

What a Memory Must Preserve

Machine Reasoning over Uncertain Observations

Michael P. Laing

`mllaing@post.harvard.edu`

ORCID 0009-0004-5593-5759

Prepared with drafting assistance from Anthropic Claude Fable 5.1

Licensed under Creative Commons Attribution 4.0 (CC BY 4.0)

September 28, 2026

Abstract

A reasoning system that acts in the world computes over remembered measured quantities: sensor readings, assay results, positions, calibrations. The companion paper [1] establishes when collapsing such a quantity to a point is guaranteed harmless — for every joint input law exactly when the downstream computation is affine, and for every downstream computation exactly when the quantity was exact — and estimates, to second order, the systematic error otherwise. This note treats the consequence that paper defers: what happens when the downstream computation is a *learned* function. The central observation requires no new mathematics. A consistent supervised learner trained on point-valued records estimates the collapsed regression function, and under the observation model of [1, Prop. 4.1] that function is, in closed form, the artifact: the classical attenuation of a regression slope by the instrument’s reliability ratio, which in the original coordinate is a concave power law absent from the system generating the data. More data and a richer, still complexity-controlled model class therefore produce faster and more faithful convergence *to the artifact*; better optimization leaves the target where it is. We state this precisely, separate the two distinct repairs it admits — calibrating the input versus computing from the posterior — and show that the geophysical correction of [2] succeeded by input calibration alone because the true response happened to be linear. We then draw the consequences for the memory layer of a reasoning system: what it must preserve, what calibration formally requires of it, and how conformance can be gauged. What is claimed is the location of the failure — the memory store, at write time — and the specification that repairs it; not the attenuation, which is classical, and not the behaviour of any deployed system.

1 The missing layer

Architectures for machine reasoning name memory as a pillar: a module that persists state between perception and planning, so that action can be conditioned on what was measured rather than only on what is currently perceived [10]. What the memory *contains* is left unspecified, and current practice supplies the default inherited from every record system since the first file format: the memory stores points. A quantity that arrived as a measurement — an instrument reading with a stated error model, an estimate with a posterior — is written

down as a number, and everything known about the number beyond its value is discarded at the moment of writing.

The companion paper [1] analyzes this act in isolation from learning. Its objects are minimal: a stored quantity is a probability measure (a *cell*) on a metrized domain D ; dependence between quantities is a coupling, carried as data; computation is pushforward of the joint measure; and *collapse* π — there taken to be the mean — maps a cell to a point. Its central theorem [1, Thm. 3.2] states that collapse commutes with a numerical derivation for every joint input law if and only if the derivation is affine, and [1, Cor. 3.4] that it commutes with every measurable derivation if and only if the cells are already exact (Dirac). Between the two sits the quantitative instrument [1, Rem. 3.5]: for a derivation f twice continuously differentiable with bounded second derivative on a neighbourhood of m containing the support of the input law, and an input concentrating at the collapsed point m — the law of $m + \varepsilon Z$, Z centered with finite second moment, $\varepsilon \rightarrow 0$ — the discrepancy, compute-then-collapse minus collapse-then-compute, is

$$\frac{1}{2} \operatorname{tr}(\nabla^2 f(m) \Sigma_\varepsilon) + o(\varepsilon^2), \quad \Sigma_\varepsilon = \varepsilon^2 \operatorname{Cov}(Z),$$

a bias, not noise: signed by curvature, scaled by the discarded covariance, and unmoved by increasing the number of sampled cases while the measurement-error mechanism is unchanged.

None of that involves learning. This note adds the learning layer, and the addition is short, because the hard work is already done: a learned function is a numerical derivation, and the theorem does not care how the derivation was obtained. What learning contributes is not a new failure mode but an industrialization of the old one — and one genuinely new twist: the learner does not merely *apply* a curved function to collapsed memory; it *estimates its target from* collapsed memory, so the artifact enters twice: through the training pairs, whose regression function it is, and through every application of the trained function to a collapsed input.

