · `selection` · readiness 4 · novelty 3

- **Baseline.** Odds ratios are estimated from measured concentrations with values below the limit of detection substituted or excluded.
- **Mechanism.** Substitution (LOD/2 or LOD/√2) biases the exposure distribution, and the bias direction depends on the proportion below LOD, which differs between cases and controls. The measured association is therefore attenuated or inflated by the case mix rather than by biology.
- **Discriminating observation.** The odds ratio should change when a censored-data (Tobit) model is used, and the change should scale with the below-LOD fraction difference.
- **Test.** Reanalysis of published biomarker studies with reported LOD fractions.
- **Prior art.** Limit-of-detection handling in epidemiology.
- **Refuted if.** Odds ratios are unchanged under censored-data modelling.

B40 — The measured gut-brain effect of a probiotic is confounded by the capsule's excipient, which is fermentable

`microbiome` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Probiotic effects are compared against a placebo capsule.
- **Mechanism.** Many placebo capsules contain fermentable excipients (maltodextrin, inulin) that alter the microbiota; the "placebo" is therefore an active comparator, and the true effect of the organism is the difference between two active interventions.
- **Discriminating observation.** The placebo arm's microbiota should shift from baseline, and the probiotic effect size should shrink when the placebo is changed to a non-fermentable excipient.
- **Test.** Reanalysis of trials reporting placebo-arm compositional change.
- **Prior art.** Excipient effects in probiotic trials.
- **Refuted if.** Placebo arms show no compositional change.

B41 — The measured protein–protein interaction affinity from yeast two-hybrid is a function of the reporter's dynamic range, not of the affinity

`molecular-biology` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Interaction strength is ranked by colony growth or by lacZ output.

- **Mechanism.** The reporter saturates at modest interaction strength; above saturation all interactions look equal, and below the detection limit all look absent. The assay's useful range is therefore one to two orders of magnitude, and rank order outside it is arbitrary.
- **Discriminating observation.** Rank order should change when the reporter's copy number or the selection stringency is changed, while a true affinity ranking should be invariant.
- **Test.** Reanalysis across stringency conditions.
- **Prior art.** Y2H quantitative limitations.
- **Refuted if.** Rank order is stringency-invariant.

B42 — The measured membrane permeability from Caco-2 assays depends on the insert's unstirred water layer

`pharmacology` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Apparent permeability P_app is computed from flux and area.
- **Mechanism.** The unstirred water layer adds a series resistance that is comparable to the monolayer's for lipophilic compounds; shaking rate and insert geometry therefore change P_app by a factor that is compound-dependent, and in-vitro/in-vivo correlations inherit the artefact.
- **Discriminating observation.** P_app should increase with shaking rate for lipophilic compounds and be unchanged for hydrophilic ones, converging to a plateau.
- **Test.** Shaking-rate series across a compound panel.
- **Prior art.** UWL effects in Caco-2.
- **Refuted if.** P_app is shaking-rate independent for lipophilic compounds.

B43 — The measured chromatin accessibility difference between conditions is driven by the sequencing depth's interaction with peak calling

`epigenomics` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Differential accessibility is called from peak counts normalised by library size.
- **Mechanism.** Peak calling is depth-dependent: low-depth libraries yield fewer, broader peaks, so the feature set differs between conditions. Differences then arise from the feature set rather than from accessibility.
- **Discriminating observation.** Reanalysis with a union peak set fixed before quantification should reduce the number of differential regions, and the reduction should scale with the depth difference.
- **Test.** Reanalysis of public ATAC-seq pairs with depth imbalance.
- **Prior art.** Differential accessibility methodology.
- **Refuted if.** Differential regions are unchanged by using a fixed union peak set.

B44 — The measured developmental timing of an organism is shifted by the temperature of the imaging stage

`developmental-biology` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Developmental stage is timed from a reference and compared across genotypes.
- **Mechanism.** Light-sheet and confocal imaging deposit heat; the sample's temperature can rise by 1–3 K, accelerating development by 10–20%. Genotypes imaged with different laser powers are therefore developmentally misaligned.
- **Discriminating observation.** Developmental rate should correlate with the imaging laser power at fixed genotype, and should match the plate-reader rate when the stage temperature is measured and controlled.
- **Test.** Laser-power series with on-sample thermometry.
- **Prior art.** Phototoxicity and heating in live imaging.
- **Refuted if.** Developmental rate is laser-power independent.

B45 — The measured mechanical stiffness of a tissue from AFM is dominated by the probe's geometry at the scale of the cell

`biophysics` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Young's modulus is from a Hertz fit to an indentation curve.
- **Mechanism.** At indentations comparable to the cortical actin layer, the response is not Hertzian; the fitted modulus depends on the contact radius and on the indentation depth. Reported stiffnesses for the same tissue can differ by a factor of three depending on probe size.
- **Discriminating observation.** Modulus should depend on probe radius and indentation depth in a way predicted by a layered model, and should converge for deep indentations with a large probe.
- **Test.** Multi-probe measurement on the same tissue.
- **Prior art.** AFM rheology of cells.
- **Refuted if.** Modulus is probe-radius independent.

B46 — The measured insect flight metabolic rate from CO_2 respirometry is biased by the chamber's washout time

`physiology` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Metabolic rate is computed from steady-state CO_2 concentration and flow.
- **Mechanism.** The chamber's washout time constant is comparable to the duration of flight bouts; the measured concentration is a convolution, and short bouts are underestimated. Reported peak rates depend on flow rate.
- **Discriminating observation.** Peak metabolic rate should increase with flow rate and saturate, with the saturation set by the washout constant.
- **Test.** Flow-rate series on the same animals.
- **Prior art.** Respirometry washout corrections.
- **Refuted if.** Peak rate is flow-rate independent.

B47 — The measured evolutionary rate from serial-passage experiments includes a selection term from the passage bottleneck

`evolution` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Mutation accumulation rates are from endpoint sequencing after serial passage.
- **Mechanism.** The bottleneck size sets the strength of drift and the efficacy of selection; small bottlenecks allow deleterious mutations to accumulate, inflating the inferred mutation rate. Passaging protocols differ in bottleneck without reporting it.
- **Discriminating observation.** The inferred rate should depend on bottleneck size and should converge to the neutral rate only for sufficiently small bottlenecks or with a selection-aware model.
- **Test.** Bottleneck series on the same organism.
- **Prior art.** Mutation-accumulation experiment design.
- **Refuted if.** Inferred rate is bottleneck-independent.

B48 — The measured vaccine immunogenicity in mice is confounded by the microbiome's batch-dependent adjuvant effect

`immunology` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Antibody titres are compared between vaccine formulations in mice from one facility.
- **Mechanism.** The commensal microbiota provides adjuvant signals that vary between vendors and between cages; titre differences between formulations can be smaller than the between-vendor variance.
- **Discriminating observation.** Titres should vary between vendors at fixed formulation, and the formulation ranking should be stable only within a vendor or after microbiota standardisation.
- **Test.** Reanalysis of multi-vendor studies.
- **Prior art.** Microbiota effects on vaccine responses.

- **Refuted if.** Between-vendor variance is negligible.

B49 — The measured protein turnover from heavy-water labelling is biased by the precursor pool's non-uniform enrichment

`proteomics` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Turnover is from the exponential rise of labelled peptide fraction, assuming the precursor pool is at the administered enrichment.
- **Mechanism.** Amino-acid precursor enrichment differs between tissues and between dietary states and is diluted by endogenous synthesis; using the administered value overestimates the precursor and therefore overestimates the turnover rate, with a tissue-dependent bias.
- **Discriminating observation.** The inferred rate should change when the measured tissue-specific precursor enrichment is used, and the change should be largest in tissues with high endogenous synthesis.
- **Test.** Reanalysis with measured precursor enrichment.
- **Prior art.** Dynamic proteomics with heavy water.
- **Refuted if.** Inferred rates are unchanged by precursor correction.

B50 — The measured efficacy of an antifungal in vivo is dominated by the compound's binding to serum proteins, which varies between species

`pharmacology` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Free-drug exposure is assumed to scale with total concentration across species.
- **Mechanism.** Protein binding differs between mouse, rat, and human serum by factors of two to ten for highly bound antifungals; the free fraction that drives efficacy therefore differs, and mouse-to-human dose translation fails in a species-dependent direction.
- **Discriminating observation.** The free fraction measured in each species' serum should predict the efficacy threshold better than total exposure, and the prediction should fail only for compounds with low binding.
- **Test.** Cross-species free-fraction measurement with efficacy reanalysis.
- **Prior art.** Free-drug exposure in PK/PD.
- **Refuted if.** Total exposure predicts efficacy equally well across binding classes.

Engineering, Earth and environmental science — 50 hypotheses

Every entry is a proposition. No experiment below has been performed.

E01 — Low-cost air-quality sensor networks report a bias that is a function of relative humidity and is absorbed into the calibration slope

`environmental-engineering` · `hidden-parameter` · readiness 5 · novelty 2

- **Baseline.** Optical particle sensors are calibrated against a reference in a dry chamber and deployed with a single slope.
- **Mechanism.** Hygroscopic growth increases the optical diameter of particles at high RH; the reported $PM_{2.5}$ then rises with humidity at constant mass. Co-location calibrations that span a narrow humidity range cannot identify the coefficient, so the deployed network's seasonal cycle contains a humidity artefact that mimics winter pollution.
- **Discriminating observation.** The sensor/reference ratio at fixed reference PM should correlate with RH with a slope that is independent of site.
- **Test.** Co-location data spanning a full humidity range.
- **Prior art.** Low-cost sensor correction literature.
- **Refuted if.** The ratio is RH-independent across sites.

E02 — Strain-based structural health monitoring attributes temperature-driven strain to damage

`civil-engineering` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Anomalies in strain or modal frequency are flagged as damage indicators.
- **Mechanism.** Thermal expansion and the temperature dependence of the elastic modulus produce daily and seasonal strain cycles that are larger than early damage signatures; a detector trained on a short baseline learns the seasonal cycle as normal and flags its return as damage.
- **Discriminating observation.** False alarms should cluster at particular seasons and should disappear when the detector includes measured temperature as a covariate.
- **Test.** Retrospective analysis of long-term monitoring records with temperature.
- **Prior art.** Environmental variability in SHM.
- **Refuted if.** Alarm rate is uncorrelated with seasonal temperature after detrending.

E03 — Battery state-of-health estimates from impedance are degenerate with the cell's compression state

`energy-engineering` · `degeneracy` · readiness 4 · novelty 3

- **Baseline.** Capacity fade is inferred from impedance growth attributed to SEI.
- **Mechanism.** Stack pressure changes the electrode's particle contacts and hence the impedance without any chemical degradation; a cell that has lost compression (from cycling-induced swelling) looks aged. The estimator cannot separate mechanical from chemical aging with a single impedance measurement.
- **Discriminating observation.** Impedance should change when stack pressure is externally varied on a cell of known chemistry, in the same direction as the "aging" signal.
- **Test.** Pressure-controlled impedance on fresh cells.
- **Prior art.** Pressure effects on cell impedance.
- **Refuted if.** Impedance is pressure-independent at fixed state of charge.

E04 — Rain-gauge calibration is biased by wind-induced undercatch in a way that increases with gauge age

`hydrology` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** Gauge catch is corrected with a wind-speed-dependent transfer function from a calibration campaign.
- **Mechanism.** The gauge's orifice and shield degrade with age (rim wear, shield deformation), changing the aerodynamic collection efficiency; the correction derived from new gauges therefore under-corrects old ones, and the network's long-term trend contains a maintenance artefact.
- **Discriminating observation.** The catch ratio against a reference gauge should decline with gauge age at matched wind speed.
- **Test.** Age-stratified co-located comparisons.
- **Prior art.** Precipitation undercatch corrections.
- **Refuted if.** Catch ratio is age-independent.

