Workshop on Trustworthy Statistical Inference for the Physical Sciences

STAMPS@CMU · May 13–15, 2026

Adaptive Frasian Inference

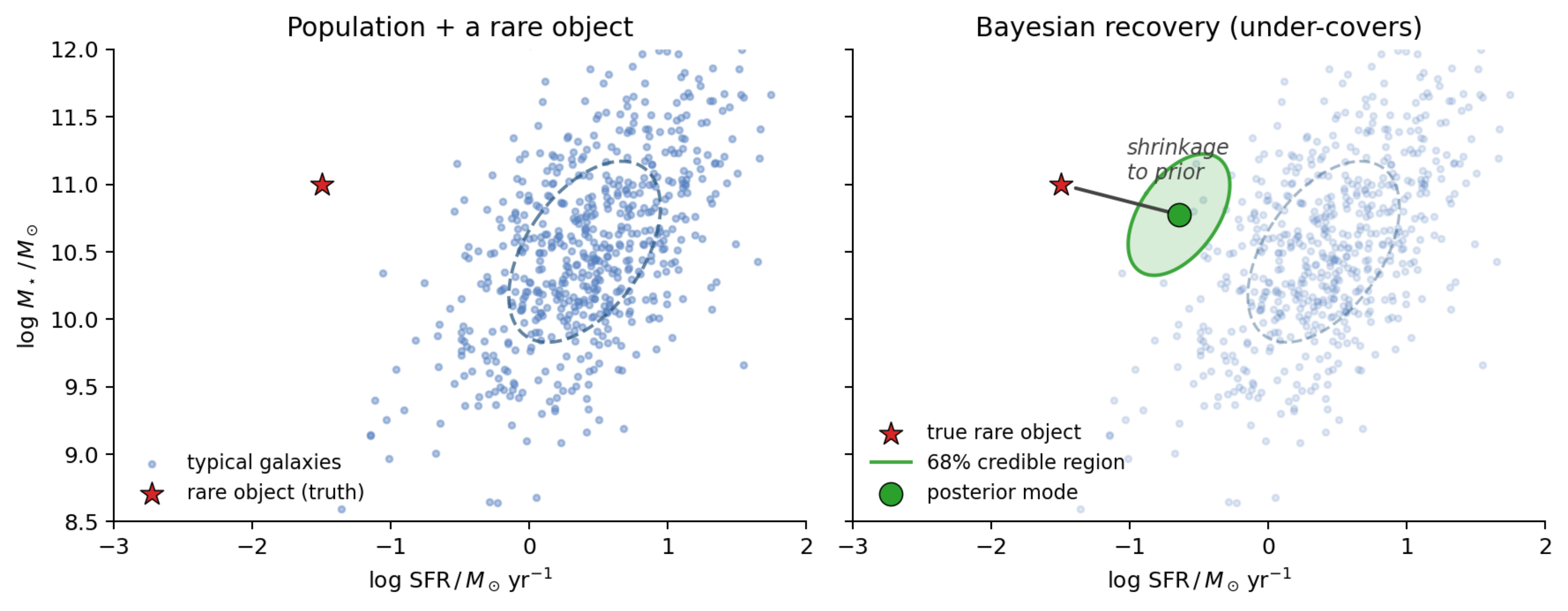
under Data–Prior Conflict

Joshua S. Speagle (沈佳士)
Department of Statistical Sciences
David A. Dunlap Department of Astronomy & Astrophysics
Dunlap Institute for Astronomy & Astrophysics
Data Sciences Institute
University of Toronto