2 What a learner learns from point-valued memory

Fix the observation structure of [1, Prop. 3.7 and Prop. 4.1], the simplest honest model of an instrument. A true quantity X has a prior; the record holds not X but an observation W related to it by a known error mechanism; a downstream quantity Y depends on X through the true response, up to outcome noise. The record layer stores the pair (W, Y) — the reading, taken at face value, and the outcome.

A supervised learner is now trained to predict Y from the record. The population version of its objective settles what it can converge to.

Proposition 1 (The learner’s target). *Let (W, Y) be square-integrable. Over all measurable $h: \mathbb{R} \rightarrow \mathbb{R}$ with $\mathbb{E}[h(W)^2] < \infty$ — outside this class the objective is infinite, so the argmin is unaffected —*

$$\arg \min_h \mathbb{E}[(Y - h(W))^2] = \mathbb{E}[Y \mid W = \cdot],$$

the collapsed regression function. In particular, under the multiplicative observation model of [1, Prop. 4.1], extended by outcome noise — $\log X \sim N(m, s^2)$, $\log W = \log X + \eta$ with $\eta \sim N(0, \sigma^2)$, and $\log Y = \log \beta + \log X + \zeta$ with $\zeta \sim N(0, \tau^2)$,¹ the three mutually independent,

¹Here τ is the outcome-noise scale, not the prior scale of [1, Prop. 3.7].

so that the true response $\mathbb{E}[Y \mid X = x] = \beta e^{\tau^2/2} x$ is exactly linear — the target is, in closed form,

$$\mathbb{E}[Y \mid W = w] = C e^{\tau^2/2} w^\lambda, \quad \lambda = \frac{s^2}{s^2 + \sigma^2} \in (0, 1),$$

with C as in [1, Prop. 4.1], which is the case $\tau = 0$: a strictly concave power law whose marginal response vanishes at large w [1, Cor. 4.2], though the true response is linear. On the log scale this is the classical attenuation of the regression slope by the reliability ratio λ — regression dilution [11, 5], $\mathbb{E}[\log X \mid \log W = u] = m + \lambda(u - m)$ — and the power law is that attenuation carried back to the original coordinate.

Proof. The first identity is the L^2 projection characterization of conditional expectation: for any candidate h , $\mathbb{E}[(Y - h(W))^2] = \mathbb{E}[(Y - \mathbb{E}[Y \mid W])^2] + \mathbb{E}[(\mathbb{E}[Y \mid W] - h(W))^2]$, minimized uniquely (up to W -null modification) at $h = \mathbb{E}[Y \mid W = \cdot]$. The closed form is [1, Prop. 4.1] times $\mathbb{E}[e^\zeta] = e^{\tau^2/2}$, the outcome noise being independent of (X, W) . $\square$

Corollary 2 (Consistency is convergence to the artifact). *Let $\hat{h}_n$ be an estimator trained on n i.i.d. records (W_i, Y_i) that is L^2 -consistent for the regression function: $\int (\hat{h}_n - \mathbb{E}[Y \mid W = \cdot])^2 d\mathcal{L}(W) \rightarrow 0$ in probability. Then under the model of Proposition 1, $\hat{h}_n$ converges in $L^2(\mathcal{L}(W))$ to the concave power law $C e^{\tau^2/2} w^\lambda$.*

Universally consistent estimators — partitioning, kernel, and nearest-neighbour regression under Stone’s conditions [3, 4] — are L^2 -consistent for every law with $\mathbb{E}[Y^2] < \infty$, and empirical-risk minimizers over a sieve are L^2 -consistent when the sieve is dense in $L^2(\mathcal{L}(W))$ and its complexity grows slowly enough in n [4, Ch. 10–12]. More data and a richer, still complexity-controlled model class sharpen the convergence — *to the artifact*.

Three readings of this corollary, in increasing order of consequence.