E05 — Tide-gauge sea-level trends include a subsidence term from the gauge's own foundation that is indistinguishable from the climate signal

`geodesy` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Tide-gauge trends are corrected with nearby GNSS vertical velocities.
- **Mechanism.** GNSS stations are anchored in bedrock while gauges are anchored in harbour structures; the two can move differently by tenths of a millimetre per year. The correction therefore removes a regional signal but not the local one, and the residual is systematic across harbours of similar construction.
- **Discriminating observation.** Gauge-minus-GNSS trends should cluster by harbour construction type.
- **Test.** Reanalysis with construction metadata.
- **Prior art.** Tide-gauge and GNSS integration.
- **Refuted if.** Residual trends are uncorrelated with construction type.

E06 — Seismic magnitude estimates from surface waves saturate at the same amplitude as the recording channel, in a magnitude-dependent way

`seismology` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Magnitudes are from maximum amplitude with a distance correction, after clipping detection.
- **Mechanism.** Clipping is detected only when the waveform is flat-topped; partial saturation of the digitizer or of the geophone's response compresses large amplitudes without visible clipping, biasing magnitudes low for the largest events and flattening the magnitude–frequency distribution.
- **Discriminating observation.** The magnitude residual against a coda-based estimate should correlate with the peak amplitude's proximity to the channel's full scale.
- **Test.** Reanalysis of stations with known full-scale values.
- **Prior art.** Magnitude saturation and clipping.
- **Refuted if.** Residuals are independent of proximity to full scale.

E07 — Radar rainfall estimates are biased by the drop-size distribution's dependence on the storm's lifecycle stage, not on its intensity

`remote-sensing` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Reflectivity is converted to rain rate with a Z–R relation fitted per region.
- **Mechanism.** Convective storms have a different drop-size distribution in their growth and decay phases at the same rain rate; a single Z–R relation averages over lifecycle, so the estimate is unbiased on average but biased by ±40% at any instant, and the bias is correlated in time.
- **Discriminating observation.** The radar/gauge ratio at fixed reflectivity should depend on the storm's lifecycle stage inferred independently (lightning rate or cloud-top temperature trend).
- **Test.** Storm-resolved radar–gauge comparison.

- **Prior art.** Z–R variability.
- **Refuted if.** The ratio is lifecycle-independent at fixed reflectivity.

E08 — The measured efficiency of a wind turbine is degraded by the nacelle anemometer's position, which is in the rotor's induction field

`energy-engineering` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Power curves are normalised by the nacelle wind speed.
- **Mechanism.** The nacelle anemometer sits behind the rotor and samples a slowed flow whose deficit depends on thrust coefficient and hence on operating point; the measured power curve therefore depends on the anemometer's location and on turbulence intensity, and manufacturers' curves are not comparable.
- **Discriminating observation.** The power curve should shift when the anemometer is moved to a mast or to a forward boom, with the shift maximised near rated power.
- **Test.** Side-by-side nacelle and mast measurements.
- **Prior art.** Nacelle anemometry corrections.
- **Refuted if.** Power curves are anemometer-position independent.

E09 — The measured soiling loss of a photovoltaic plant is dominated by the cleaning protocol's resuspension, not by deposition

`energy-engineering` · `model-artifact` · readiness 4 · novelty 2

- **Baseline.** Soiling loss is inferred from the transmittance drop between clean and soiled reference coupons.
- **Mechanism.** The reference coupon's surface energy differs from the module glass, and cleaning with water leaves a residue; the inferred loss therefore includes a cleaning artefact that accumulates over cycles and is larger for dry climates.
- **Discriminating observation.** Inferred soiling rate should depend on the cleaning method and on the coupon material at the same site.
- **Test.** Coupon-material and cleaning-method comparison.
- **Prior art.** Soiling measurement standards.
- **Refuted if.** Inferred rate is method-independent.

E10 — Acoustic leak detection in pipes is confused by the pipe's burial depth, which shifts the dispersion in a way that mimics a different leak size

`engineering` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Leak rate is estimated from the acoustic amplitude at a fixed attenuation coefficient.
- **Mechanism.** Burial depth and soil moisture change the attenuation of the leak's acoustic emission by an order of magnitude; the same leak size produces amplitudes differing by 10×, and the inversion attributes the difference to leak size.
- **Discriminating observation.** Inferred leak size should correlate with measured burial depth and soil moisture at fixed true leak (from excavation records).
- **Test.** Excavation-verified data set.
- **Prior art.** Acoustic leak localisation.
- **Refuted if.** Inferred size is independent of burial depth.

E11 — The measured thermal comfort of a building is a function of the sensor's radiant environment, not of the air temperature it reports

`building-science` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** Comfort models use air temperature and humidity from wall-mounted sensors.
- **Mechanism.** Wall sensors are radiatively coupled to the wall; in a poorly insulated wall the sensor reads a temperature biased toward the outdoor value, so the control system's setpoint error is a function of the wall's U-

value rather than of the room.

- **Discriminating observation.** The sensor reading should differ from a shielded, aspirated reference by an amount proportional to the wall's U-value and the indoor–outdoor difference.
- **Test.** Side-by-side sensor comparison across wall types.
- **Prior art.** Sensor placement in HVAC control.
- **Refuted if.** Readings are wall-type independent.

E12 — The measured capacity fade of a flow battery includes a crossover term that is a function of the pump's duty cycle

`energy-engineering` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Capacity fade is attributed to active-species degradation.
- **Mechanism.** Species crossover through the membrane depends on the concentration gradient at the membrane, which depends on flow rate and on the state of charge's spatial uniformity; a low-flow protocol allows stratification and increases local crossover. Reported fade rates therefore track the pump protocol.
- **Discriminating observation.** Fade rate should decrease with increasing flow rate above a threshold and should correlate with the measured crossover current.
- **Test.** Flow-rate series on the same cell.
- **Prior art.** Crossover in redox flow batteries.
- **Refuted if.** Fade rate is flow-rate independent.

E13 — The measured seismic site response is biased by the reference site's own amplification

`seismology` · `circularity` · readiness 4 · novelty 3

- **Baseline.** Site amplification is the spectral ratio of a target station to a reference station on rock.
- **Mechanism.** The reference station is never perfectly rigid; its amplification correlates with the target's because both are driven by the same regional wavefield. Standard spectral ratios therefore underestimate the true amplification and the underestimation grows as the reference degrades.
- **Discriminating observation.** Amplification should increase when the reference is replaced by a borehole sensor at depth, with the increase tracking the reference's measured downhole/urface ratio.
- **Test.** Sites with paired surface and borehole references.
- **Prior art.** Site-response estimation.
- **Refuted if.** Amplification is reference-independent.

E14 — The measured glacier mass balance from stake networks is biased by the stakes' own effect on local melt

`glaciology` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Stake emergence gives point mass balance, interpolated over the glacier.
- **Mechanism.** A dark stake absorbs radiation and enhances local melt, creating a depression that the next measurement records as additional ablation; the bias scales with stake albedo and with the number of seasons.
- **Discriminating observation.** Balance should be more negative at stakes than at nearby un-staked ice, with the difference scaling with stake albedo.
- **Test.** Paired staked and reference plots.
- **Prior art.** Stake-based mass balance.
- **Refuted if.** Staked and unstaked plots agree within error.

E15 — The measured ozone trend from sondes is affected by the sensor's pump efficiency, which degrades with age and with altitude

`atmospheric-science` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Ozone partial pressure is computed from the electrochemical cell's current with a constant pump correction.
- **Mechanism.** Pump efficiency falls with the reduced pressure and with contamination; the correction is a function of the sonde's age and of the launch profile, so long-term records from a changing fleet contain an instrument-generation trend that is confounded with the atmospheric trend.
- **Discriminating observation.** The sonde-to-satellite or sonde-to-lidar ratio should differ between sonde generations after correction.
- **Test.** Intercomparison campaigns across generations.
- **Prior art.** Ozonesonde homogenisation.
- **Refuted if.** Ratios are generation-independent after standard correction.

E16 — The measured soil moisture from cosmic-ray neutron sensors depends on the vegetation's biomass, which changes with the season

`hydrology` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Neutron intensity is inverted to soil moisture with a calibration that includes a biomass correction from a single sampling.
- **Mechanism.** The sensor's footprint integrates hydrogen from soil, vegetation, and the atmosphere; seasonal biomass growth changes the vegetation hydrogen pool by an amount comparable to a 0.05 change in volumetric moisture. A static biomass correction therefore produces a spurious seasonal cycle.
- **Discriminating observation.** The residual between sensor and gravimetric moisture should correlate with measured biomass at fixed true moisture.
- **Test.** Sites with paired biomass and gravimetric sampling.
- **Prior art.** Cosmic-ray neutron soil moisture.
- **Refuted if.** Residuals are uncorrelated with biomass.

E17 — The measured aerodynamic roughness of a city from tower data depends on the tower's own wake

`atmospheric-science` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Roughness length is from wind-profile fitting above the canopy.
- **Mechanism.** The tower and its instrumentation disturb the flow; the measured profile is inside the tower's wake for some wind directions, which biases the fitted roughness in a direction that depends on the tower's porosity.
- **Discriminating observation.** Fitted roughness should vary with wind direction relative to the tower, with a minimum in the direction of least wake.
- **Test.** Direction-resolved analysis on a single tower.
- **Prior art.** Tower-flow distortion.
- **Refuted if.** Roughness is direction-independent.

E18 — The measured groundwater recharge from the water-table fluctuation method is biased by the specific yield's dependence on the water table's depth

`hydrology` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Recharge is the water-table rise times a constant specific yield.
- **Mechanism.** Specific yield decreases with depth to the water table because the capillary fringe occupies a larger fraction of the profile; using a constant value overestimates recharge for deep water tables and underestimates it for shallow ones.
- **Discriminating observation.** The ratio of recharge to water-table rise should depend on depth to water table across sites.
- **Test.** Multi-site analysis with depth-stratified specific yield.
- **Prior art.** Water-table fluctuation method.

- **Refuted if.** The ratio is depth-independent.

E19 — The measured corrosion rate of buried pipeline steel from ER probes is dominated by the soil's aeration cycle

`engineering` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Electrical-resistance probes give a corrosion rate averaged over the deployment.
- **Mechanism.** Corrosion is controlled by the cathodic reaction, which depends on oxygen supply; seasonal water-table fluctuations change the aeration and hence the rate by an order of magnitude. A probe's average therefore reflects the deployment season.
- **Discriminating observation.** Probe-derived rates should differ between installations at the same site made in different seasons, and should correlate with measured redox potential.
- **Test.** Multi-season probe deployments.
- **Prior art.** Soil corrosion monitoring.
- **Refuted if.** Rates are deployment-season independent.

E20 — The measured fatigue life of a welded joint depends on the residual stress field, which the test specimen does not reproduce

`mechanical-engineering` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** S–N curves from standard coupons are applied to structures.
- **Mechanism.** Residual stress shifts the effective stress ratio; a coupon stress-relieved by its small size has a different crack-growth driving force than a full-scale weld. Life predictions therefore depend on the specimen's manufacturing history rather than on the material.
- **Discriminating observation.** Fatigue life should change when the specimen's residual stress is deliberately varied at fixed geometry and material.
- **Test.** Residual-stress-controlled coupon series.
- **Prior art.** Residual stress in fatigue design.
- **Refuted if.** Life is residual-stress independent.

E21 — The measured performance of a heat exchanger is degraded by fouling that the test protocol itself induces on the first run

`thermal-engineering` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Overall heat-transfer coefficient is measured on a clean exchanger and derated with a fouling factor.
- **Mechanism.** The first run deposits a layer whose thermal resistance is comparable to the design fouling factor; the "clean" reference is therefore taken too early, and the fouling factor double-counts.
- **Discriminating observation.** The coefficient should drop between the first and second run at identical conditions, and should recover partially after cleaning.
- **Test.** Repeated runs on the same exchanger.
- **Prior art.** Fouling factor methodology.
- **Refuted if.** Coefficient is run-order independent.