with James Carzon, Ricardo Baptista, & Ann Lee

Slides generated with Claude Opus 4.7

The interesting galaxies are the ones the prior fights



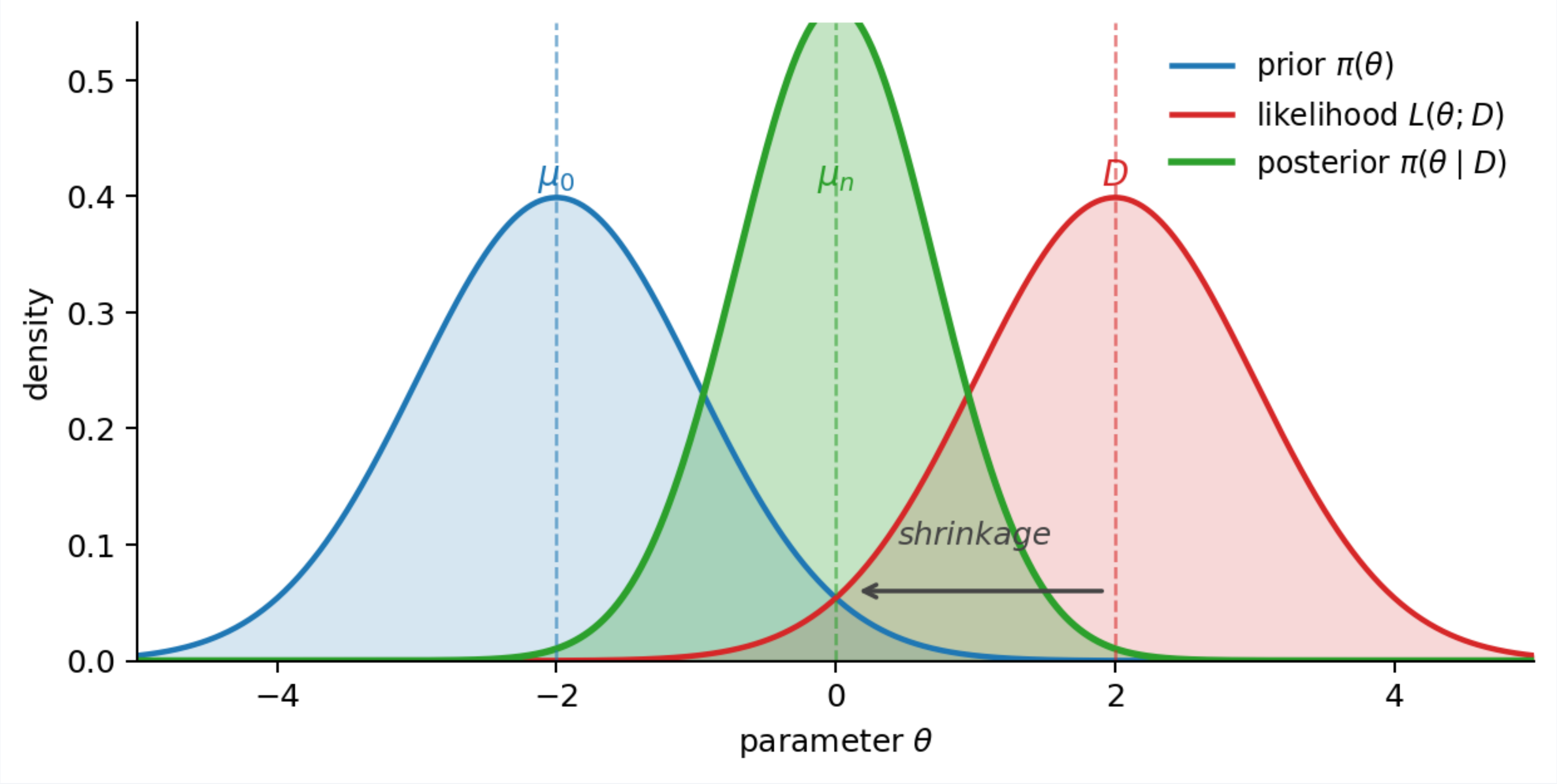
Rare = where the **science** is

Population prior **fights** the data

Posterior shrinks to population mean — bias

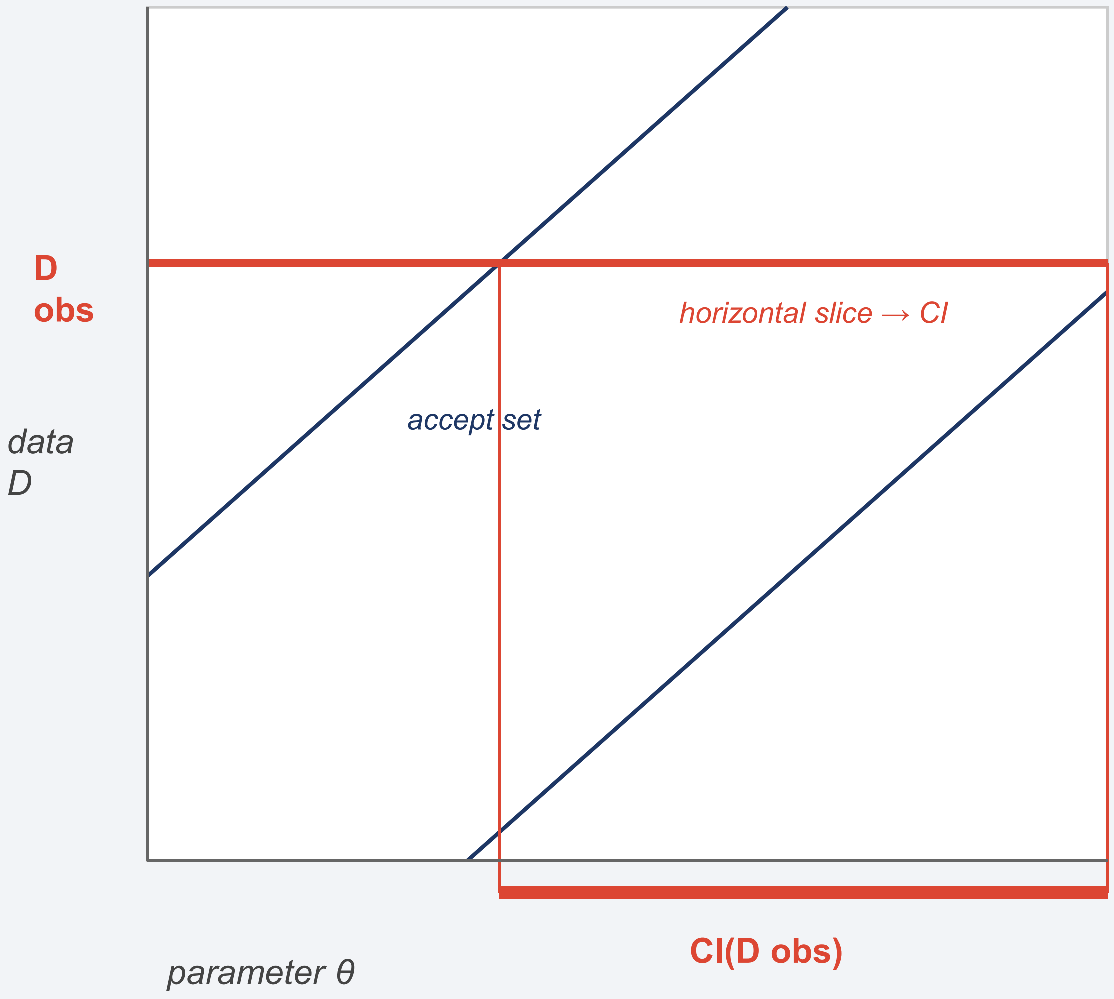
Prospector: Johnson et al. 2021, Leja et al. 2017 · Prospector-β: Wang et al. 2023 · atypicals: Suess et al. 2022, Narayanan et al. 2024

The prior has information — but neither side uses it well



- **Bayes** lets the prior dominate — biased point
- **Standard frequentist** throws the prior away
- **Want**: use prior when it helps, robust when it doesn't

Test inversion: CIs from p-values

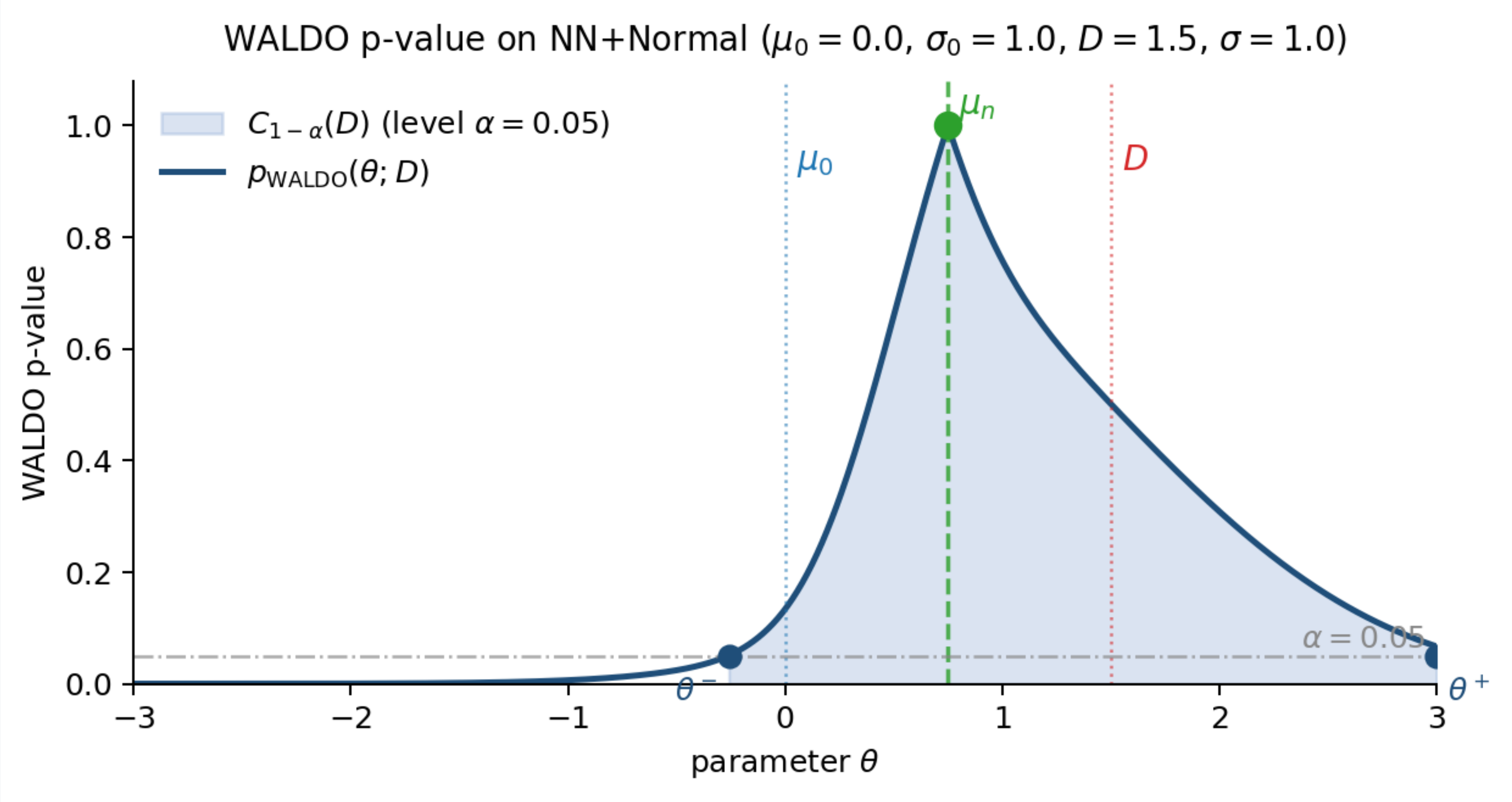


Casella & Berger 2002 — test inversion (textbook)

$CI(D) = \{ \theta : p(\theta; D) \geq \alpha \}$

- Frequentist machinery, **no Bayesianism required**
- The **test statistic** sets the CI shape
- So: pick a statistic that **uses the prior**