First, the artifact is the ceiling of learning, not a small-sample pathology. The concavity is a property of the population law of the records, so no quantity of data removes it; the estimator’s L^2 error decomposes into estimation error — distance to $\mathbb{E}[Y \mid W]$ — plus Bayes error under the record, and learning effort spends itself entirely on the first term while the second, and the shape of the target, are fixed by the record layer. The learner is behaving correctly. It is answering, with increasing precision, a question the memory has already answered wrongly.

Second, the statistician’s defense is structurally absent. The statistical literature’s repair for errors-in-variables presumes an analyst who suspects the problem and reaches outside the record for the error model. A training pipeline is the industrialization of not suspecting: the record is the interface, the error model is not in it, and the pipeline’s entire design goal is faithful fit to the record. Where the vigilant analyst was an unreliable defense, the pipeline is a reliable non-defense. What the analyst reaches for is a number the record does not hold — the reliability ratio λ , or equivalently the error variance σ^2 — and §4 shows that, once the outcome is noisy too, no functional of the record’s law supplies it.

Third, the second entry iterates: the artifact re-enters the record. The trained model’s outputs — predictions, scores, imputations — are written back as points, becoming inputs to further inference and further training. Each subsequent curved consumer of those points pays its own second-order correction [1, Rem. 3.5], and the corrections do not cancel: they are signed by curvature, and the sign structure of a pipeline is not random. The bias-propagation calculus of Bouleau [9], in which a bias recursion (first derivative times inherited bias, plus

half the second derivative times variance) rides alongside the variance recursion, is the correct bookkeeping for such chains, and it presupposes exactly what the record layer discarded: the variance to put in the recursion.

Remark 3 (Dependent records). Corollary 2 assumes i.i.d. records, while the core’s motivating couplings — shared instruments, calibrations, clocks — make real records dependent. If the record process is stationary, the marginal law of (W_i, Y_i) , and hence the target $\mathbb{E}[Y | W = \cdot]$, is the same as in the i.i.d. case; dependence changes the estimators for which consistency is known and the rates at which it holds, not the target. Where consistency holds for stationary mixing or ergodic records [4, Ch. 27], it holds to the same artifact. The coupling the memory failed to carry does not rescue the learner that trained without it.

Remark 4 (Outcome-side error). The model of Proposition 1 carries outcome noise, and the target barely sees it: additive noise $Y' = Y + \zeta$ with $\mathbb{E}[\zeta | W] = 0$ and $\mathbb{E}\zeta^2 < \infty$ leaves the target unchanged, $\mathbb{E}[Y' | W] = \mathbb{E}[Y | W]$; the multiplicative noise of the model scales it by the constant $e^{\tau^2/2}$; in either case the noise, when independent of (X, W) as in the model, or for additive noise merely uncorrelated with Y given W , inflates only the variance *around* the target. Outcome noise widens the scatter about the target; input collapse *moves* the target. The pathology this note treats is input-side by nature, not by simplification. The outcome noise is kept in the exhibit for a different reason, given in §4: it is what makes the record’s law insufficient to recover the response by regression in either direction.

3 Two repairs, and why they are not the same

Proposition 1 admits two distinct repairs, and the distinction is invisible in the flagship empirical case precisely because that case is the lucky one.

The first repair is *input calibration*: replace the recorded w by the collapsed posterior $\mu(w) = \mathbb{E}[X | W = w]$ and apply the true (or separately estimated) response to it. This is regression calibration [5], the modern form of Spearman’s attenuation correction, and it is what [2] applied to recover linearity from twenty-five years of geophysical records.

The second repair is *posterior computation*: carry the whole conditional cell $\mathcal{L}(X | W = w)$ and compute the pushforward mean $\mathbb{E}[g(X) | W = w]$ — the object the memory layer of [1] computes by construction.