E22 — The measured efficiency of an electrolyser is biased by the bubble layer, and the bias depends on the current density's history

`energy-engineering` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Polarisation curves are measured by stepping current upward and recorded as steady state.
- **Mechanism.** Bubble coverage depends on nucleation site history; stepping upward leaves a supersaturated electrolyte that produces more bubbles at a given current than a downward step, so the two directions give different curves. The reported curve is a protocol artefact.

- **Discriminating observation.** Upward and downward polarisation curves should differ by more than the measurement error, with the hysteresis scaling with current density.
- **Test.** Direction-resolved polarisation on one cell.
- **Prior art.** Bubble effects in electrolyser testing.
- **Refuted if.** Upward and downward curves coincide.

E23 — Train delay propagation is measured with a threshold that removes the very events that carry the amplification

`transport-engineering` · `selection` · readiness 4 · novelty 3

- **Baseline.** Delay propagation is quantified by the ratio of delay at successive stations, excluding delays below a punctuality threshold.
- **Mechanism.** Small delays are the ones that amplify most (they have slack to lose); excluding them conditions the estimate on the tail and biases the amplification factor downward.
- **Discriminating observation.** The amplification factor should increase as the threshold is lowered, converging only at zero threshold.
- **Test.** Reanalysis of automatic train-running data at several thresholds.
- **Prior art.** Delay propagation analysis.
- **Refuted if.** The factor is threshold-independent.

E24 — The measured comfort of an indoor environment is confounded by the survey's own recruitment, which selects people who are uncomfortable

`building-science` · `selection` · readiness 3 · novelty 3

- **Baseline.** Comfort votes are collected from occupants and regressed on measured conditions.
- **Mechanism.** Response rates are higher among those dissatisfied; the sample's dissatisfaction is therefore higher than the population's, and the regression's intercept is biased while its slope is nearly unbiased. Predicted comfort temperatures are then wrong by a fixed offset.
- **Discriminating observation.** The intercept should shift with response rate while the slope stays constant, and both should stabilise when non-responders are sampled by a separate mechanism.
- **Test.** Response-rate-stratified surveys.
- **Prior art.** Non-response bias in comfort surveys.
- **Refuted if.** Intercept is response-rate independent.

E25 — The measured durability of a concrete from accelerated carbonation is not transferable because the accelerating CO_2 concentration changes the reaction product

`civil-engineering` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Carbonation depth follows a t^0.5 law whose rate constant is fitted at elevated CO_2 and extrapolated to ambient.
- **Mechanism.** At high CO_2 the reaction product is a different polymorph and the pore blocking differs, so the rate constant's dependence on concentration is not the square root that the extrapolation assumes.
- **Discriminating observation.** The fitted rate constant should scale with CO_2 concentration with an exponent different from 0.5, and the exponent should differ between binders.
- **Test.** Multi-concentration carbonation series.
- **Prior art.** Accelerated carbonation testing.
- **Refuted if.** The exponent is 0.5 for all binders.

E26 — The measured noise of a MEMS accelerometer includes a bias-drift term that is a function of the package's thermal gradient, not of the sensor

`instrumentation` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Bias stability is quoted from Allan deviation at controlled ambient temperature.
- **Mechanism.** On-chip thermal gradients generate thermoelectric voltages in the metallisation; the gradient depends on the package's mounting and on the PCB's power dissipation. The quoted stability therefore reflects the test fixture.
- **Discriminating observation.** Bias drift should correlate with the measured on-chip gradient rather than with ambient temperature, and should change when the mounting is changed.
- **Test.** Multi-mount comparison with on-chip thermometry.
- **Prior art.** Thermal effects in MEMS inertial sensors.
- **Refuted if.** Drift is mount-independent.

E27 — The measured sound insulation of a partition depends on the receiving room's absorption, which is not reported

`acoustics` · `boundary-condition` · readiness 5 · novelty 2

- **Baseline.** Sound reduction index is measured in reverberant rooms per standard, with a correction for receiving-room absorption.
- **Mechanism.** The correction assumes a diffuse field; in rooms with non-uniform absorption the correction is biased, and the bias is larger at low frequencies where modal behaviour dominates. Reported indices for the same partition can differ by several decibels.
- **Discriminating observation.** Measured index should depend on the receiving room's absorption distribution, not only on its average.
- **Test.** Inter-laboratory comparison with absorption mapping.
- **Prior art.** Acoustic transmission suite standards.
- **Refuted if.** Indices are absorption-distribution independent.

E28 — The measured state of charge of a lithium cell from open-circuit voltage requires a relaxation time that is longer than most protocols allow

`energy-engineering` · `time-dependent` · readiness 5 · novelty 3

- **Baseline.** OCV is read after a fixed rest and mapped to SOC.
- **Mechanism.** Diffusion relaxation in thick electrodes takes hours; a 1 h rest leaves a voltage offset whose sign depends on the prior current direction, so the SOC error is hysteresis-like and depends on the cycling history rather than on the cell.
- **Discriminating observation.** OCV at fixed SOC should depend on the preceding current direction and should converge only after multi-hour rests.
- **Test.** Rest-duration series after charge and after discharge.
- **Prior art.** OCV–SOC hysteresis.
- **Refuted if.** OCV is history-independent after 1 h.

E29 — The measured fire resistance of a composite is dominated by the test furnace's oxygen availability

`fire-engineering` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Fire resistance ratings come from furnace tests with a standard time–temperature curve.
- **Mechanism.** The furnace's ventilation condition changes the equivalence ratio and hence the heat flux and the combustion chemistry of the specimen; ratings obtained in different furnaces differ by minutes, which is comparable to the design margin.
- **Discriminating observation.** Time to failure should differ between furnaces at matched nominal curve, and should correlate with measured heat flux rather than with nominal temperature.
- **Test.** Inter-laboratory furnace comparison with heat-flux gauges.
- **Prior art.** Fire-test variability.
- **Refuted if.** Failure times are furnace-independent at matched heat flux.

E30 — The measured particle emission of a brake depends on the dynamometer's cooling airflow, which dilutes and ages the aerosol

`environmental-engineering` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Emission factors are from dynamometer tests with a standard dilution.
- **Mechanism.** Semivolatile species partition between particle and gas phase on a timescale set by dilution and temperature; a different dilution ratio changes the measured PM mass by tens of percent. Emission factors are therefore dilution-defined.
- **Discriminating observation.** Measured PM should depend on dilution ratio in a way predicted by partitioning theory, converging only at high dilution.
- **Test.** Dilution-ratio series on the same brake cycle.
- **Prior art.** Dilution effects in brake wear measurement.
- **Refuted if.** Emission factors are dilution-independent.

E31 — The measured thermal conductivity of an insulation is reported for a specimen thickness that suppresses convection

`building-science` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** κ is measured on a thin slab in a guarded hot plate and applied to thick layers.
- **Mechanism.** At the application thickness, natural convection within the porous structure and at the boundaries becomes significant; the thin-specimen value therefore underestimates the in-service conductivity, and the discrepancy grows with thickness and with the temperature difference.
- **Discriminating observation.** Measured κ should increase with specimen thickness above a threshold, and the threshold should depend on the pore size.
- **Test.** Thickness series at several temperature differences.
- **Prior art.** Convection in porous insulation.
- **Refuted if.** κ is thickness-independent up to the application thickness.

E32 — The measured thermal performance of a battery pack is dominated by contact resistance that varies with the cell's swelling over life

`energy-engineering` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Pack thermal models use a fixed cell-to-coolant conductance.
- **Mechanism.** Cycling swells the cell, increasing contact pressure and conductance; models fitted on fresh packs over-predict temperatures late in life, and safety margins are set from the fresh-pack fit.
- **Discriminating observation.** Measured contact conductance should increase with cycle count, in step with measured cell thickness.
- **Test.** Ageing series with thickness and thermal-impedance measurement.
- **Prior art.** Thermal contact in battery packs.
- **Refuted if.** Conductance is age-independent at fixed thickness.

E33 — The measured vibration serviceability of a floor is biased by the test's excitation method, which is not the excitation in service

`civil-engineering` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Natural frequencies and damping are from a hammer or shaker test and used to predict response to walking.
- **Mechanism.** Damping is amplitude-dependent for floors with non-structural partitions; a low-amplitude modal test measures a damping far below the in-service value, so the predicted resonant response is too high, and the calibration of the design guide inherits the bias.
- **Discriminating observation.** Damping should increase with excitation amplitude in floors with partitions, and predicted responses should match measured walking responses only when amplitude-matched.

- **Test.** Amplitude-series modal testing.
- **Prior art.** Amplitude-dependent damping in floors.
- **Refuted if.** Damping is amplitude-independent.

E34 — The measured efficiency of a thermoelectric generator in a system test includes a parasitic that is attributed to the material

`energy-engineering` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** System efficiency is compared with the material's zT to validate the material.
- **Mechanism.** Contact resistance, thermal shunting through the insulation, and the parasitic load of the converter are lumped into an apparent material degradation; the discrepancy is then attributed to a materials problem and drives the wrong optimisation.
- **Discriminating observation.** The system/material gap should correlate with independently measured contact and shunt terms across systems.
- **Test.** Reanalysis of module data with contact measurements.
- **Prior art.** TEG system efficiency.
- **Refuted if.** The gap is uncorrelated with parasitics.

E35 — The measured aerosol optical depth from a sun photometer is affected by the instrument's temperature dependence in a way that mimics a diurnal cycle

`atmospheric-science` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** AOD is from the Langley calibration and the instrument's internal temperature is logged but not used.
- **Mechanism.** Filter transmission and detector responsivity drift with temperature; the diurnal temperature swing produces an apparent diurnal AOD cycle that is currently interpreted as boundary-layer growth.
- **Discriminating observation.** AOD residuals should correlate with the instrument's internal temperature after removing air-mass effects, with a coefficient that is filter-specific.
- **Test.** Reanalysis of long AOD records with temperature telemetry.
- **Prior art.** Sun-photometer temperature characterisation.
- **Refuted if.** Residuals are uncorrelated with internal temperature.

E36 — The measured urban heat island from satellite skin temperature is biased by the surface's emissivity, which is not retrieved

`remote-sensing` · `assumption` · readiness 4 · novelty 3

- **Baseline.** Land surface temperature is from a split-window algorithm with an assumed emissivity.
- **Mechanism.** Emissivity differs between asphalt, roofs, and vegetation by 0.05–0.15, producing brightness-temperature errors of several kelvin; the urban–rural difference therefore partly reflects material composition rather than heat storage.
- **Discriminating observation.** The urban–rural difference should change when emissivity is retrieved per pixel rather than assumed, and the change should correlate with land-cover composition.
- **Test.** Reanalysis with a per-pixel emissivity product.
- **Prior art.** Emissivity in LST retrieval.
- **Refuted if.** The difference is emissivity-independent.

E37 — The measured earthquake recurrence from paleoseismic trenches is biased by the trench's detection threshold

`geology` · `selection` · readiness 3 · novelty 3

- **Baseline.** Recurrence intervals are from the number of events in a stratigraphic section divided by its duration.
- **Mechanism.** Small displacement events are not preserved or not recognised; the completeness threshold depends on sedimentation rate and on the trench's exposure. Sections with different rates therefore give

different “recurrence”, and the difference is a detection effect.

- **Discriminating observation.** Recurrence should correlate with sedimentation rate at fixed fault, converging when only large events are counted.
- **Test.** Multi-site compilation with rate metadata.
- **Prior art.** Paleoseismic completeness.
- **Refuted if.** Recurrence is sedimentation-rate independent.