WALDO: a *Frasian* test statistic

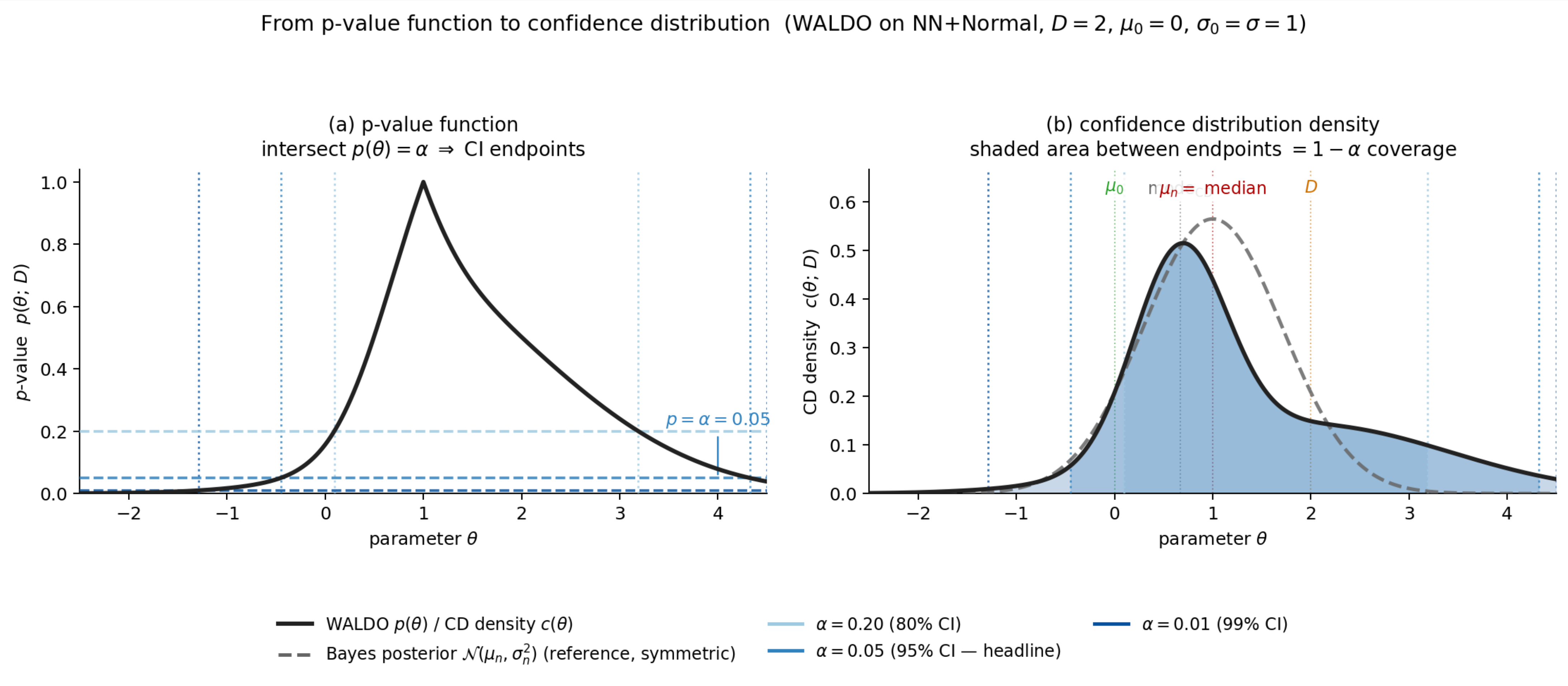


$$T = (\theta - \mu_{post})^2 / \sigma^2_{post}$$

- Statistic from the **posterior**
- Calibrated **frequentist**
- *Frasian* = freq. test + Bayes statistic

Frasian lineage: *Fraser 1968 · Dempster 1974, 1997 · Aitkin 1991, 1997 · Wasserman 2011 · Dalmasso et al. 2020 (LF2I) · Masserano et al. 2023 (WALDO) · Carzon et al. 2025*

Stacking CIs across α gives a confidence distribution

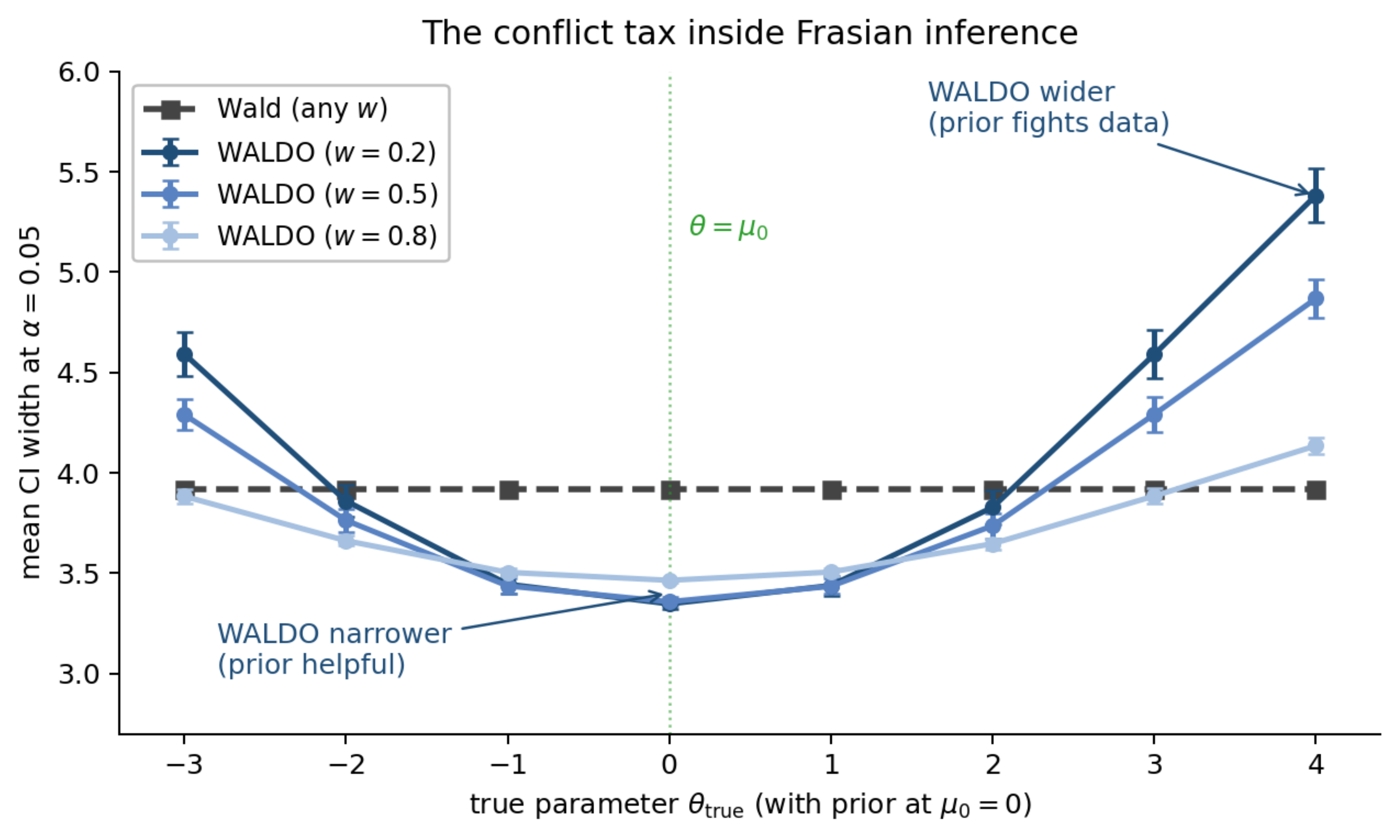


Same coverage guarantee, posterior-shaped density

WALDO CD is asymmetric: mode toward μ_0, median at μ_n

CD lineage: *Xie & Singh 2013 — CD review · Schweder & Hjort 2016 Ch. 3 — calibration, Ch. 4 — fusion · Singh, Xie & Strawderman 2005 — confidence fusion*

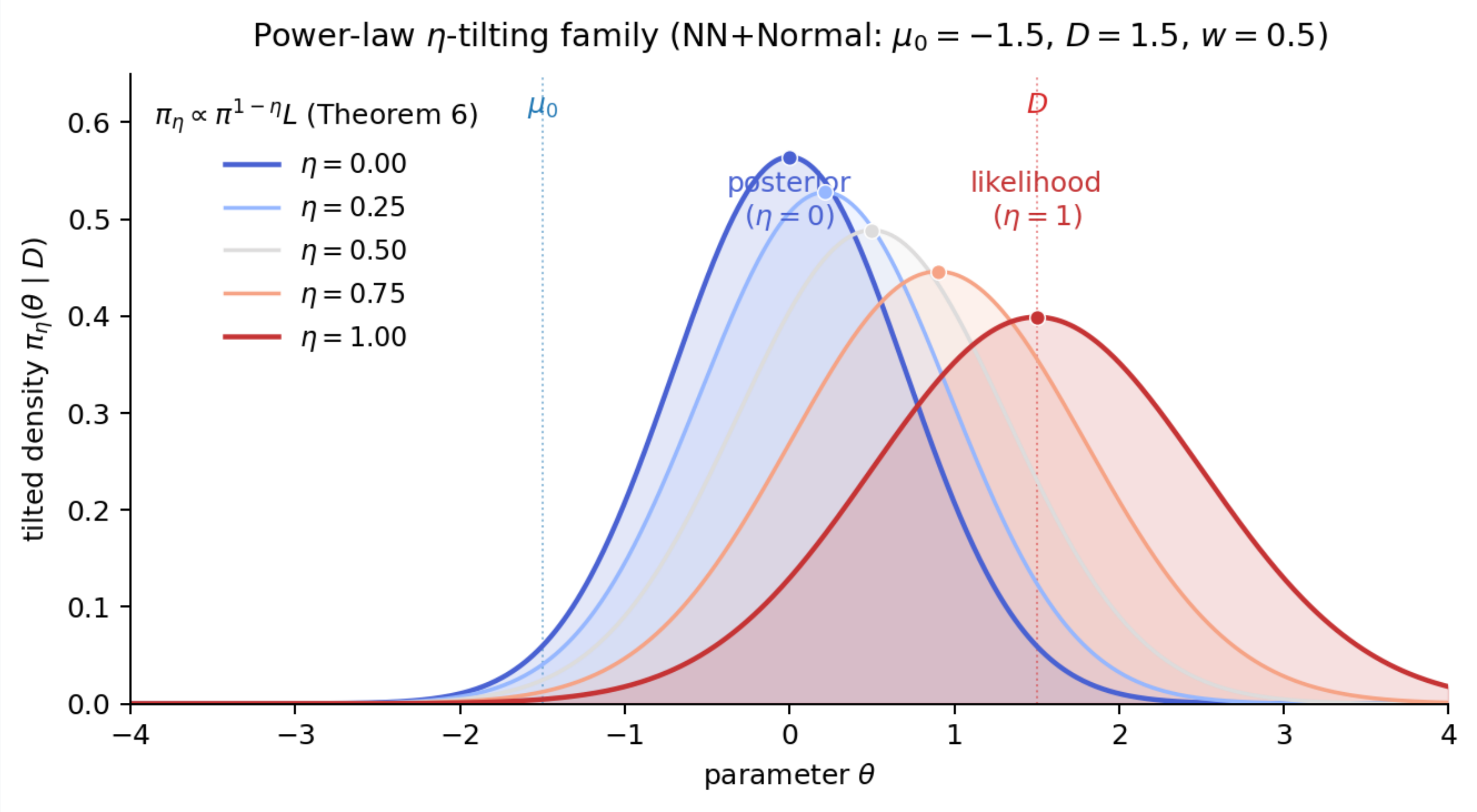
Coverage holds — but width pays a conflict tax