Proposition 5 (The repairs differ by the second-order correction). *Let the response satisfy $\mathbb{E}[Y | X, W] = g(X)$,² with g twice continuously differentiable with bounded second derivative, and write $\mu(w) = \mathbb{E}[X | W = w]$, $v(w) = \text{Var}(X | W = w)$. Then, to second order in the posterior dispersion (rigorously, for a family of posteriors concentrating at $\mu(w)$, in the sense of [1, Rem. 3.5]),*

$$\underbrace{\mathbb{E}[g(X) | W = w]}_{\text{posterior computation}} - \underbrace{g(\mu(w))}_{\text{input calibration}} = \frac{1}{2} g''(\mu(w)) v(w) + o(v(w)).$$

Fix an error model with likelihood $k(w | x)$ whose ratio $k(w | x)/k(w | y)$ takes every value in $(0, \infty)$ as w varies, for each $x \neq y$ (the Gaussian additive and multiplicative models of [1,

²The outcome depends on the reading only through the quantity, and any outcome noise has conditional mean zero. The exact case is $Y = g(X)$; the model of Proposition 1 with $\tau > 0$ is an instance, with $g(x) = \beta e^{\tau^2/2} x$.

Props. 3.7 and 4.1] do). Then $\mathbb{E}[g(X) \mid W = w] = g(\mu(w))$ for every prior with finite second moment and every w if and only if g is affine.

Proof. Taylor expansion of g about $\mu(w)$ under the conditional law, exactly as in [1, Rem. 3.5] with the posterior cell in the role of the input cell. Sufficiency of affinity is the easy direction of [1, Thm. 3.2] applied conditionally. For necessity, take the two-point prior at $x \neq y$: the posterior at w is two-point with weight $\lambda(w) \in (0, 1)$ on x , and $\lambda(w)$ ranges over $(0, 1)$ as w varies, by the likelihood-ratio hypothesis. The identity then forces $g(\lambda x + (1 - \lambda)y) = \lambda g(x) + (1 - \lambda)g(y)$ for all $\lambda \in (0, 1)$ — Jensen’s functional equation, the two-point argument in the proof of [1, Thm. 3.2] — so g is affine. $\square$

Outside the boundedness hypothesis the correction can still resum exactly, needing no Taylor control at all: for $g = \exp$ with Gaussian posterior in the exponent — the latent taken to be $\log X$, the repairs compared being $\exp(\mathbb{E}[\log X \mid W = w])$ and $\mathbb{E}[X \mid W = w]$ — it is the Gaussian moment generating function, contributing the factor $e^{\text{Var}(\log X \mid W = w)/2}$ exactly, as in [1, §4.1]. Affinity of g is a property of the coordinate in which the posterior mean is taken. The model of Proposition 1 is affine in X and exponential in $\log X$: input calibration in X — the coordinate [2] used — is exact, while input calibration in $\log X$ pays the factor $e^{\lambda\sigma^2/2}$ that [1, §4.1] isolates in the constant C . The first repair is a special case in a second sense: it requires a coordinate in which the response is affine, and a learned response offers none.

The reading of Proposition 5 matters more than its (elementary) content. The geophysical correction succeeded by input calibration alone *because the true response was linear*: $g = \beta \text{id}$ makes the correction term vanish identically, so collapsing the posterior after conditioning was free — [1, Thm. 3.2] operating in the analyst’s favor for once. Had the response been curved — a threshold, a ratio, a learned model — calibrating the input would have repaired the observation failure while leaving the transformation failure intact, and the two failure modes, which [1, §4.1] shows are mathematically distinct effects of the single decision to replace a measure by a point, would have required the full posterior. The exhibit of Proposition 1 is, for the same reason and in the coordinate X in which g is affine, entirely the observation failure: the correction of Proposition 5 vanishes, and what the learner learns is $\beta e^{\tau^2/2}$ times the posterior mean of X — attenuation — and nothing else. The factor $e^{\lambda\sigma^2/2}$ that [1, §4.1] books as the curvature of $\exp$ is the transformation failure of the other coordinate, $\log X$, as above; [1, Thm. 3.2] bites only where the response, or the use made of the learned response downstream, is curved in the coordinate the memory holds. A reasoning system cannot assume its downstream uses are linear; its uses are, overwhelmingly, learned and curved. For such a system the first repair is a special case, and the second is the specification: *the memory must carry the cell, because no shallower object supports the general repair.*