E38 — The measured heat flux from a geothermal well is biased by the circulation’s own thermal disturbance

`geophysics` · `time-dependent` · readiness 4 · novelty 2

- **Baseline.** Heat flux is from a temperature log after drilling.
- **Mechanism.** Drilling fluid circulation resets the temperature profile; the re-equilibration time scales with the borehole radius squared over thermal diffusivity and can exceed the logging delay. The measured gradient therefore reflects the disturbance, and the bias is larger for larger holes.
- **Discriminating observation.** The inferred flux should decrease with successive logs over months and should converge; the convergence time should scale with hole diameter.
- **Test.** Repeated logging series.
- **Prior art.** Borehole thermal re-equilibration.
- **Refuted if.** Flux is stable from the first log.

E39 — The measured scour depth around a bridge pier from a single survey is a snapshot of a process with hysteresis

`civil-engineering` · `time-dependent` · readiness 3 · novelty 3

- **Baseline.** Scour is measured once and compared with a design value.
- **Mechanism.** Scour holes refill during low flow; the measured depth depends on the time since the last peak flow, so a survey after a quiet period underestimates the maximum. Design comparisons are therefore biased by survey timing.
- **Discriminating observation.** Scour depth should decay after peak flow at a rate set by the sediment’s settling velocity, and repeated surveys should show the decay.
- **Test.** Post-flood survey series.
- **Prior art.** Scour monitoring.
- **Refuted if.** Scour depth is stable between floods.

E40 — The measured thermal comfort of a naturally ventilated building depends on the wind pressure coefficient, which is assumed, not measured

`building-science` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Ventilation rates are from models using tabulated wind pressure coefficients.
- **Mechanism.** Coefficients depend on the surrounding obstructions and change with wind direction; a 20% error in the coefficient changes the predicted ventilation rate by a comparable amount, and the error is absorbed into the occupant’s adaptive behaviour, biasing comfort models.
- **Discriminating observation.** Predicted and measured ventilation rates (from tracer gas) should diverge systematically with wind direction in a way predicted by the coefficient’s direction dependence.
- **Test.** Tracer-gas campaigns with directional analysis.
- **Prior art.** Wind pressure coefficients in ventilation models.
- **Refuted if.** Predicted rates match measured across directions.

E41 — The measured thermal runaway onset of a cell depends on the calorimeter's crucible, not on the cell chemistry

`energy-engineering` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** Onset temperature is from accelerating rate calorimetry with a standard sample holder.
- **Mechanism.** The holder's thermal mass and the sample's contact area set the heat-loss term; onset is defined by a self-heating rate threshold, which is a heat-balance quantity. Different holders therefore give different onset temperatures for the same cell.
- **Discriminating observation.** Onset should shift with holder mass at fixed cell, in the direction predicted by the heat balance.
- **Test.** Holder-mass series.
- **Prior art.** ARC methodology.
- **Refuted if.** Onset is holder-independent.

E42 — The measured sediment oxygen demand in a lake is biased by the chamber's stirring, which sets the diffusive boundary layer

`limnology` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Benthic chambers measure oxygen drawdown over time.
- **Mechanism.** The diffusive boundary layer thickness depends on stirring; a thick layer limits flux and the measured demand underestimates the true one. Reported demands therefore reflect chamber design.
- **Discriminating observation.** Measured demand should increase with stirring and saturate at the mass-transfer limit; the saturation should depend on the sediment's roughness.
- **Test.** Stirring series across sediment types.
- **Prior art.** Benthic chamber methodology.
- **Refuted if.** Demand is stirring-independent.

E43 — The measured methane emission from a landfill by the tracer-release method is biased by the release height relative to the source

`environmental-engineering` · `boundary-condition` · readiness 3 · novelty 3

- **Baseline.** Emission is the known tracer release rate times the measured ratio of methane to tracer downwind.
- **Mechanism.** The ratio assumes the tracer and methane disperse identically; a tracer released at a different height or with a different buoyancy samples a different part of the plume, so the ratio depends on the release geometry.
- **Discriminating observation.** The inferred emission should change with tracer release height at fixed source, converging only when the release matches the source's effective height.
- **Test.** Height-series releases at one site.
- **Prior art.** Tracer-release emission measurement.
- **Refuted if.** Emission is release-height independent.

E44 — The measured bearing capacity of a pile from a load test includes a setup effect that is a function of the soil's thixotropy

`geotechnical` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Capacity is measured at a fixed time after driving and applied to long-term design.
- **Mechanism.** Pore-pressure dissipation and thixotropic recovery increase capacity with time; measurements at 7 days are below the long-term value, and the fraction recovered depends on the soil's plasticity. Design therefore depends on the test's timing.
- **Discriminating observation.** Capacity should increase logarithmically with time since driving, with a slope that correlates with the soil's plasticity index.
- **Test.** Multi-time load tests on instrumented piles.

- **Prior art.** Pile setup.
- **Refuted if.** Capacity is time-independent beyond one day.

E45 — The measured air-change rate of a building from a blower-door test is not the in-service rate, because the test changes the pressure field it measures

`building-science` · `boundary-condition` · readiness 5 · novelty 3

- **Baseline.** ACH50 is converted to natural infiltration by a leakage–infiltration ratio.
- **Mechanism.** The conversion assumes the leakage distribution and the stack effect are unchanged, but the blower door pressurises the whole envelope uniformly while in service the driving pressure varies by height. The conversion therefore depends on the building's leak distribution, which the test does not measure.
- **Discriminating observation.** Predicted infiltration should diverge from tracer-gas measurements in buildings with vertically uneven leakage, in a direction predicted by the leak distribution.
- **Test.** Paired blower-door and tracer-gas measurements with leak localisation.
- **Prior art.** Leakage–infiltration conversions.
- **Refuted if.** Predicted and measured rates agree regardless of leak distribution.

E46 — The measured extreme rainfall trend depends on the record's averaging interval, which changed with automation

`hydrology` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Extreme indices are computed from fixed-interval maxima.
- **Mechanism.** Automation changed the integration interval (e.g. from tipping-bucket counts to 1-minute samples), and the maximum over a shorter interval is systematically higher for the same storm. The trend in extremes therefore contains an instrument-transition step.
- **Discriminating observation.** The extreme index should show a step at the automation date that is absent in co-located manual records.
- **Test.** Parallel-record analysis across transition dates.
- **Prior art.** Instrument inhomogeneity in climate series.
- **Refuted if.** No step is detectable at transition dates after homogenisation.

E47 — The measured thermal conductivity of a soil from a probe is biased by the probe's contact resistance, which changes with drying

`geotechnical` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Thermal conductivity is from a transient line-source probe with a fixed geometry factor.
- **Mechanism.** As the soil dries it shrinks away from the probe, adding a contact resistance that grows with dryness; the measured conductivity falls faster than the true one, and the effect is confounded with the material's own moisture dependence.
- **Discriminating observation.** Measured conductivity should depend on the probe's diameter and on the installation method at matched moisture.
- **Test.** Multi-probe comparison at controlled moisture.
- **Prior art.** Soil thermal conductivity probes.
- **Refuted if.** Conductivity is probe-geometry independent.

E48 — The measured noise exposure from a wearable dosimeter is biased by the body's acoustic shadow

`occupational-health` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** Dosimeters are worn on the shoulder and assumed to sample the worker's exposure.
- **Mechanism.** The torso and head shadow high frequencies by several decibels, and the shadow depends on the source's direction relative to the body; exposure metrics therefore depend on posture and on the dominant

source direction.

- **Discriminating observation.** Measured level should change with the dosimeter's position on the body and with the worker's orientation to the source.
- **Test.** Multi-position dosimetry with source tracking.
- **Prior art.** Dosimeter placement effects.
- **Refuted if.** Levels are position-independent within 1 dB.

E49 — The measured thermal performance of a solar collector is reported at a flow rate that is not the operating one

`energy-engineering` · `boundary-condition` · readiness 4 · novelty 2

- **Baseline.** Efficiency is from a standard test at a specified flow rate.
- **Mechanism.** The collector's efficiency factor depends on the flow rate through the plate-to-tube bond resistance; at lower operating flow the efficiency drops, and the standard test's flow is above typical operation. Ratings therefore overstate field performance.
- **Discriminating observation.** Efficiency should decrease with flow rate below the test value in a way predicted by the fin efficiency.
- **Test.** Flow-rate series on one collector.
- **Prior art.** Collector test standards.
- **Refuted if.** Efficiency is flow-rate independent.

E50 — The measured seismic vulnerability of a building stock is biased by the inspection's targeting, which selects visibly damaged buildings

`earthquake-engineering` · `selection` · readiness 3 · novelty 3

- **Baseline.** Vulnerability curves are from post-earthquake damage surveys.
- **Mechanism.** Surveyors are directed to buildings that are visibly affected; the surveyed sample is enriched in weak structures, so the fragility curve is shifted toward higher vulnerability. Extrapolating to the stock overestimates loss.
- **Discriminating observation.** Fragility should depend on the survey's sampling frame, and should shift when a random sample of the stock is surveyed.
- **Test.** Comparison of targeted and random samples after one event.
- **Prior art.** Post-earthquake survey methodology.
- **Refuted if.** Targeted and random samples give the same fragility.

Cross-cutting methodology — 50 hypotheses

These concern assumptions that are shared across the natural sciences, and the measurement artefacts they produce. Every entry is a proposition. No experiment below has been performed.

X01 — Finite-window decay exponents are properties of the window, and the offset is identifiable from curvature

`methodology` · `identifiability` · readiness 5 · novelty 4

- **Baseline.** A power law $y = At^{(-\alpha)}$ is fitted to a decay over whatever range is available, and α is reported as a system property.
- **Mechanism.** With an unsubtracted additive offset c, the local log–log slope is $s(x) = -\alpha/(1+Be^{\alpha x})$, $B = c/A$: a logistic in ln t of width $1/\alpha$. A straight-line fit averages that logistic, so α_{fit}/α depends on where the window sits relative to the crossover $t_c = (A/c)^{1/\alpha}$, and equals exactly 1/2 when the window is centred on it.
- **Discriminating observation.** α should shift systematically when a free offset is allowed; the shift should be predicted by the fitted t_c.
- **Test.** Re-fit published decays with a free offset; complete derivation, Cramér–Rao bounds and Monte Carlo in investigation I1.
- **Prior art.** I1 of this collection; baseline-subtraction practice.
- **Refuted if.** Published exponents are unchanged by a free offset and show no curvature.

X02 — Regression dilution biases every calibration curve fitted to a diluted standard series

`methodology` · `errors-in-variables` · readiness 5 · novelty 2

- **Baseline.** Calibration slopes are fitted by ordinary least squares with the standard concentration treated as exact.
- **Mechanism.** Pipetting error makes the true concentration differ from the nominal one; OLS with error only in y is unbiased, but the standard practice of fitting y on x with x uncertain, or of using the inverse calibration for prediction, attenuates the slope by $1/(1+\lambda)$ with $\lambda = \sigma^2_x/\sigma^2_{est}$.
- **Discriminating observation.** The calibration slope should increase when the standards are prepared by independent weighings rather than serial dilution, with the increase predicted by the measured pipetting variance.
- **Test.** Serial vs independent standard series on the same instrument.
- **Prior art.** Regression dilution in analytical chemistry.
- **Refuted if.** Slopes agree within error between the two preparation routes.

X03 — Spectroscopic baselines fitted over a fixed window absorb real bands and bias the peak parameters

`methodology` · `model-artifact` · readiness 5 · novelty 3

- **Baseline.** A polynomial baseline is fitted to regions chosen by eye, then the peak is fitted on the residual.
- **Mechanism.** Window choice changes the baseline curvature, which changes the peak's width and amplitude in a correlated way; the fitted area can vary by 10–30% between analysts on the same spectrum while every fit looks good.
- **Discriminating observation.** Inter-analyst variance in fitted peak area should exceed the fit covariance by a factor that scales with window length.
- **Test.** Multi-analyst re-fitting of a shared spectral set.
- **Prior art.** Baseline-correction methodology in spectroscopy.
- **Refuted if.** Inter-analyst variance is within the reported fit errors.