- **Coverage** holds everywhere ✓
- **Width** grows with |Δ|
- **~30%** wider than Wald at moderate conflict
- The ***conflict tax***

Empirical: Normal–Normal sandbox audit, α = 0.05, prior strengths w ∈ {0.2, 0.5, 0.8}, sweep over true θ

A **power-law** η-knob between WALDO and Wald



$$\pi_\eta(\theta|D) \propto \pi(\theta)^{1-\eta} \cdot L(\theta;D)$$

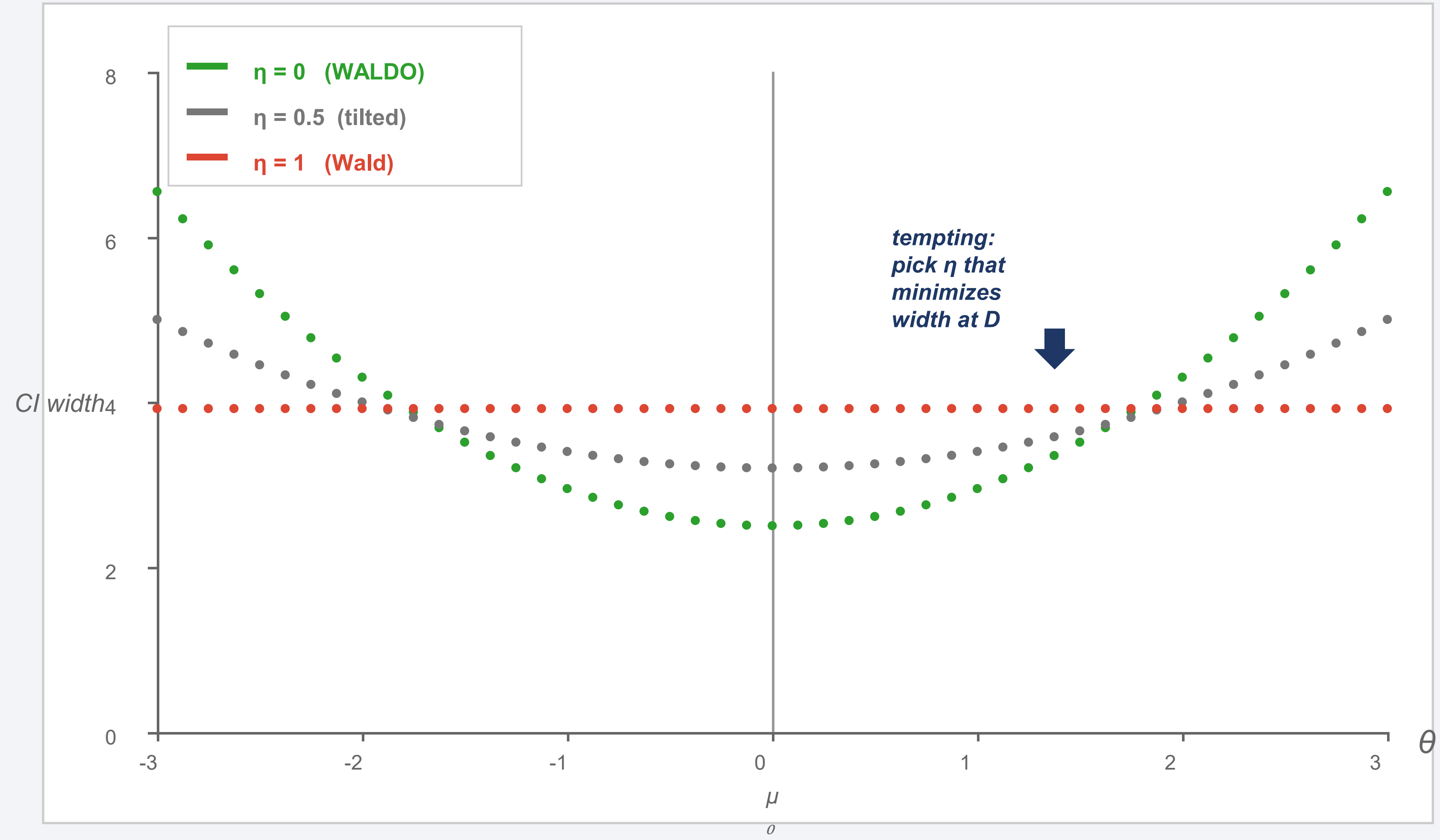
- **η = 0**: full prior (WALDO)
- **η = 1**: no prior (Wald)

How to pick η(·) — later

Power-law lineage: *Amari & Nagaoka 2000 — e-geodesic · Friel & Pettitt 2008 — tempered posteriors · Bissiri, Holmes & Walker 2016 — generalized Bayes · Grünwald & van Ommen 2017 — safe/fractional Bayes*

Each η gives a different CI

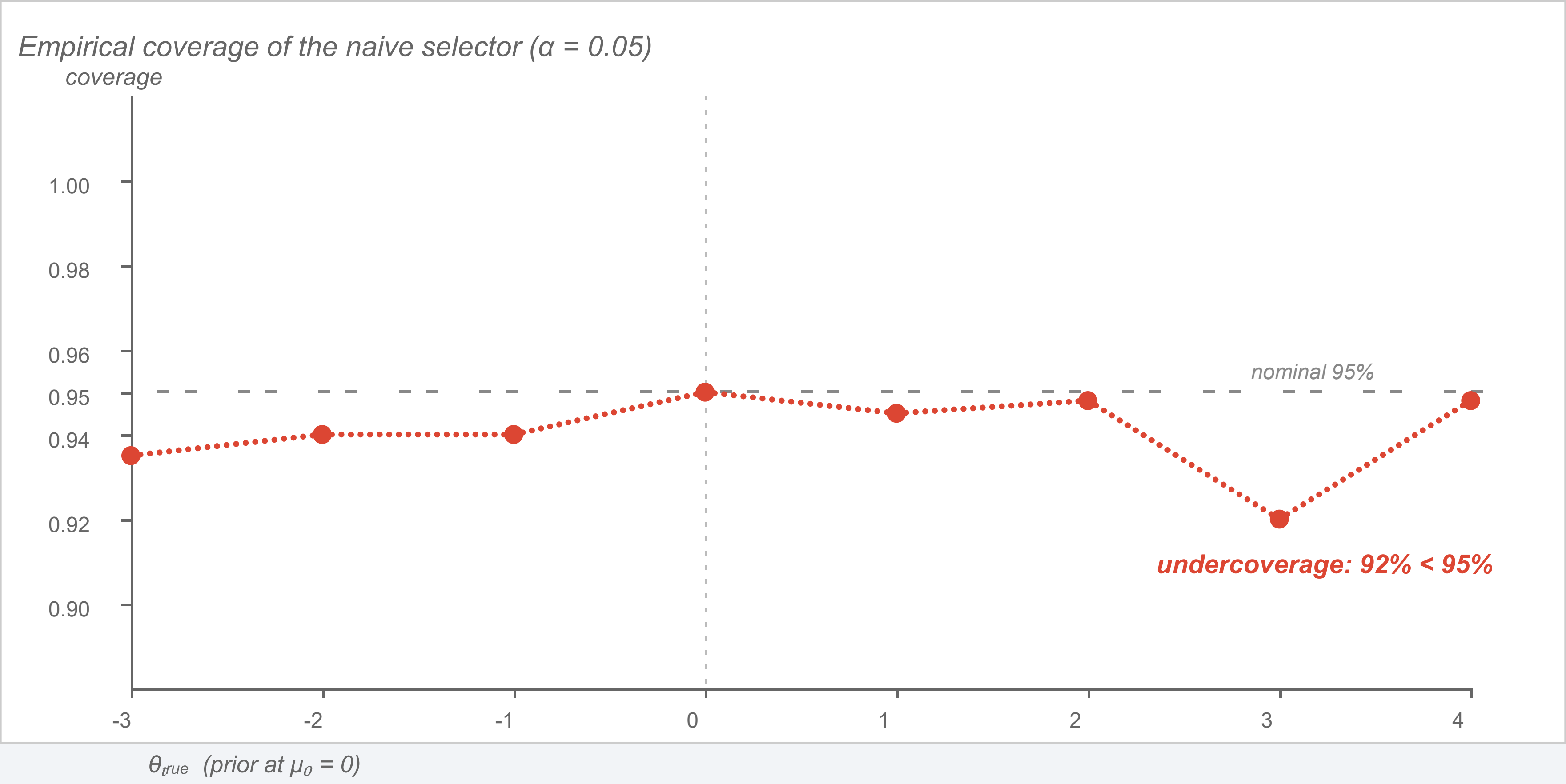
CI width as a function of θ, for fixed η



- Different **η** → different CI width profile
- All are **calibrated** — they trade width for shape
- Tempting: at each **D**, pick the tightest η
- *(spoiler — this breaks calibration)*

Schematic: CI width for power-law family on NN+Normal at fixed (μ_0, σ_0, σ, w).