4 Calibration, and what it formally requires of memory

A system is calibrated — in the forecaster’s sense [6, 7, 8], distinct from the input calibration of §3 and from the instrument calibrations that couple records — when its expressed confidence about a quantity matches its epistemic state about that quantity. Which quantity matters. Confidence about the outcome given the record, $\mathcal{L}(Y \mid W = w)$, is a functional of the population law of the records and is learnable from point-valued records by any consistent conditional-

distribution estimator; a point-valued memory does not prevent calibration about Y . Confidence about the remembered quantity X is different, for two reasons the record layer cannot cure.

Proposition 6 (Constancy on fibers). *Let a record be a measurable map r from states of knowledge $\mu \in \mathcal{P}(D)$, D the metrized domain of §1, to a record space R , and let a confidence rule be any measurable $c: R \rightarrow [0, 1]$, however trained. If $r(\mu_1) = r(\mu_2)$ then $c(r(\mu_1)) = c(r(\mu_2))$; hence for any target functional $q: \mathcal{P}(D) \rightarrow [0, 1]$ (for instance $q(\mu) = \mu(A)$ for a Borel set $A \subseteq D$) with $q(\mu_1) \neq q(\mu_2)$, the rule is miscalibrated on at least one of μ_1, μ_2 .*

Proof. $c \circ r$ is constant on the fiber $r^{-1}(r(\mu_1))$. □

The point-valued record is the map $\mu \mapsto \pi\mu$ (or $\mu \mapsto \delta_w$ at observation), whose fibers contain every measurement state with the same collapsed value, tight or vague. The proposition bites when the record holds distinct measurement states in one fiber — heterogeneous instruments, calibrations, exposures — with no recorded proxy for the dispersion. It says nothing when dispersion is homogeneous across records, and there the second reason applies: the posterior $\mathcal{L}(X \mid W = w)$ depends on the error model, which the record does not carry, and the error model is not in general a functional of the record’s population law: with the response and the error mechanism both unconstrained, any reparametrization of X leaves $\mathcal{L}(W, Y)$ unchanged. The model of Proposition 1 shows where the boundary lies. Let the analyst know the Gaussian forms of the prior and of the errors but not the response, $\log Y = \log \beta + b \log X + \zeta$ with b unknown (the model has $b = 1$). With the outcome exact, $\tau = 0$, the record identifies everything: the reverse regression of $\log W$ on $\log Y$ has slope exactly $1/b$, and then $\sigma^2 = \text{Var} \log W - b^{-1} \text{Cov}(\log W, \log Y)$ — through structure imported from outside the record, namely that the outcome is the exact variable. With outcome noise, $\tau > 0$, the record’s law — Gaussian, so nothing beyond its moments — fixes three second moments, $\text{Var} \log W = s^2 + \sigma^2$, $\text{Var} \log Y = b^2 s^2 + \tau^2$, $\text{Cov}(\log W, \log Y) = b s^2$, against four unknowns, and b is confined to the interval between the direct regression’s slope, $b\lambda$, and the reciprocal of the reverse regression’s, $b + \tau^2/(b s^2)$ — the Gini–Frisch bounds [12, 13] — with no functional of $\mathcal{L}(W, Y)$ picking the point (the Gaussian case of Reiersøl [14]); the error model, σ^2 or τ^2 , picks it. Either way, whatever confidence about X a point-valued system expresses is borrowed from structure the memory did not store. A confabulated confidence is the memory-layer analogue of a hallucination: an epistemic state reported as certain.

The measure-valued memory of [1] is the structure on which calibration about the remembered quantity is expressible, and three of its operations are the ones a reasoning loop needs.