X04 — Aliasing of a slow drift into a fast periodic signal manufactures a resonance that is not in the system

`methodology` · `time-dependent` · readiness 4 · novelty 3

- **Baseline.** Spectral peaks are interpreted as resonances once they exceed a significance threshold.
- **Mechanism.** Sampling a monotonic drift at a fixed cadence produces a low-frequency ramp whose spectral leakage appears as a peak at the drift's beat frequency; windowing redistributes but does not remove it. The peak frequency then depends on the sampling cadence, not on the system.
- **Discriminating observation.** The peak frequency should move when the sampling cadence is changed, while a true resonance stays fixed.
- **Test.** Multi-cadence resampling of an existing record.
- **Prior art.** Spectral leakage and aliasing.
- **Refuted if.** Peak frequencies are cadence-invariant.

X05 — Instrument response deconvolution amplifies noise in a way that creates spurious fast components

`methodology` · `ill-posed` · readiness 4 · novelty 3

- **Baseline.** Deconvolution is regularised and the regularisation strength is chosen to make the residual structureless.
- **Mechanism.** The chosen regularisation is a bias–variance trade selected on the residual, which is insensitive to bias; fast components are systematically suppressed or, with too little regularisation, synthesised from noise. The number of resolvable components is therefore a property of the regulariser.
- **Discriminating observation.** The recovered component count should change with regularisation strength and with sampling rate for the same underlying data.
- **Test.** Regularisation sweep on a dataset with a known synthetic truth.
- **Prior art.** Regularised deconvolution.
- **Refuted if.** Component counts are stable across regularisation strengths spanning a decade.

X06 — Model averaging over a discrete set of hypotheses produces confidence intervals that are too narrow when the set is incomplete

`methodology` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Bayesian model averaging gives posteriors that account for model uncertainty within the candidate set.
- **Mechanism.** The posterior mass is normalised within the set, so the variance reflects only the spread among candidates; if the true model is outside the set, the reported uncertainty is arbitrarily small and the interval excludes the truth with high probability.
- **Discriminating observation.** Coverage should degrade as the candidate set is made less complete, while remaining nominal for complete sets.
- **Test.** Coverage simulation with a deliberately incomplete model set.
- **Prior art.** Model-averaging coverage.
- **Refuted if.** Coverage stays nominal when the truth is excluded from the set.

X07 — Cross-validation on data that were preprocessed before splitting leaks information and inflates reported skill

`methodology` · `circularity` · readiness 5 · novelty 2

- **Baseline.** Normalisation, feature selection, or imputation are applied to the full dataset, then k-fold CV is run.
- **Mechanism.** The preprocessing statistics contain information from the held-out fold, so the CV estimate is optimistically biased; the bias grows with the number of features and with the strength of the preprocessing.

- **Discriminating observation.** Reported skill should drop when preprocessing is nested inside each fold, with the drop scaling with the number of features.
- **Test.** Re-run published pipelines with nested preprocessing.
- **Prior art.** Data leakage in predictive modelling.
- **Refuted if.** Skill estimates are unchanged by nesting.

X08 — Rounding and digitisation of published values create apparent structure in meta-analysis

`methodology` · `artefact` · readiness 5 · novelty 3

- **Baseline.** Reported means and uncertainties are treated as exact draws from the underlying distribution.
- **Mechanism.** Values are rounded to the reporting precision, so the measurement error distribution is uniform within a bin rather than Gaussian; meta-analytic weights computed from the reported σ therefore mis-weight studies with different reporting precision, and heterogeneity statistics acquire spurious structure.
- **Discriminating observation.** Heterogeneity should correlate with reporting precision across studies of the same quantity.
- **Test.** Reanalysis of published meta-analytic datasets with precision metadata.
- **Prior art.** Rounding in meta-analysis.
- **Refuted if.** Heterogeneity is uncorrelated with reporting precision.

X09 — Normalising to a reference sample imports that sample's error into every subsequent measurement

`methodology` · `hidden-parameter` · readiness 5 · novelty 2

- **Baseline.** Measurements are expressed relative to a reference or control to remove batch effects.
- **Mechanism.** The reference's error is common to all normalised values and therefore does not average out; it appears as a correlation between nominally independent samples, which is then mistaken for biological or physical structure.
- **Discriminating observation.** Inter-sample correlations should persist after randomising which sample is used as reference, if they are real, and should vanish or invert if they are reference artefacts.
- **Test.** Reference-permutation analysis on existing data.
- **Prior art.** Compositional and reference normalisation.
- **Refuted if.** Correlations are invariant to reference permutation.

X10 — Compositional data analysed with standard statistics produce correlations that are forced by the closure constraint

`methodology` · `artefact` · readiness 5 · novelty 2

- **Baseline.** Fractions or percentages are analysed with correlation, PCA, or regression directly.
- **Mechanism.** Components constrained to sum to one have spurious negative correlations; the strength depends on the number of components and on the total, so a compositional dataset analysed naively always shows a "trade-off" between components.
- **Discriminating observation.** Spurious correlations should appear in a synthetic closed dataset with independent components, and should vanish under log-ratio transformation.
- **Test.** Simulation plus reanalysis of a published compositional dataset.
- **Prior art.** Compositional data analysis.
- **Refuted if.** Correlations are unchanged by log-ratio transformation.

X11 — Censoring at a detection limit biases trends in a direction set by the trend itself

`methodology` · `selection` · readiness 4 · novelty 3

- **Baseline.** Values below a limit are replaced by a constant or excluded.

- **Mechanism.** Substitution compresses the low end; if the quantity is increasing over time, the fraction censored falls, so the substituted values occupy less of the record and the trend is steepened (or flattened, depending on substitution) by an amount that depends on the censoring fraction.
- **Discriminating observation.** The trend should change when a censored-data likelihood is used, and the change should scale with the censoring fraction's trend.
- **Test.** Reanalysis of environmental time series with reported detection limits.
- **Prior art.** Censored-data methods.
- **Refuted if.** Trends are unchanged under censored likelihoods.

X12 — Two laboratories measuring the same quantity with the same method disagree because of a shared systematic that neither can see

`methodology` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** Interlaboratory scatter is interpreted as random; a consistent offset is interpreted as one lab's error.
- **Mechanism.** A method-dependent systematic (e.g. a calibration standard's purity) is common to both labs and therefore invisible in the comparison; the comparison can only bound the *differential* error, and the common part propagates into any value traceable to that method.
- **Discriminating observation.** Comparison against a method with a different systematic (a primary method) should shift both labs together.
- **Test.** Reanalysis of key comparisons with a primary-method link.
- **Prior art.** Metrological traceability.
- **Refuted if.** Primary-method comparison leaves both labs unchanged.

X13 — Non-stationarity makes any estimated autocorrelation time too short, and hence every error bar too small

`methodology` · `time-dependent` · readiness 5 · novelty 3

- **Baseline.** Error bars on a time average use an autocorrelation time estimated from the same record.
- **Mechanism.** A slowly varying mean looks like long correlation at long lags but is fitted away by detrending; detrending removes exactly the variance that dominates the uncertainty, so the estimated correlation time is short and the error bar is small by the ratio of the trend's variance to the residual's.
- **Discriminating observation.** Error bars should grow when estimated from an ensemble of records rather than from one record's autocorrelation.
- **Test.** Comparison of single-record and ensemble error estimates on replicated measurements.
- **Prior art.** Non-stationary time-series uncertainty.
- **Refuted if.** Single-record and ensemble error estimates agree.

X14 — A statistically significant difference between two noisy measurements is not evidence that they measure different things

`methodology` · `inference` · readiness 5 · novelty 2

- **Baseline.** A tension is declared when the difference exceeds a threshold in units of the combined uncertainty.
- **Mechanism.** The threshold is a false-positive rate, not a model-selection criterion; with many comparisons, and with uncertainties that are themselves estimated, the expected number of "tensions" is nonzero even when both methods are unbiased, and the excess is driven by the uncertainty distribution's tails.
- **Discriminating observation.** The number of declared tensions should match the false-positive expectation when a null quantity is measured many times by both methods.
- **Test.** Null-comparison audit within a single field.
- **Prior art.** Significance thresholds in measurement comparison.
- **Refuted if.** Null comparisons yield far fewer tensions than the false-positive rate.

X15 — Fitting in log space when the noise is additive biases low-amplitude values upward and flattens trends

`methodology` · `model-artifact` · readiness 5 · novelty 3

- **Baseline.** Data spanning decades are fitted as ln y vs x because the scatter looks homoscedastic in log space.
- **Mechanism.** With additive Gaussian noise, ln y is negatively biased by half the relative variance; the bias is largest where y is smallest, so a log-space fit bends the low end upward and reduces the fitted slope. The distortion is systematic, not random.
- **Discriminating observation.** The fitted slope should change when the fit is performed in linear space with the correct error model, and the change should scale with the relative noise at the faint end.
- **Test.** Refit published data with the correct likelihood.
- **Prior art.** Log-transform bias.
- **Refuted if.** Slopes agree between log-space and correct-likelihood fits.

X16 — Smoothing before differentiating manufactures the feature that the derivative is used to locate

`methodology` · `ill-posed` · readiness 5 · novelty 2

- **Baseline.** Peaks in a derivative spectrum are located after Savitzky–Golay smoothing.
- **Mechanism.** The smoothing kernel's sidelobes create derivative structure at the kernel's width; a feature whose true width is below the kernel appears at the kernel scale, so the reported position and width describe the filter.
- **Discriminating observation.** The feature's width should scale with the smoothing window, and should converge only for windows narrower than the true width.
- **Test.** Window-series reanalysis.
- **Prior art.** Savitzky–Golay artefacts.
- **Refuted if.** Feature width is window-independent.

X17 — Error bars from repeated measurements under-reproduce the true error because repetition shares the systematic

`methodology` · `hidden-parameter` · readiness 4 · novelty 3

- **Baseline.** Uncertainty is the standard error of repeated measurements on one instrument.
- **Mechanism.** Repeats share the calibration, the sample preparation, and the operator; the standard error therefore estimates precision, not accuracy. The ratio of reproducibility to repeatability is a measurable quantity that is rarely reported, and it sets the true uncertainty.
- **Discriminating observation.** The ratio should exceed one by an amount that depends on the number of shared steps, and should approach one only for fully independent replicates.
- **Test.** Nested variance analysis on a published interlaboratory dataset.
- **Prior art.** Repeatability vs reproducibility.
- **Refuted if.** The ratio is unity for measurements sharing calibration.

X18 — Subsampling to balance a dataset discards the very regime where the model fails

`methodology` · `selection` · readiness 4 · novelty 3

- **Baseline.** Class-imbalanced data are downsampled so that the fit is not dominated by the majority class.
- **Mechanism.** The discarded majority samples are concentrated where the model is well constrained; after downsampling, the fit is dominated by the sparse region and the model's behaviour in the discarded regime is unconstrained but is reported as if fitted.
- **Discriminating observation.** Predictions in the majority regime should degrade after downsampling, and the degradation should be invisible in cross-validation that also uses the downsampled set.
- **Test.** Downsampled vs weighted-likelihood comparison.
- **Prior art.** Class imbalance in regression.

- **Refuted if.** Predictions are unchanged between downsampling and weighting.