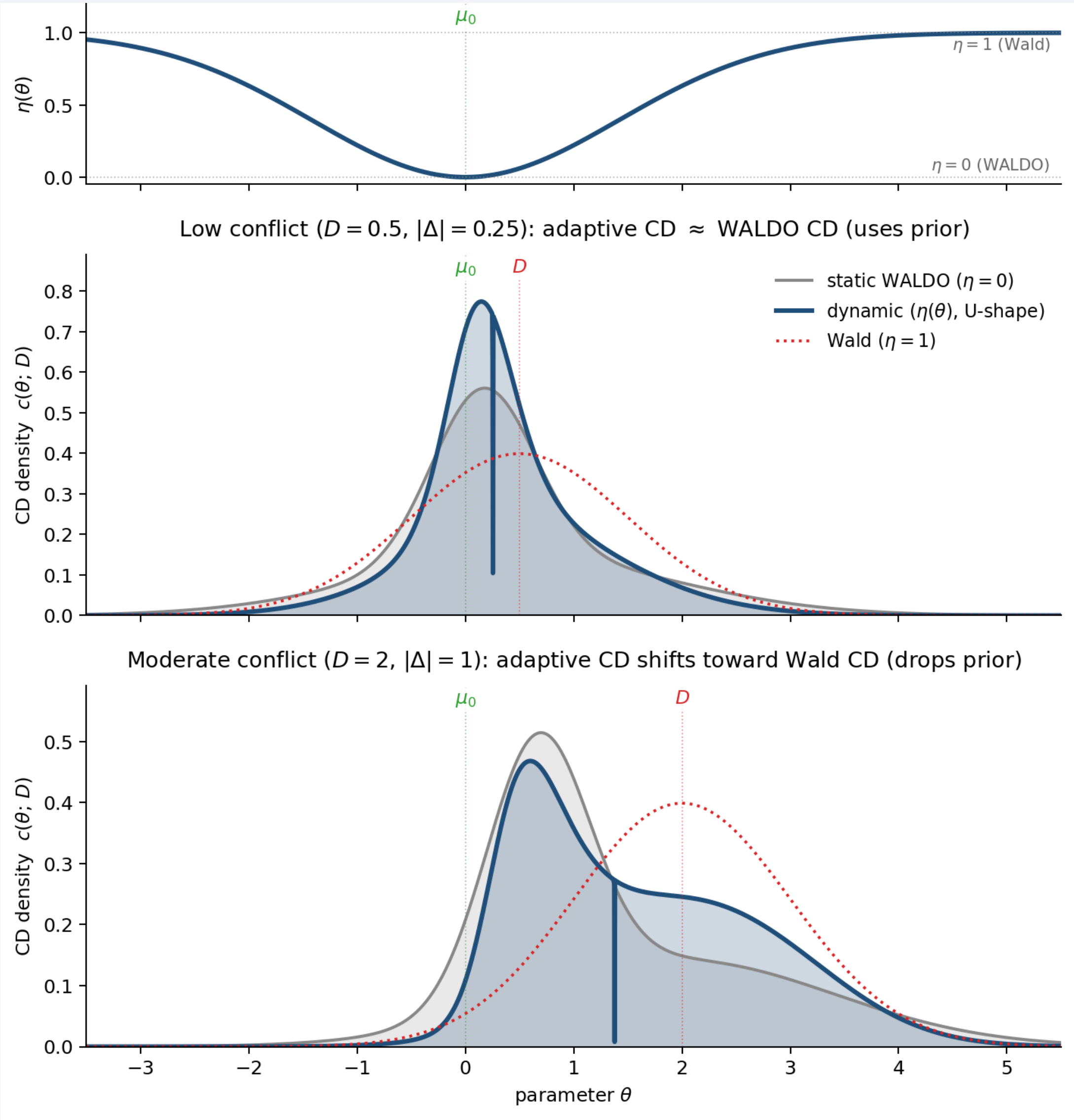
The naive recipe breaks



- Recipe: pick $\eta^* = \text{argmin width}(\eta; D)$
- Empirical coverage: **~93%** — below nominal 95%
- Data picked the test, then tested itself
- **Post-selection inference** — guarantee gone

Berk et al. 2013 — valid post-selection inference · Fithian, Sun, Taylor 2014 — exact PSI

Dynamic tilting: η(θ), not η(D)

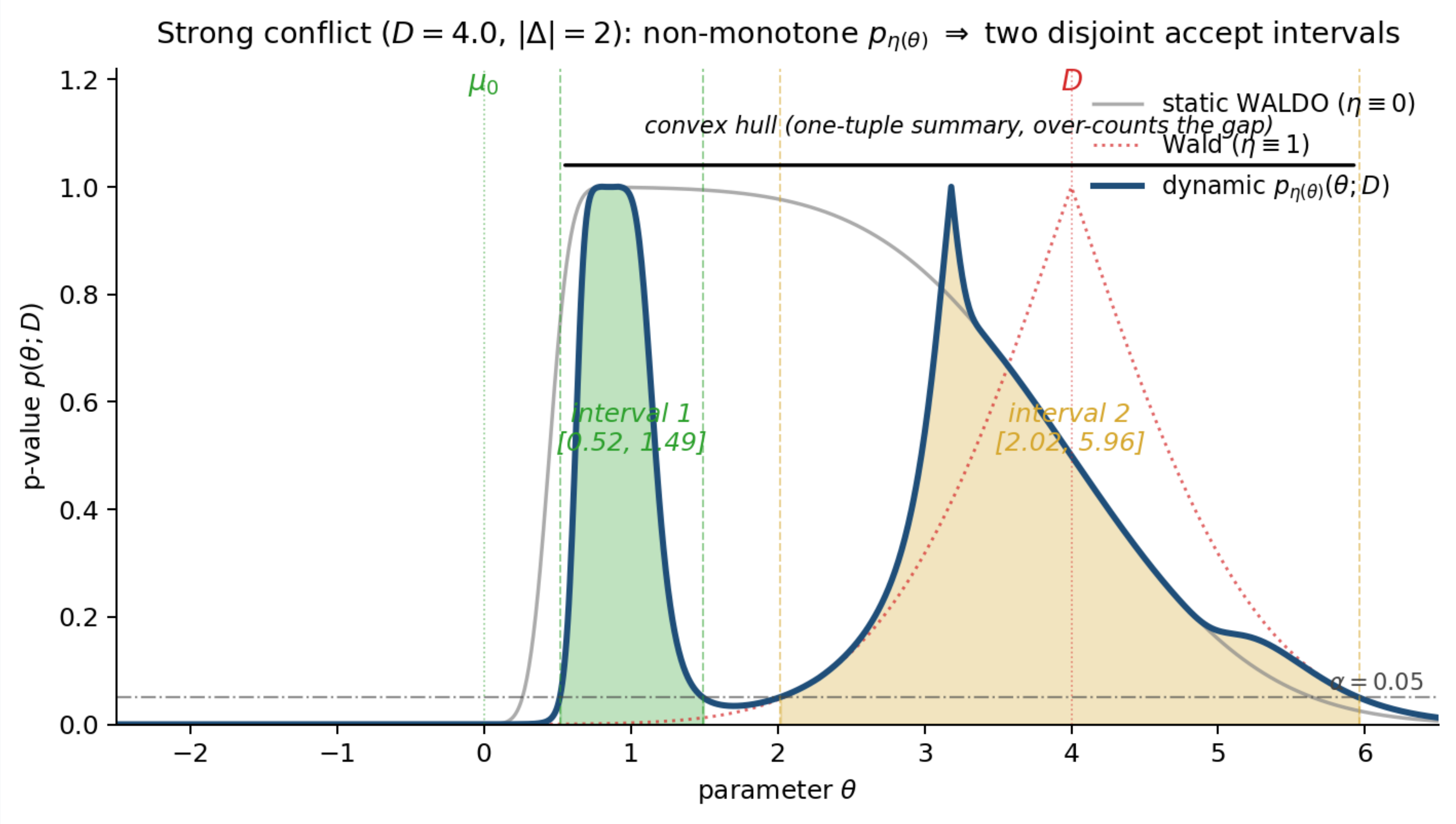


- **Fix:** let η depend on θ alone — **not D**
- p-value at each fixed η stays U[0,1] ✓ **calibrated**

- Natural shape: U-curve — use prior near μ_0, drop it at extremes?
- Would allow CD to **adapt** to local prior–data conflict

Connection: η(θ) is the θ-dependent fusion weight in a Schweder–Hjort fused CD. Refs: Schweder & Hjort 2016, Ch. 4 · Singh, Xie & Strawderman 2005 — confidence fusion.