Conditioning is the operation that earns confidence. Repeated measurement of the same quantity, honestly combined, shrinks the cell in expectation, and with it the covariance every downstream second-order correction consumes; where a stage’s Hessian is definite the correction is monotone in that covariance. This is an operation the measure-carrying model admits and the collapsed one does not [1, §3.1]; its specification — the version of the conditional kernel and where the prior lives in the schema — is [1, Prob. 4], and it is not among the five operations closed in [1, Prop. 2.10]. A million readings recorded as points can be averaged, and the average is a better point; what the point-valued layer cannot record is what the averaging earned — the posterior width — nor that the readings were of one quantity, which is the coupling. A million observations conditioned into one cell carry both. The averaged point does not escape Proposition 1 either: a stored mean of a few readings is a reading with a smaller error variance,

and its reliability ratio is still below one — as documented in quantitative genetics, where selection gradients fitted to individual means of a few repeated measurements remain attenuated [15]. A memory that cannot condition cannot get more confident for the right reasons.

The decision operator is where deferral ends. An action is a point; collapse is deferred to the moment of decision, not eliminated [1, Prob. 6]. The formal object that problem poses is $\arg \min_a \mathbb{E}[\ell(a, X)]$ under the carried joint measure — and it is decision, not reporting, that forces a point: a distribution can be *presented* (a density, a quantile band) by marginalizing the answer cell, which preserves its law, discards only the carried coupling, and is strictly finer than collapsing it. In this vocabulary any output that must be a point is a decision; the reporting layer as such never needs collapse.

Audited collapse is the log. When a collapse does occur, the memory should record — the provenance [1, Prob. 6] asks for — what was collapsed, under what loss, with what dispersion discarded, turning the dangerous operation into an auditable one. For a learning pipeline the audit has a precise target: the accumulated second-order corrections along the chain, which can be estimated to leading order exactly where the memory still holds the covariance they consume [1, Rem. 3.5]; a non-asymptotic single-stage bound and a composition rule for the chain are the open items of [1, Prob. 5].

5 Conformance for learning pipelines

The companion paper is a specification, not a system: it states what any conforming memory must preserve and provides the instruments to gauge what an implementation loses [1, §5]. For learning pipelines the conformance questions become concrete, and they are checks, not aspirations:

Dirac conformance. On exact data — Dirac cells — and for a fixed seed, data order, and numerics, the pipeline must produce the output of its point-valued counterpart, exactly: the learning-layer form of the Dirac, discrete-metric restriction of [1, Prop. 2.7], which is stated there for the algebra and is invoked here by analogy. A pipeline that behaves differently on certainty has invented information.

Affine transparency. Through affine stages — sums, scalings, weighted aggregations, linear layers taken alone — means and covariances must transport exactly, with no approximation budget spent: [1, Thm. 3.2] marks the region where exactness is achievable, and an implementation that pays error there is nonconforming by choice.

Curvature accounting. At every non-affine stage — and a trained network interleaves them with its affine layers — the implementation owes an estimate of the second-order correction it induces, with [1, Rem. 3.5] as the estimator and the carried covariance as its input. The benchmark shape is inherited from [1, Prob. 7]: synthetic tasks with known ground truth. Under the model of Proposition 1 the pass criteria are slopes, not equalities: the fitted log–log slope of the point-valued pipeline’s prediction against the recorded reading w must converge to λ , and the slope of the response a conforming pipeline learns for the carried quantity X must converge to 1, each within a stated tolerance as n grows; and at each curved stage the reported second-order estimate must match the measured gap between the stage’s output mean and the pushforward mean to the order of the discarded variance. Proposition 1 predicts exactly what the point-valued pipeline will learn; the check is whether the conforming pipeline does not.

These three are tests: behavioural, run on benchmark shapes, and what they certify is an implementation. They do not certify a result. That is the audit of [1, Prob. 6 and §5]: a presented result carries the term of the algebra that produced it, and a small reference evaluator re-derives it from that term within the tolerance the plan declares. A pipeline can pass every test above and present a result whose derivation no one can state, and a term can check statically and still be evaluated badly; the per-stage estimate the third check owes is what the term carries; neither certificate does the other’s job, and a conforming pipeline owes both.