X19 — Extrapolating a fitted trend beyond the calibration range is undetectable from the fit quality alone

`methodology` · `extrapolation` · readiness 4 · novelty 3

- **Baseline.** A model fitted over a range is used outside it, justified by a high R^2.
- **Mechanism.** R^2 is computed inside the range and is insensitive to the model's behaviour outside it; any smooth function can fit a finite range well while diverging outside, so the fit quality carries no information about extrapolation error.
- **Discriminating observation.** Hold-out tests at the range's edges should predict the extrapolation error, while in-range R^2 should not correlate with it.
- **Test.** Edge-hold-out on published fits.
- **Prior art.** Extrapolation risk.
- **Refuted if.** In-range R^2 predicts out-of-range error.

X20 — Consensus values from discrepant measurements depend on the weighting scheme more than on the data

`methodology` · `model-disagreement` · readiness 5 · novelty 3

- **Baseline.** A recommended value is a weighted mean with weights from the reported uncertainties, sometimes with a scale factor.
- **Mechanism.** When the measurements are discrepant, the weighted mean is dominated by the smallest quoted uncertainty; a 20% change in that uncertainty moves the consensus by more than the discrepancy it was meant to resolve.
- **Discriminating observation.** The consensus should move by more than its stated uncertainty when the smallest weight is perturbed within its own uncertainty.
- **Test.** Sensitivity analysis on published evaluations.
- **Prior art.** Data evaluation and weighting.
- **Refuted if.** Consensus is stable under weight perturbation within quoted errors.

X21 — Detrending with a fixed polynomial order removes a real low-frequency component and leaves a residual that looks like noise

`methodology` · `model-artifact` · readiness 4 · novelty 2

- **Baseline.** Trends are removed with a linear or quadratic fit before spectral or variability analysis.
- **Mechanism.** A physical low-frequency component (e.g. a multi-year oscillation) is partially absorbed by the polynomial; the residual has reduced power at low frequency, which is then reported as a lack of variability, and the absorbed power is redistributed into an apparent trend.
- **Discriminating observation.** The low-frequency power should depend on the detrending order and should recover when the trend is modelled physically rather than removed.
- **Test.** Order-sweep on a long record.
- **Prior art.** Detrending in climate and geophysical series.
- **Refuted if.** Low-frequency power is detrending-order independent.

X22 — Averaging spectra before fitting changes the fitted parameters nonlinearly

`methodology` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Replicate spectra are averaged to improve signal-to-noise, then fitted.
- **Mechanism.** The model is nonlinear in its parameters, so the fit of the mean is not the mean of the fits; with heterogeneous line widths, averaging narrows the apparent distribution and the fitted width approaches the narrowest component.

- **Discriminating observation.** Fitted widths should decrease with the number of averaged replicates when the replicates differ, while the mean of individual fits stays constant.
- **Test.** Reanalysis of replicate spectral sets.
- **Prior art.** Averaging before nonlinear fitting.
- **Refuted if.** Widths from averaged and per-replicate fits agree.

X23 — Instrument linearity is verified over a range that excludes the operating point

`methodology` · `blind-spot` · readiness 5 · novelty 2

- **Baseline.** Linearity is demonstrated with a dilution series spanning the calibration range.
- **Mechanism.** Detector nonlinearity is strongest at the extremes; if the operating point lies outside the verified range, or if the verification used a different gain or integration time, the check does not constrain the operating regime.
- **Discriminating observation.** Response should deviate from linearity when the gain or integration time is changed at a fixed true value.
- **Test.** Gain- and integration-time series at fixed standard.
- **Prior art.** Instrument linearity verification.
- **Refuted if.** Response is gain- and integration-independent across the operating range.

X24 — Anomalies detected by residual analysis are conditioned on the model that produced the residuals

`methodology` · `circularity` · readiness 4 · novelty 3

- **Baseline.** Outliers and anomalies are defined as residuals beyond a threshold from a fitted model.
- **Mechanism.** The model is fitted including the anomaly, which pulls the fit toward it and reduces its own residual; the detection power therefore decreases as the anomaly's leverage increases, and the anomalies that are detected are systematically those with low leverage (i.e. the less physically extreme ones).
- **Discriminating observation.** Injection-recovery tests should show detection efficiency decreasing with the injected anomaly's leverage.
- **Test.** Injection-recovery on the published pipeline.
- **Prior art.** Robust regression and outlier detection.
- **Refuted if.** Detection efficiency is leverage-independent.

X25 — Two datasets that agree within their error bars can still be inconsistent when the errors are correlated

`methodology` · `inference` · readiness 4 · novelty 3

- **Baseline.** Consistency is judged by the difference divided by the quadrature sum of uncertainties.
- **Mechanism.** If both datasets share a calibration or a template, their errors are positively correlated; the variance of the difference is then smaller than the quadrature sum, so the quoted tension is underestimated. Conversely, anticorrelated errors inflate an apparent tension.
- **Discriminating observation.** The tension should change when the shared systematic is included with its measured correlation, in a direction set by the sign of the correlation.
- **Test.** Reanalysis of published comparisons with correlation budgets.
- **Prior art.** Correlated systematics in data combination.
- **Refuted if.** Tensions are unchanged by including measured correlations.

X26 — Interpolation between calibration points creates structure at the scale of the spacing

`methodology` · `artefact` · readiness 4 · novelty 2

- **Baseline.** A smooth interpolant is used between sparse calibration points.

- **Mechanism.** Any interpolant imposes its own smoothness; if the true response has structure between points, the interpolant either misses it or, with splines, oscillates at the spacing scale. Mapped fields therefore contain artefacts at the sampling scale that look like physical features.
- **Discriminating observation.** The feature's spatial scale should match the sampling spacing and should move when the sampling is changed.
- **Test.** Resampling analysis.
- **Prior art.** Interpolation artefacts in mapping.
- **Refuted if.** Features are invariant to sampling density.

X27 — Selecting the number of components by the scree plot is equivalent to thresholding noise at an arbitrary level

`methodology` · `arbitrary-choice` · readiness 4 · novelty 2

- **Baseline.** The number of retained components is chosen where the eigenvalue spectrum flattens.
- **Mechanism.** For a noise-only spectrum the eigenvalues follow a Marchenko–Pastur distribution whose edge is a computable function of the aspect ratio; a scree-plot choice near that edge selects noise, and the choice moves with sample size. The retained "structure" is therefore a sample-size artefact.
- **Discriminating observation.** The retained component count should grow with sample size for pure noise data.
- **Test.** Component-count vs sample-size on surrogate noise.
- **Prior art.** Marchenko–Pastur thresholding.
- **Refuted if.** Component count is stable in sample size on noise.

X28 — Correcting for a nuisance variable that is measured with error introduces more bias than it removes

`methodology` · `errors-in-variables` · readiness 4 · novelty 3

- **Baseline.** A covariate is regressed out to control for a confounder.
- **Mechanism.** If the covariate is measured with error, residual confounding remains and the correction also removes part of the signal; the net bias can exceed the uncorrected bias, and the direction depends on the error variance ratio.
- **Discriminating observation.** The effect estimate should move non-monotonically as the covariate's measurement error increases, with a minimum bias at an intermediate error level.
- **Test.** Simulation calibrated to the measured covariate reliability.
- **Prior art.** Measurement error in covariate adjustment.
- **Refuted if.** Bias decreases monotonically with correction.

X29 — Peak-picking software's default parameters encode a prior on peak width that becomes the reported result

`methodology` · `model-artifact` · readiness 5 · novelty 2

- **Baseline.** Automated peak detection is treated as objective.
- **Mechanism.** The minimum peak width and the smoothing window are priors; peaks narrower than the window are merged, and the reported count depends on the default. Multi-software comparisons therefore measure software, not samples.
- **Discriminating observation.** Peak counts should differ systematically across software defaults on the same raw data, with the difference concentrated in the narrow-peak regime.
- **Test.** Multi-software reanalysis.
- **Prior art.** Reproducibility of automated peak detection.
- **Refuted if.** Peak counts are identical across defaults.

X30 — Physical constants extracted from a global fit absorb the fit's unmodelled physics and appear anomalously precise

`methodology` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** A parameter's uncertainty comes from the fit's covariance matrix.
- **Mechanism.** The covariance reflects only the parameters in the model; unmodelled physics is absorbed into whichever parameter is most flexible, which then appears both precise and shifted. The precision is therefore a statement about the model's rigidity, not about the measurement.
- **Discriminating observation.** The parameter should shift when the model is extended, by more than its quoted uncertainty, and the shift should be absorbed by the parameter with the largest flexibility.
- **Test.** Model-extension sensitivity on published fits.
- **Prior art.** Model dependence of fitted constants.
- **Refuted if.** Parameters are stable under physically motivated model extensions.

X31 — Uncertainty propagation through a nonlinear forward model is asymmetric, so the quoted symmetric error bar is wrong at the precision claimed

`methodology` · `nonlinearity` · readiness 4 · novelty 3

- **Baseline.** Uncertainties are propagated linearly and quoted as symmetric.
- **Mechanism.** For a forward model with curvature, the posterior is skewed; the symmetric interval is then either conservative on one side or excludes the truth on the other. At the precision where systematics are subdominant, this skew is the leading error.
- **Discriminating observation.** The bias between the posterior mean and the maximum-likelihood value should scale with the model's second derivative, and the coverage of the symmetric interval should deviate from nominal.
- **Test.** Coverage test on a calibrated forward model.
- **Prior art.** Nonlinear error propagation.
- **Refuted if.** Coverage is nominal for models with large curvature.

X32 — Data-driven background estimates are trained on the signal region's sidebands, which the signal contaminates

`methodology` · `circularity` · readiness 4 · novelty 3

- **Baseline.** Background is extrapolated from sidebands into the signal region.
- **Mechanism.** The signal's tail leaks into the sidebands; the extrapolated background is then biased high, which reduces the apparent signal. The effect scales with the signal's tail fraction, which is precisely what the analysis is trying to measure.
- **Discriminating observation.** The extracted signal should change when the sideband definition is moved, in a direction predicted by the signal's tail.
- **Test.** Sideband-definition sweep on published analyses.
- **Prior art.** Sideband background estimation.
- **Refuted if.** The signal yield is sideband-definition independent.

X33 — Statistically optimal weighting is biased when the uncertainties are estimated from the same data

`methodology` · `circularity` · readiness 4 · novelty 3

- **Baseline.** Inverse-variance weighting is applied with uncertainties estimated per point.
- **Mechanism.** Points with downward-fluctuating noise get smaller estimated uncertainties and larger weights, so the weighted mean is biased toward them; the bias is first order in the noise-to-signal ratio and is absent only when uncertainties are known a priori.

- **Discriminating observation.** The weighted mean should differ from an unweighted mean by an amount that scales with the uncertainty estimation error, with a sign fixed by the noise distribution's skew.
- **Test.** Simulation with data-estimated uncertainties.
- **Prior art.** Bias from estimated weights.
- **Refuted if.** Weighted and unweighted means agree at the predicted precision.

X34 — Reproducibility failures in a field are concentrated in quantities whose uncertainty is dominated by a single unmeasured nuisance parameter

`methodology` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Failed replications are attributed to error or to fraud.
- **Mechanism.** Quantities whose uncertainty budget is dominated by one uncontrolled parameter (sample storage, alignment, purification) have a heavy-tailed reproducibility distribution; the tail is a property of the budget's structure and is predictable from the reported budget.
- **Discriminating observation.** The replication-failure rate should correlate with the fraction of the uncertainty budget attributed to a single nuisance term.
- **Test.** Reanalysis of replication studies with published budgets.
- **Prior art.** Reproducibility meta-research.
- **Refuted if.** Failure rate is uncorrelated with budget structure.

X35 — Selecting a representative sample by its measured values makes the sample unrepresentative of the population it is meant to represent

`methodology` · `selection` · readiness 4 · novelty 3

- **Baseline.** A "representative" subset is chosen to match the population on measured covariates.
- **Mechanism.** Matching on measured values conditions on noise in those values; the selected subset is biased on the unmeasured covariates that correlate with the noise, a form of collider bias. The subset can then be less representative than a random one.
- **Discriminating observation.** Unmeasured covariate distributions should differ between matched and random subsets even when measured covariates match exactly.
- **Test.** Matched vs random subset comparison with held-out covariates.
- **Prior art.** Selection on observables.
- **Refuted if.** Held-out covariate distributions match.