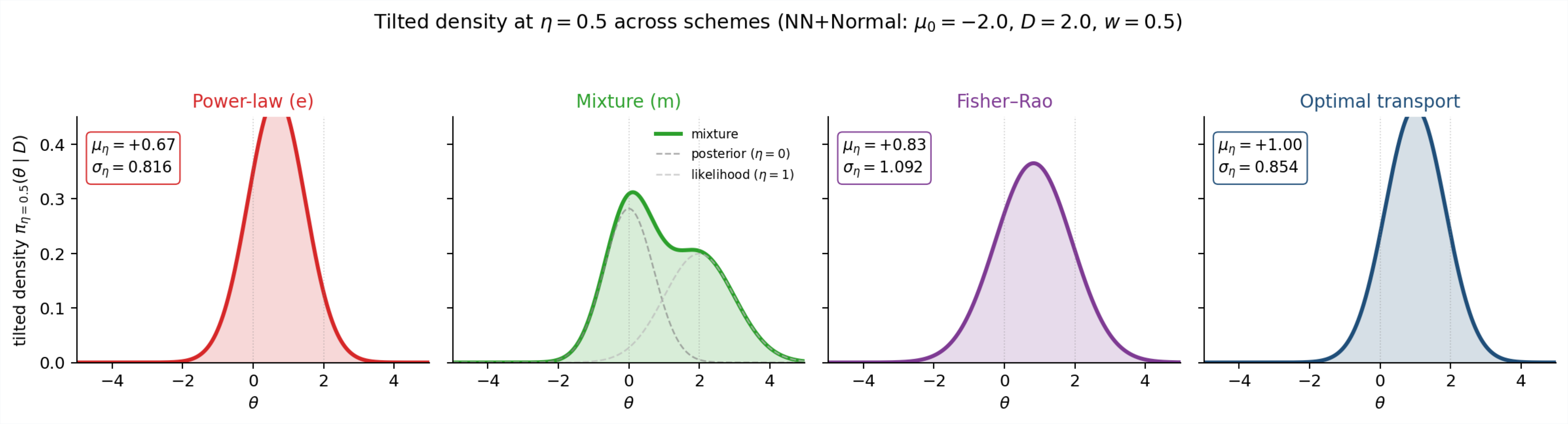
Consequences: disjoint accept regions



- Strong conflict → **non-monotone** p(θ; D)
- **Accept region:** union of **two disjoint** intervals
- One near **μ_0**, one near **D**
- Framework returns both — not a convex hull
- ***Coverage still exact*** — *the truth of the test*

Multi-region as fused-CD with strong conflict: Schweder & Hjort 2016, Ch. 4. Disjoint accept regions in prior–data conflict: Mandelkern 2002; Feldman & Cousins 1998 (related precedent in physical-sciences CIs).

Multiple ways to interpolate between prior and likelihood



PL $\pi_n \propto \pi^{1-n} \cdot L$
log-linear (e-geodesic)

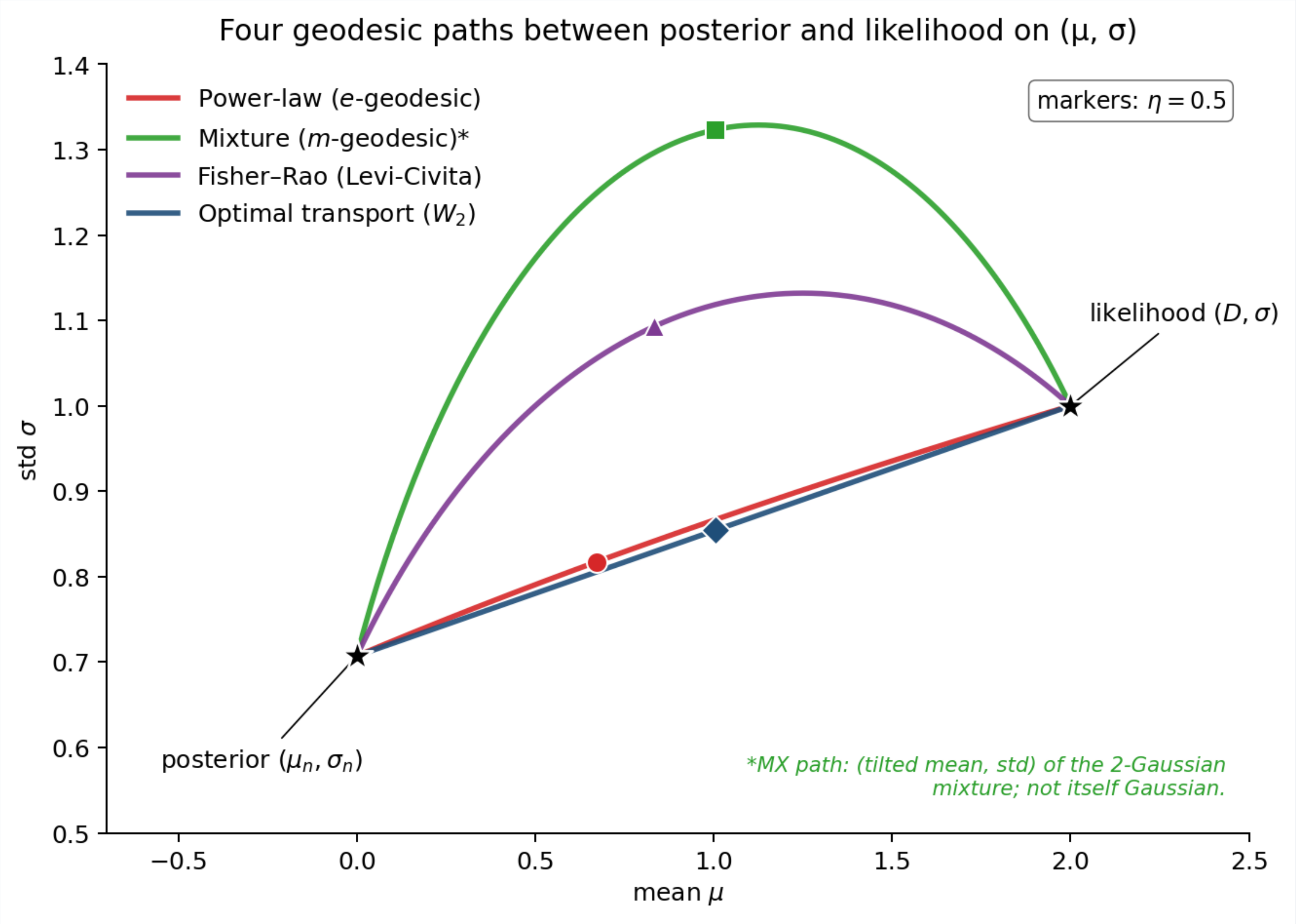
MX $\pi_n = (1-\eta)\pi_a + \eta L/Z$
linear in densities (m-geodesic)

FR $ds^2 = (d\mu^2 + 2d\sigma^2)/\sigma^2$
Fisher–Rao metric on (μ, σ)

OT $F^{-1}{}_n(u) = (1-\eta)F^{-1}{}_a + \eta F^{-1}{}_l$
linear in quantiles (W_2)