6 Relation to uncertainty in representation

Predictive architectures that operate in a learned latent space handle irreducible uncertainty by *abstraction*: the encoder discards detail that cannot be predicted, and a latent variable ranges over the plausible futures that remain [10]. That is an answer for the representation layer, and this note does not touch it. The stored-quantity layer is a different place with a different correct answer: *conservation* — carry what cannot be resolved, as a measure, with its dependence, and collapse only where [1, Thm. 3.2] licenses it or where a decision demands it. That layer is not hypothetical for a reader who runs an embedding store: [1, Prob. 8] treats the store as exactly this layer — each vector a Dirac cell on a metrized domain, its encoder’s dispersion observable at write time and recorded nowhere — so a retriever tuned against stored point embeddings is in the situation of Proposition 1, with the embedding in the role of the observation W . The two commitments are the same epistemic stance applied at two layers, and they compose: an architecture that abstracts in representation while its memory stores measured quantities as points has solved uncertainty where it is cheap to acknowledge and reintroduced it where it is silent. The regime where the composition matters most is the one the representation layer does not address: when what is uncertain is not the dynamics but the remembered *state* — the assay value, the field strength, the position — epistemic uncertainty lives in the memory or nowhere, unless the architecture treats state estimation itself as a prediction and routes its uncertainty through the latent, a choice [10] leaves open rather than makes.

The closing register of [1] applies here with one word changed. Point-valued practice is not an approximation of correct machine reasoning over uncertain observations; it is a different computation, and Proposition 1 exhibits the difference as the thing a learner will, with perfect fidelity and increasing confidence, learn.

Canonical reference for this paper and its companion: <https://realql.org>.

References

- [1] M.P. Laing, “Measure-Relational Algebra: The Core,” 2026, <https://doi.org/10.5281/zenodo.22999911> (all versions); served at https://realql.org/mra_core.pdf. Canonical reference: <https://realql.org>.
- [2] N. Sivadas, D. Sibeck, V. Subramanyan, M.-T. Walach, D.S. Ozturk, B. Ferdousi, B. Michotte de Welle, “Regression to the mean can explain saturation of geomagnetic storms,” *Nature* **655**, 1143–1147 (2026), doi:10.1038/s41586-026-10757-4.

- [3] C. J. Stone, “Consistent nonparametric regression,” *Ann. Statist.* **5**, 595–645 (1977).
- [4] L. Györfi, M. Kohler, A. Krzyżak, H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer, 2002.
- [5] R. J. Carroll, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed., Chapman & Hall/CRC, 2006.
- [6] A. P. Dawid, “The well-calibrated Bayesian,” *J. Amer. Statist. Assoc.* **77**, 605–610 (1982).
- [7] T. Gneiting, F. Balabdaoui, A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *J. R. Statist. Soc. B* **69**, 243–268 (2007).
- [8] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, “On calibration of modern neural networks,” *Proc. ICML*, PMLR **70**, 1321–1330 (2017).
- [9] N. Bouleau, *Error Calculus for Finance and Physics: The Language of Dirichlet Forms*, de Gruyter, 2003.
- [10] Y. LeCun, “A Path Towards Autonomous Machine Intelligence,” OpenReview, 2022.
- [11] C. Spearman, “The proof and measurement of association between two things,” *Amer. J. Psychol.* **15**, 72–101 (1904).
- [12] C. Gini, “Sull’interpolazione di una retta quando i valori della variabile indipendente sono affetti da errori accidentali,” *Metron* **1**(3), 63–82 (1921).
- [13] R. Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publ. No. 5, University Institute of Economics, Oslo, 1934.
- [14] O. Reiersøl, “Identifiability of a linear relation between variables which are subject to error,” *Econometrica* **18**(4), 375–389 (1950).
- [15] N. J. Dingemanse, Y. G. Araya-Ajoy, D. F. Westneat, “Most published selection gradients are underestimated: why this is and how to fix it,” *Evolution* **75**(4), 806–818 (2021), doi:10.1111/evo.14198.