X36 — Reanalysing published summary statistics cannot recover the effect that the original analysis reported when the original model was nonlinear

`methodology` · `identifiability` · readiness 4 · novelty 3

- **Baseline.** Meta-analysis combines published effect sizes assuming each is an estimate of the same parameter.
- **Mechanism.** For nonlinear models the reported effect size depends on the model's other parameters, which vary between studies; combining them estimates a mixture, not a common parameter, and the mixture's mean depends on the distribution of the nuisance parameters.
- **Discriminating observation.** Pooled estimates should change when studies are stratified by their nuisance parameters, even for the same underlying effect.
- **Test.** Stratified reanalysis of a published meta-analysis.
- **Prior art.** Meta-analysis of nonlinear models.
- **Refuted if.** Pooled estimates are stratification-independent.

X37 — A null result from an underpowered experiment is not evidence of absence, and its inclusion in a meta-analysis biases the pooled effect

`methodology` · `selection` · readiness 5 · novelty 2

- **Baseline.** All studies, including nulls, are pooled.
- **Mechanism.** A null from an underpowered study is consistent with a large effect; pooling treats it as a point estimate near zero, which pulls the pooled effect down by an amount that depends on the power distribution rather than on the effect.
- **Discriminating observation.** The pooled effect should change when studies are weighted by power rather than by precision.
- **Test.** Power-weighted reanalysis.
- **Prior art.** Power in meta-analysis.
- **Refuted if.** Pooled effect is weighting-scheme independent.

X38 — Calibration transfer between instruments assumes a linear mapping that fails precisely where the instruments differ most

`methodology` · `model-disagreement` · readiness 4 · novelty 3

- **Baseline.** A transfer function (slope and offset) maps one instrument's response to another's.
- **Mechanism.** Instrument differences are usually wavelength- or energy-dependent, so a two-parameter transfer is exact only near the calibration points; at the spectral extremes the transfer error exceeds the between-instrument difference it was meant to remove.
- **Discriminating observation.** Transfer residuals should grow toward the ends of the calibration range, and should be correlated with the instruments' response-difference spectrum.
- **Test.** Residual analysis on a transfer dataset.
- **Prior art.** Calibration transfer in spectroscopy.
- **Refuted if.** Transfer residuals are uniform across the range.

X39 — Excluding "failed" runs on the basis of a quality metric that depends on the treatment creates the treatment effect

`methodology` · `collider` · readiness 4 · novelty 3

- **Baseline.** Runs failing a QC threshold are excluded before analysis.
- **Mechanism.** The QC metric depends on the sample's state, which the treatment changes; exclusion is therefore differential, and the surviving groups differ in a way that mimics or masks the effect.
- **Discriminating observation.** The effect size should change when failures are retained with a censored model, and the change should scale with the differential failure rate.
- **Test.** Reanalysis with and without QC exclusion.
- **Prior art.** Differential exclusion bias.
- **Refuted if.** Effect sizes are unchanged by retention.

X40 — Sampling at a fixed interval synchronised to a periodic driver aliases the response to DC

`methodology` · `aliasing` · readiness 4 · novelty 3

- **Baseline.** Periodic responses are averaged over an integer number of cycles to remove them.
- **Mechanism.** If the sampling interval is an integer fraction of the drive period, the response appears constant and is removed by the average; the residual contains only the harmonics that are not aliased, which is a subset determined by the sampling, not by the system.
- **Discriminating observation.** Changing the sampling phase should reveal the response that the synchronised average removed, with an amplitude equal to the true one.
- **Test.** Phase-shifted resampling of an existing record.
- **Prior art.** Synchronous averaging artefacts.
- **Refuted if.** Phase-shifted analyses recover the same response amplitude.

X41 — Prior elicitation from expert opinion encodes the experts' shared training, so the prior is tighter than the knowledge

`methodology` · `hidden-parameter` · readiness 3 · novelty 3

- **Baseline.** Expert priors are treated as independent and combined.
- **Mechanism.** Experts trained on the same literature share errors; the combined prior's width shrinks as 1/√n under an independence assumption that is false, producing overconfidence that propagates into the posterior.
- **Discriminating observation.** The prior's width should stop shrinking once the experts' shared source is accounted for, and the posterior should widen accordingly.
- **Test.** Elicitation with source attribution and correlation modelling.
- **Prior art.** Expert elicitation methodology.
- **Refuted if.** Combined prior width shrinks as 1/√n after source de-duplication.

X42 — Reporting the best of several analysis choices is a selection effect that no correction in the paper can undo

`methodology` · `selection` · readiness 4 · novelty 3

- **Baseline.** The analysis pipeline is chosen during the study and reported as if fixed.
- **Mechanism.** The reported result is the maximum over an unrecorded set of choices; its distribution is the maximum of that set, so its uncertainty is understated by an amount set by the number of choices and their correlation. Only pre-registration or multiverse reporting recovers the correct distribution.
- **Discriminating observation.** The reported effect should sit at the edge of the multiverse distribution, and the multiverse's width should exceed the reported uncertainty.
- **Test.** Multiverse reanalysis of a published result.
- **Prior art.** Multiverse analysis and specification search.
- **Refuted if.** Reported effect is central in the multiverse with matching width.

X43 — A detection threshold calibrated on simulations inherits the simulation's physics, so the threshold is a property of the simulator

`methodology` · `circularity` · readiness 4 · novelty 3

- **Baseline.** Detection significance is calibrated on injected signals in simulated data.
- **Mechanism.** If the simulator lacks a physical effect present in the data, the background distribution is misestimated and the threshold corresponds to a different false-positive rate in reality; the discrepancy is unobservable from within the simulation.
- **Discriminating observation.** The false-positive rate measured on real, signal-free data should differ from the simulated one, with the difference scaling with the missing physics.
- **Test.** Data-driven background estimation versus simulation.
- **Prior art.** Simulation-based calibration.
- **Refuted if.** Real and simulated false-positive rates agree.

X44 — Combining measurements with a common systematic double-counts it when the correlation is assumed zero and under-counts when assumed one

`methodology` · `model-artifact` · readiness 4 · novelty 3

- **Baseline.** Correlations between datasets are set to 0 or 1 by convention.
- **Mechanism.** The true correlation is intermediate and often measurable from the shared inputs; either convention biases the combined uncertainty in a known direction, and the bias is largest when the shared systematic dominates.
- **Discriminating observation.** The combined uncertainty should be non-monotonic in the assumed correlation, with a minimum at the true value.
- **Test.** Reanalysis with a measured correlation.

- **Prior art.** Correlation in data combination.
- **Refuted if.** Combined uncertainty is correlation-insensitive when the shared term dominates.

X45 — Time-averaging a nonlinear response function does not commute with averaging the input

`methodology` · `nonlinearity` · readiness 5 · novelty 3

- **Baseline.** A measured mean response is converted to a mean input using the calibration curve.
- **Mechanism.** For a convex calibration, the mean response corresponds to an input above the mean input (Jensen's inequality); the error is second order in the input's variability and first order in the calibration's curvature. Averaged records therefore report a systematically different mean.
- **Discriminating observation.** The inferred mean should differ between high- and low-variability records at the same true mean, with the difference scaling as the variance times the curvature.
- **Test.** Reanalysis of records with known variability.
- **Prior art.** Jensen's inequality in instrument calibration.
- **Refuted if.** Inferred means are variability-independent.

X46 — Sparse sampling of a threshold crossing biases the estimated crossing time toward the sampling grid

`methodology` · `artefact` · readiness 4 · novelty 2

- **Baseline.** Crossing times are recorded at the sampling resolution.
- **Mechanism.** With a threshold crossing detected between samples, the recorded time is the first sample above threshold, which is late by half a sampling interval on average; if the sampling interval changed over the record (instrument upgrades), the bias changes and creates a trend.
- **Discriminating observation.** Crossing times should shift when the record is resampled at a finer grid by interpolation, with the shift scaling with the interval.
- **Test.** Resampling analysis on a long threshold-crossing record.
- **Prior art.** Threshold-crossing timing.
- **Refuted if.** Crossing times are resampling-invariant.

X47 — Normalising by a denominator measured with error creates a spurious negative correlation between ratio and denominator

`methodology` · `artefact` · readiness 5 · novelty 2

- **Baseline.** Ratios (per-cell, per-mass, per-area) are correlated with the denominator across samples.
- **Mechanism.** Division by a noisy denominator induces a negative correlation between the ratio and the denominator even when numerator and denominator are independent (Pearson's spurious correlation). The effect's magnitude is set by the denominator's coefficient of variation.
- **Discriminating observation.** The negative correlation should vanish when the denominator is measured more precisely or when the analysis uses a log-ratio model.
- **Test.** Precision-improvement comparison.
- **Prior art.** Spurious correlation of ratios.
- **Refuted if.** The correlation is unchanged by improving denominator precision.

X48 — Quoting a detection limit from the blank's standard deviation assumes the blank is representative of the sample matrix

`methodology` · `hidden-parameter` · readiness 5 · novelty 2

- **Baseline.** LOD = 3σ_blank/slope.
- **Mechanism.** The blank lacks the matrix that suppresses or enhances the signal; in complex matrices the LOD can be several times higher, so reported values between the two limits are not distinguishable from zero.

- **Discriminating observation.** The LOD measured in matrix should exceed the clean-blank LOD by a factor that correlates with the matrix's suppression factor.
- **Test.** Matrix-matched LOD determination.
- **Prior art.** Detection limits in trace analysis.
- **Refuted if.** Matrix and clean-blank LODs agree.

X49 — A monotonic but nonlinear sensor response makes any two-point calibration exact only at the points

`methodology` · `nonlinearity` · readiness 5 · novelty 2

- **Baseline.** Two-point calibration is used for a sensor assumed linear.
- **Mechanism.** A monotonic nonlinearity passes exactly through both calibration points and deviates in between; the maximum interpolation error is at the midpoint and is proportional to the curvature, so a sensor that "calibrates perfectly" can be wrong by the full nonlinearity in service.
- **Discriminating observation.** The interpolation error should be revealed by a mid-range check standard, and should scale with the span.
- **Test.** Multi-point check on a two-point-calibrated sensor.
- **Prior art.** Calibration practice.
- **Refuted if.** Mid-range check agrees with two-point interpolation within the stated error.

X50 — The assumption that a measured quantity is extensive fails at the nanoscale, and molar quantities are reported anyway

`methodology` · `boundary-condition` · readiness 4 · novelty 3

- **Baseline.** Per-mole or per-mass quantities are computed by dividing by the amount of material, assuming extensivity.
- **Mechanism.** At nanoscale, surface and interface contributions do not scale with volume; a quantity per mole therefore depends on particle size, and a size series will show an apparent size effect that is a definitional artefact rather than new physics.
- **Discriminating observation.** The per-mole quantity should scale with the surface-to-volume ratio, and should collapse when normalised by total surface area rather than by mass.
- **Test.** Size-series reanalysis with surface normalisation.
- **Prior art.** Extensivity at the nanoscale.
- **Refuted if.** Per-mole quantities are size-independent.