Amari & Nagaoka 2000 — e-/m-geodesics · Villani 2003 — W_2 optimal transport

Each family is a path through distribution space

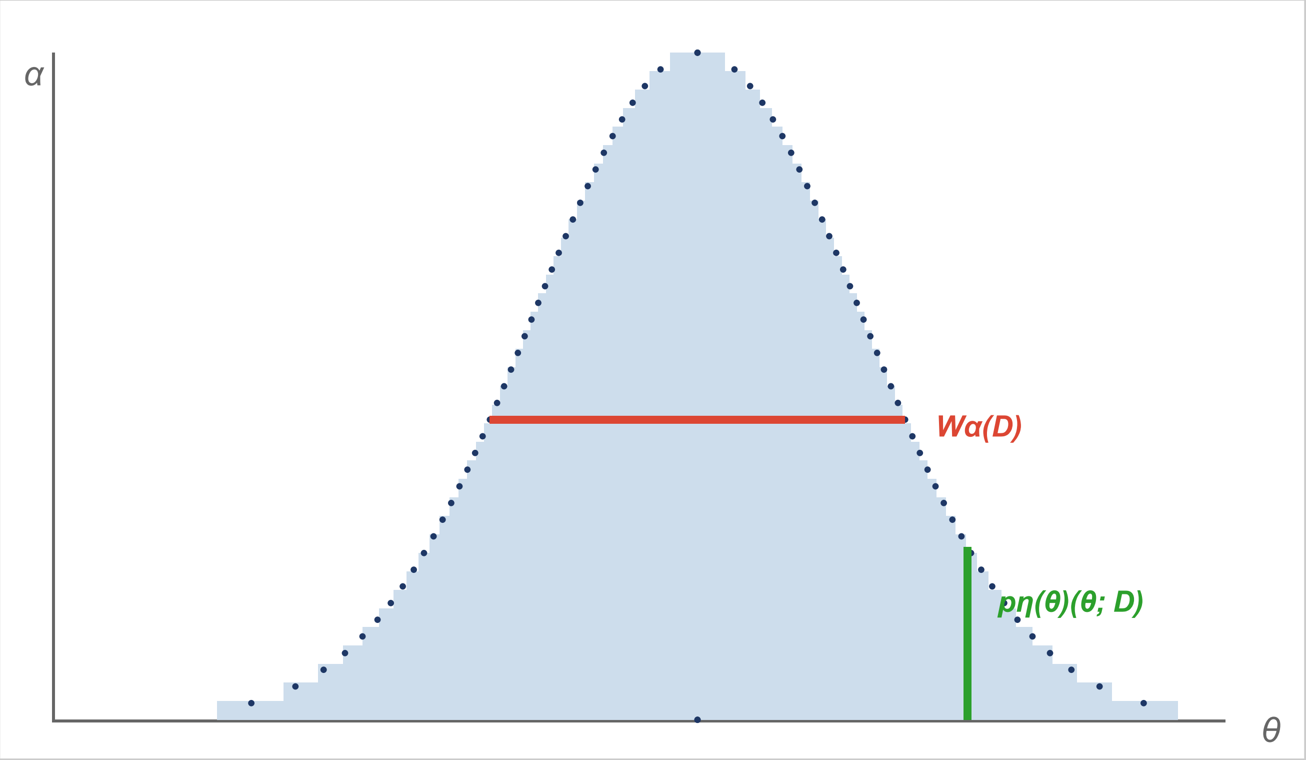


- η traces a **path** from posterior to likelihood
- **PL** ≈ **OT** here — Gaussian coincidence
- **MX** bulges up: mixture variance > either endpoint
- **FR** arcs through Fisher–Rao curvature
- Choice = **smoothness × admissibility**, not validity

Amari & Nagaoka 2000 — e- and m-geodesics on statistical manifolds · Villani 2003 — optimal transport, W_2 geodesics · Csiszár 1975 — information divergences.

What objective are we optimizing?

- A good cross-θ functional is **average CI width over a calibration set**
  - Swap order of integration **(Fubini)**
  - → trainable as **integrated p-value**



integrate by rows = average width · by columns = integrated p-value

Fubini swap

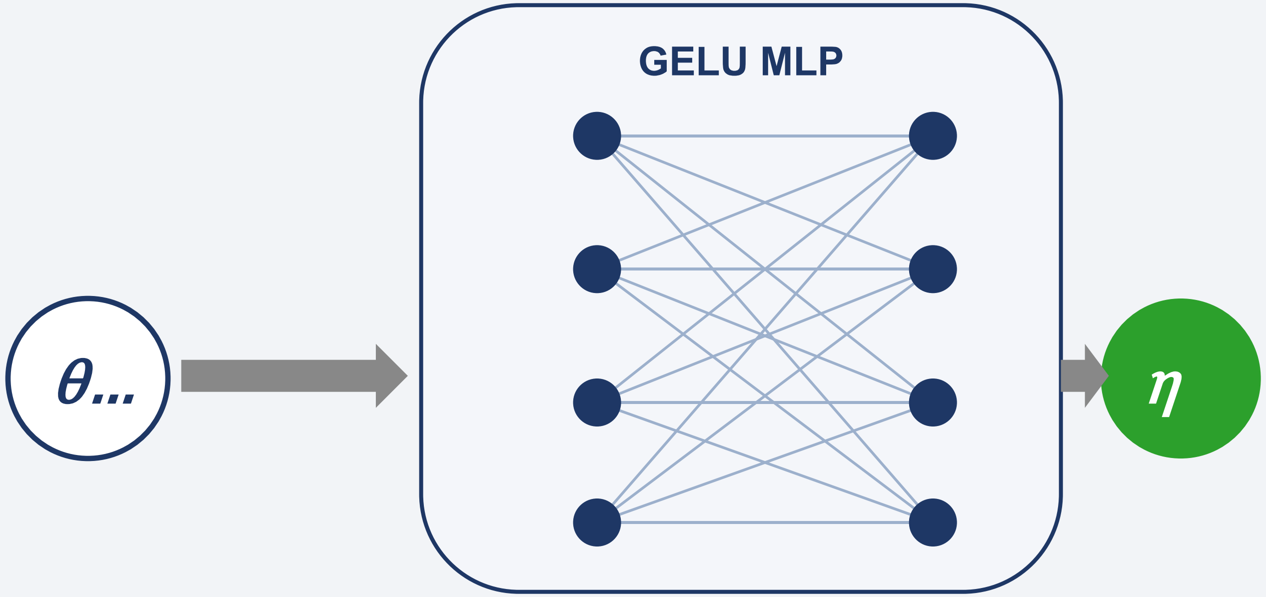
$$\int_{\theta} p_{\eta(\theta)}(\theta; D)\, d\theta = \int_{0}^{1} W_{\alpha}(D; \eta)\, d\alpha$$

integrated p-value over θ = *average CI width across α-levels*

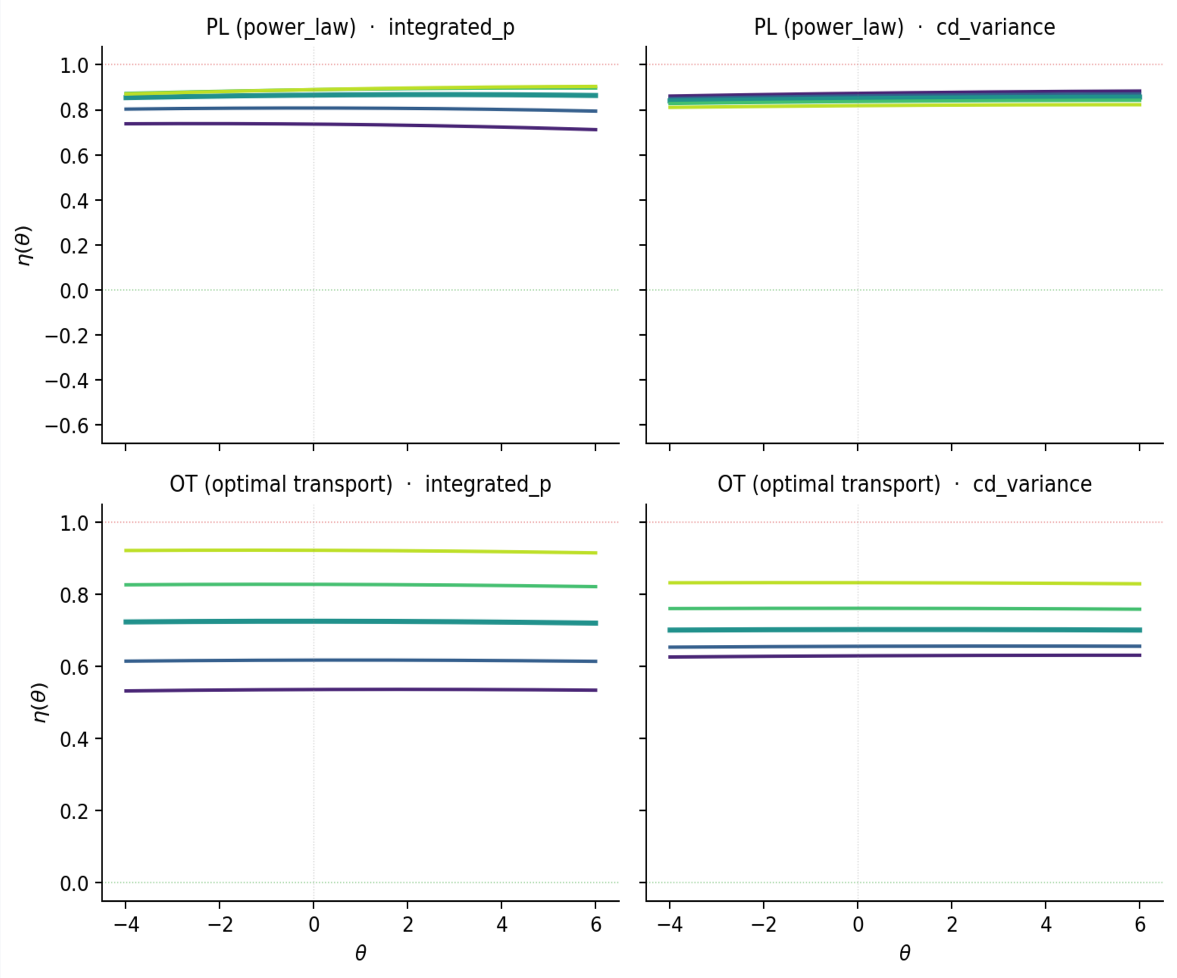
Alternative losses for ablation: `cd_variance` *(variance of the CD density)* · `static_width` *(CI width at a reference θ_0)*