Part IV — Method

What was built

1. **A source base.** A programmatic harvest of 817 arXiv records submitted between January 2024 and December 2026, drawn from 49 queries aimed at measurement systematics rather than at subject matter: “systematic error”, “measurement bias”, “unobservable”, “blind spot”, “cannot be distinguished”, “degeneracy”, “identifiability”, “hidden variable”, “nuisance parameter”, “model discrepancy”, “interlaboratory”, “boundary condition”, “contact resistance”, “nonlinear impedance”, “third harmonic”, “small-signal approximation”, “Kramers–Kronig”, “instrument response”, “deconvolution”, “drift”, “aging”, “hysteresis”, “model misspecification”, and 26 others. The harvester is `src/harvest_arxiv.py`; the raw pack is `data/arxiv_pack.json`.
2. **A prior-art screen.** 358 unique prior Incoprea hypothesis titles were extracted from the existing export bundles, and every new title was compared against them by string similarity. Two substantive overlaps were found and replaced (see `CORRECTIONS.md`).
3. **Hypothesis construction.** 300 hypotheses were written directly, in six domain batches: physics and astronomy (50), chemistry and electrochemistry (50), materials (50), biology and medicine (50), engineering and Earth science (50), cross-cutting methodology (50). Each was required to have a named baseline, a mechanism, a discriminating observation, an available test, prior art, and an explicit refutation. The parser `src/parse_hyp.py` enforces the schema and rejects missing fields.
4. **Selection and refinement.** 30 were selected for development on three criteria: the mechanism had to be expressible quantitatively; the blind spot had to be invisible to the instrument or test that is normally used as a check; and the refutation had to be cheap. Refinement required labelling each part of the claim as an identity, an artefact, or a genuine prediction.
5. **Investigation.** Three were taken to full mathematical treatment with validated code: finite-window decay identifiability (I1), spot-size bias in thermal metrology (I2), and Kramers–Kronig blindness (I3). Every number quoted in the refined set is regenerated by `investigations/R1_refinement_numbers.py`.

What “readiness” means

Score	Meaning
5	A reanalysis of published data, or a two-point change to an existing measurement, would settle it
4	Existing public data or an existing apparatus can confirm or kill it
3	Needs a dedicated measurement campaign or a new analysis pipeline
2	Needs new instrumentation
1	Needs the idea made precise before it needs an experiment

Readiness is a judgement about cost, not importance. Of the 300: 45 are scored 5, 201 are scored 4, and 54 are scored 3.

What “novelty” means

Score	Meaning
5	A mechanism not present in the literature as far as the source base shows, with a new observable
4	A known mechanism applied where it is not currently applied, with a quantitative prediction
3	A known mechanism given a specific, testable form in a new domain
2	A known effect quantified for a domain that has not quantified it
1	A restatement or a small extension

Novelty 4–5 is claimed for 61 hypotheses; the rest are mostly 3. This is deliberate: most real progress in measurement science comes from taking a known mechanism seriously in a place it has not been taken seriously, not from inventing a new mechanism.

The three labels used in the refined set

- **Identity.** True by construction. The logistic slope law in R01, the Bessel conductance ratio in R02, the closed-form temperature rise in R03, and the series-resistance formula in R25 are identities. They are useful and they are not discoveries. Presenting them as discoveries is the most common way this kind of work goes wrong.
- **Model artefact.** A property of the analysis rather than of the world. R07 (spurious ratio correlation), R09 (log-space bias), R21 (weighting), and R20 (error-bar deflation) are artefacts in this sense: they are real effects on reported numbers, produced entirely by the analysis.
- **Genuine prediction.** A statement that could be false. R01's claim that the offset is recoverable from curvature, R02's claim that the KK test cannot detect the bias, R03's claim that $\kappa_app \propto s^2$ at large spot size, R11's frequency-dependence ratio, and R26's area-scaling collapse are all of this kind.

Limits

- **No experiment was performed.** Every result is analytic or numerical. The collection contains no new measurement, and nothing here should be cited as evidence that a proposed effect was observed.
- **The source base is arXiv-weighted.** arXiv covers physics, mathematics, materials, and quantitative biology well and clinical medicine poorly. Hypotheses in the medical entries rest on general methodological literature rather than on a dedicated source harvest.
- **The prior-art screen is lexical.** String similarity catches phrasing overlap, not conceptual overlap. Two overlaps were found; others may remain.
- **Quantitative predictions are model-dependent.** The numbers in the refined set are computed from stated models with stated parameters. They are illustrations of magnitude and scaling, not predictions for named materials or systems. Where a number depends on an assumed parameter (for instance the substrate conductance G in I2), the assumption is stated.
- **The readiness scores are optimistic by design.** A 4 means someone could test it soon, not that the test is easy or cheap.
- **Adversarial reading.** Several entries take the contrarian side of a live dispute (P11 on the neutron lifetime, P20 on clock self-gravity, P36 on the proton radius). They are marked, and each carries the observation that would show it is wrong.

Reproducing the collection

```
cd blindspot-2026-11-01
../.venv-research/bin/python src/parse_hyp.py            # rebuild CSV/JSON, validate schema
```

```
../.venv-research/bin/python investigations/I1_offset_window_bias.py
../.venv-research/bin/python investigations/I2_boundary_bias.py
../.venv-research/bin/python investigations/I3_kk_blindspot.py
../.venv-research/bin/python investigations/R1_refinement_numbers.py
python3 build/build_doc.py                                  # assemble and typeset
```

The arXiv harvest is not required to rebuild the document; `data/arxiv_pack.json` is included. Re-running `src/harvest_arxiv.py` will produce a different pack because arXiv grows, so the pack shipped here is the one this collection was built from.

Part V — Corrections and negative results

This file records everything in the collection that changed during construction, every claim that failed a quantitative check, and every place where a hypothesis was narrowed rather than defended. It is deliberately kept separate from the hypotheses so that a reader can see what the evidence did to the original idea.

1. Two hypotheses were replaced before publication

C38 (chemistry). The original entry claimed that electrochemical ammonia synthesis rates near the detection limit are dominated by ambient and membrane-derived contamination. A check against the prior Incoprea corpus found an existing hypothesis making essentially the same claim ("Most reported electrochemical ammonia synthesis is contamination"). Rather than republish it with a new number, C38 was replaced with a different electrochemistry hypothesis: that scanning electrochemical microscopy assigns a finite rate constant to insulating regions because its feedback theory assumes a conductive substrate. The replacement has a different mechanism (diffusion-limited approach curve fitted with a conducting-substrate model), a different observable (tip-radius dependence of the apparent rate over a known insulator), and a different refutation.

P20 (physics). The original entry attributed the residual muon $g-2$ discrepancy to a scheme mismatch in the intermediate window between lattice and data. A prior Incoprea hypothesis already assigns the residual to lattice hadronic-vacuum-polarization systematics. The overlap was judged too close. P20 was replaced with a different precision-physics hypothesis: that optical-clock comparisons inherit an unmodelled self-gravity term from the apparatus mass. The replacement is testable with a ballast-variation experiment and does not depend on the $g-2$ literature.

2. A hypothesis failed its own quantitative check: R04 / B01

Original claim. Plate-reader OD600 compresses above ~0.3–0.6 because of detector saturation and stray light, suppressing fitted growth rates.

What was computed. A stray-light model, measured OD = $-\log_{10}(10\text{^}(-\text{OD_true}) + S)$, with the bias in fitted μ obtained from a log–log regression over four candidate windows (`R1_refinement_numbers.py`, section A).

Result. - $S = 10^{-3}$, window OD 0.2–1.0: bias **−0.1%** - $S = 10^{-2}$, window OD 0.2–1.0: bias **−1.1%** - $S = 10^{-3}$, window OD 2.0–3.0: bias **−8.0%** - $S = 10^{-2}$, window OD 2.0–3.0: bias **−23.4%**

Conclusion. Stray light cannot explain compression in the OD 0.3–0.6 range. The hypothesis as originally stated is **not supported**. It was narrowed: the mechanism is real but confined to OD > 1.5, and any sub-exponential growth below OD 1 must have a different cause. The entry in `hyp/04-biology.md` now carries the corrected numbers and names the surviving candidate mechanisms (detector nonlinearity, meniscus and fill volume, evaporation, physiological shift) along with the tests that separate them.

Why it is kept. A hypothesis that loses cleanly is the point of the collection. The failed mechanism was removed from the suspect list for the low-OD regime, which is a real gain.

3. A hypothesis was corrected in direction but not in kind: R21 / X33

Original claim. Inverse-variance weighting with variances estimated from the same data biases the weighted mean toward points whose noise fluctuated low.

What was computed. 40 000 Monte Carlo trials with four measurements and 2–30 degrees of freedom per variance estimate.

Result. The mean is unbiased (-0.000 to 0.005). The *spread* of the estimator exceeds its nominal value: 0.686 at 2 dof, 0.582 at 5, 0.539 at 10, 0.511 at 30, against a nominal 0.5. The quoted uncertainty is therefore 73%, 86%, 93% and 98% of the true spread.

Conclusion. The heading was wrong, the phenomenon is real. The refined entry now claims uncertainty understatement rather than mean bias, with the dof-dependent magnitude.

4. A hypothesis overstated its own effect size: R20 / X13

Original claim. Non-stationarity makes every estimated autocorrelation time too short, and hence every error bar too small, by a large factor.

What was computed. For a linear trend of amplitude a against noise σ, the ratio of the true standard error to the naive one is $\sqrt{(a^2/12 + \sigma^2)}/\sigma$.

Result. 1.01 at $a/\sigma = 0.5$; 1.04 at 1; 1.15 at 2; 1.53 at 4.

Conclusion. The effect is real but modest — under a factor of two even when the trend is four times the noise. The refined entry says so explicitly and restricts the claim to regimes where the trend dominates.

5. A claim about test power was softened after computation

In I1 the initial framing assumed that a broad distribution of relaxation times would be indistinguishable from a power law over a short window. Direct evaluation of the log-normal Laplace transform showed the opposite: a log-normal spectrum produces a strongly curved decay, and a 0.5-decade window is enough to reject a single power law at 1% noise when the window sits on the spectral knee. The investigation was rebuilt around the offset term instead, where the degeneracy is genuine and the Cramér–Rao bound actually diverges. The discarded framing is recorded here rather than deleted.

6. Numerical errors caught during development

- **I2, Hankel quadrature.** The first version of the validation integral was missing a factor of two and disagreed with the closed form by exactly 0.5 at every λ. The closed form was correct; the check was fixed. Agreement is now 2×10^{-13}.
- **I3, spectral derivative.** The first harmonic-balance implementation divided the derivative of η by N, making the capacitive current 128× too small. The symptom was a fitted C_dl of 0.156 μF against a true 20 μF. Fixed by using the exact rfft/irfft derivative.
- **I3, Bessel comparison.** The first comparison used the thermal voltage RT/F where the Tafel parameter RT/(αF) was required, so the analytic and numerical curves differed by a factor of two in x. The identity was right; the comparison was wrong.
- **I3, Lin–KK.** The first Lin–KK implementation used an unbounded least-squares Voigt fit and reported a residual of 0.17 even for an exactly linear spectrum. Rather than use a fragile test, the final version compares the measured spectrum directly against the exact Debye form, which gives 10^{-7} and is unambiguous.
- **R1, DSC lag.** The first version double-converted scan rate and produced onset shifts of 12–200 K. Replaced with a time-constant formulation giving 0.4–6.7 K.
- **R1, TDTR.** The first sensitivity model was two series resistances, which is trivially degenerate at all frequencies. Replaced with a thermal-quadrupole multilayer model that includes the film's heat capacity and reproduces the expected frequency dependence.

7. Things that were not done

- No experiment was performed. No apparatus was built. No sample was measured.
- No published dataset was reanalysed directly. The reanalysis predictions in I2 and R03 are stated as tests, not as results.
- The arXiv records cited were retrieved through the arXiv API and their abstract pages returned HTTP 200 on 1 November 2026. Their content was read at abstract level; no claim in this collection depends on a number taken from a paper without saying so.
- The 300 hypotheses were checked against 358 prior Incoprea hypothesis titles with a string-similarity screen. Two overlaps were found and replaced (section 1). The similarity screen cannot catch semantic overlap with different wording, so residual overlap remains possible.