Parameterize η(θ) with a small neural net

- η(·) is a **function** — needs an approximator
- Small **GELU MLP** mapping θ → η
- Condition on **(μ_0, σ_0, σ)** — one net per regime
- Second head **ValidityNet**: soft admissibility



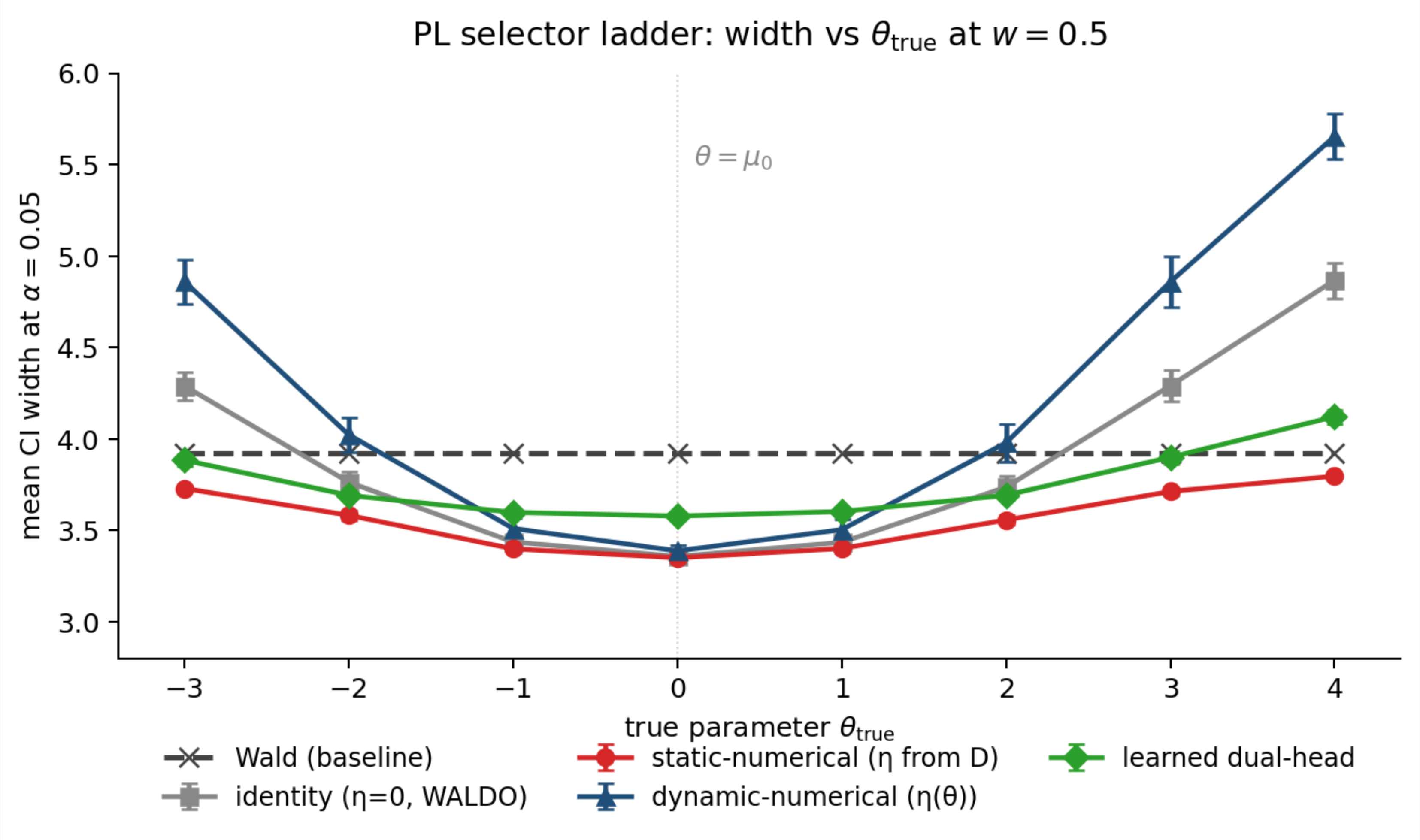
Both schemes adapt — just at different η



Top (PL): η lives near its admissibility cap (~0.8) · **Bottom (OT):** η spans 0.4–1.0 across σ_0

Same headline (calibrated, narrower than Wald) — different transit point between posterior and likelihood

Calibrated and narrower than Wald



Coverage ✓ nominal at every θ

Width ✓ 3–9% narrower than Wald at |θ| ≤ 2

~5% wider only at |θ|=4 — no post-selection tax

(the static-numerical and dynamic-numerical show what happens in this case)

Only the learned selector hits this target

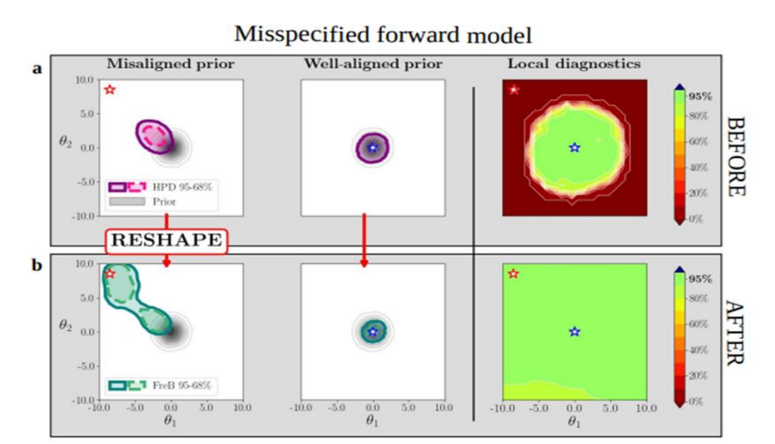
Summary and outlook

The reframe

- Bayesian **shape** + frequentist **calibration** — not an either/or
- Treat the prior as **flexible meta-belief** — a dial, not a commitment

What we built

- **η-tilting**: a tunable interpolation between posterior and likelihood
- **η(θ)** learned globally — calibration preserved by Neyman inversion
- Result: **calibrated and tighter than Wald** — geometry-agnostic



What's next

- SBI under misspecification — *Carzon et al. 2025, Ingram et al. (in prep.)*
- Multi-dim θ, non-conjugate likelihoods, OOD detection

Figure: Carzon et al. 2025 — frequentist reshaping (FreB) of Bayesian credible regions under misspecified forward models

References / Q&A

WALDO and Frasian lineage

- Masserano, Dorigo, Izbicki, Kuusela, Lee 2023 — *WALDO*
- Dalmasso et al. 2020 — *LF2I*
- Fraser 1968 — *structural inference*
- Dempster 1974, 1997 — *posterior likelihood ratio*
- Aitkin 1991, 1997 — *posterior Bayes factors*
- Wasserman 2011 — *coined "Frasian"*

Post-selection inference

- Berk, Brown, Buja, Zhang, Zhao 2013 — *valid PSI*
- Fithian, Sun, Taylor 2014 — *exact PSI*

Confidence distributions & fusion

- Schweder & Hjort 2016 — *Confidence, Likelihood, Probability*
- Xie & Singh 2013 — *CD review*
- Singh, Xie, Strawderman 2005 — *confidence fusion*

Geometry of distributions

- Amari & Nagaoka 2000 — *Methods of Information Geometry*
- Villani 2003 — *Topics in Optimal Transportation*
- Behboodian 1970 — *Gaussian mixture bimodality*

Prior-data conflict

- Evans & Moshonov 2006; Marshall & Spiegelhalter 2007

Applications & ongoing work

- Johnson, Leja et al. 2021; Leja et al. 2017 — *Prospector*
- Wang et al. 2023 — *Prospector-β*
- Suess et al. 2022; Narayanan et al. 2024 — *atypical galaxies*
- Carzon, Speagle, Lee, Izbicki, Masserano et al. 2025 — *FreB / trustworthy SBI*
- **Ingram et al. (in prep.) —** *selection/calibration*
- Cranmer, Brehmer, Louppe 2020 — *SBI overview*

Foundations

- Casella & Berger 2002 — *Statistical Inference (textbook)*