

What a Deployment Correlates

Shared Causes and the Records That Cannot Carry Them

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 — draft 7

Abstract

When one model’s outputs are written into many records — across cells, tables, institutions — every entry inherits a component of the model’s error, and the entries become coupled through a cause none of them stores. This note works out three consequences, using only the machinery of [1] and [2]. First, aggregation hits a floor: averaging n model-derived entries does not converge on the truth but on the model, with residual variance equal to the shared component — old covariance algebra whose consequence here is new in kind, because the floor is computable at write time from a coupling that point-valued practice discards at write time. Second, the floor propagates: through the second-order correction of [1, Rem. 3.5], the discarded shared covariance contributes a non-vanishing bias at every consumer curved at the mean, and — the sharpest form of the result — the persistent part is carried *entirely* by the off-diagonal entries, so a record carrying per-cell uncertainties but no coupling is asymptotically as blind as a point record. Third, institutions that never exchange a byte are coupled by a shared upstream model, so every downstream computation that assumes cross-institutional independence — risk pooling, meta-analysis, ensemble voting, redundancy engineering — silently applies the product default of [1, Rem. 2.5] at deployment scale. The algorithmic-monoculture literature has established the phenomenon’s welfare consequences and, recently, its empirical reality; what it does not occupy, and what this note claims, is the storage framing: the coupling as a first-class object a record layer must carry, keyed by its cause, with the floor and the propagated bias computable from what is carried. The purest instance is n instances of one model at one version, the case now called an agentic swarm: n cells written by one instrument, their agreement one judgment at the floor, and one estimate whose error is independent of theirs priced against all of them by a computable weight. Not claimed: the magnitude of the floor or of the bias in any particular deployment, which depends on the shared component the record did not carry.

1 Introduction

The core paper [1] locates a defect at the write: collapsing an uncertain quantity to a point does not commute with nonlinear computation, and the discrepancy is systematic. Its learning companion [2] follows the defect into the trained model: a consistent learner converges to the

collapsed record’s regression function, artifact included. Both papers treat one record, one pipeline.

This note treats the third layer, which appears the moment a model is *deployed*: one instrument writing into many records at once. A clinical model imputes fields across hospital systems; a defect-detection model scores wafers across an industry’s fabs; a language model drafts text that lands in millions of documents, some of which become training data, records, and evidence. Every value so produced carries a component of the instrument’s error, and that component is *shared*: one draw, copied everywhere the instrument wrote.

The limiting case has acquired a name. Frontier development now dispatches thousands of instances of one model to collaborate on a problem [5, 27], a practice described as abstracting oversight away through sheer complexity [5]. Under the model of §2 the complexity is beside the point: n instances of one model at one version share one draw of that model’s systematic error, so their pooled judgment has a floor that n does not lower, and the record of their outputs does not carry the number that would say so. We use “swarm” only in the abstract, here, and once more to name the instance in §6; the paper’s term is many instances of one model, which is what the mathematics sees.

The mathematics of shared error is a century old and we make no claim on it. The claim of this note is about *records*. The shared component induces a coupling between cells — exactly the object [1, Def. 2.3] exists to carry — whose cause is knowable at write time, whose magnitude is estimable at write time, and which point-valued practice discards at write time. Every consequence below is a consequence of that discard, and every repair is a statement about what the record layer must carry. The deployment phenomenon itself has an established literature under the name *algorithmic monoculture*, engaged in §8; the storage framing is the residual this note occupies.

Results are numbered continuously; Propositions 1 and 3 are the load-bearing pair.

2 The shared-cause model

Fix n true quantities $X_1, \dots, X_n$ with joint covariance Σ_X , and written entries

$$W_i = X_i + \eta_i + \xi, \quad i = 1, \dots, n, \quad (1)$$

where the η_i are centered idiosyncratic errors (independent of each other and of everything else, variance σ^2) and ξ is the centered shared component (variance $\tau^2 > 0$, independent of the X_i and η_i): one draw of the deployed instrument’s systematic error, written into all n cells. Nothing requires ξ to be additive or scalar in applications; (1) is the minimal model in which every phenomenon of interest already occurs. Throughout, “entry” is the scalar W_i and “record” is the stored object that holds entries and, when it does, their coupling. We call the joint law of the errors $(\eta_i + \xi)_i$, keyed to its cause, the *error coupling*; the joint covariance Σ_n of the entries (§4) carries it together with Σ_X .

Two features distinguish the deployment setting from generic common-mode error [21, §5.2]. First, the cause of ξ is an identifiable entity — a model, at a version, under a decoding regime — so the induced coupling is *cause-keyed* in the sense of [1, Problem 2]: the error-borne covariance $\text{Cov}(W_i, W_j) - \text{Cov}(X_i, X_j) = \tau^2$ ($i \neq j$) factors through the cause table. Second, τ^2 is estimable when the entry is written — it is the deployed model’s systematic-error scale, the

object a calibration certificate would certify — so everything below is computable prospectively, not only forensically. Point-valued practice stores W_i alone; the coupling, its cause, and τ^2 are discarded at the moment of writing.

3 The aggregation floor

Proposition 1 (Aggregation floor). *Under (1), with $\bar{W}$ and $\bar{X}$ the means of the W_i and X_i ,*

$$\text{Var}(\bar{W} - \bar{X}) = \frac{\sigma^2}{n} + \tau^2 \xrightarrow{n \rightarrow \infty} \tau^2 > 0.$$

Moreover, the floor is determined by the error coupling and not by the marginals: the marginal laws of the W_i are unchanged by replacing ξ with i.i.d. copies ξ_i , under which the floor is 0. Within (1) the floor is the error-borne covariance $\tau^2 = \text{Cov}(W_i, W_j) - \text{Cov}(X_i, X_j)$, $i \neq j$: it is carried by the error coupling keyed to its cause, not by the marginals, and not by the joint law of the written entries alone, which fixes $\text{Cov}(W_i, W_j)$ but not its split into $\text{Cov}(X_i, X_j)$ and τ^2 .

Proof. $\bar{W} - \bar{X} = \bar{\eta} + \xi$, with the two terms independent; the variance is $\sigma^2/n + \tau^2$. For the second claim, both constructions give each W_i the law of X_i plus an independent error of variance $\sigma^2 + \tau^2$; they differ only in the joint law, and the same computation with ξ_i i.i.d. gives $(\sigma^2 + \tau^2)/n \rightarrow 0$. For the last clause take, first, $\Sigma_X = 0$ with errors of covariance $\sigma^2 I_n + \tau^2 J_n$ and, second, $X_i = c$ for a shared c of variance $\tau^2/2$ and mean equal to the constants of the first, with errors of covariance $\sigma^2 I_n + (\tau^2/2)J_n$: both are within (1), the entries have the same joint covariance (in the Gaussian case the same joint law), and the floors are τ^2 and $\sigma^2/n + \tau^2/2$. $\square$

Averaging n entries produced through one instrument does not converge on the truth; it converges on the instrument. The floor does not need the η_i independent of one another: for any idiosyncratic errors independent of ξ , $\text{Var}(\bar{W} - \bar{X}) = \text{Var}(\bar{\eta}) + \tau^2 \geq \tau^2$. Instances that cross-check or average one another's outputs keep this hypothesis, since differences and averages of outputs carry ξ only through the term the identity already counts; an instance that shrinks its output toward the others' level correlates its error with ξ and is outside the model, having traded the floor for a bias. The covariance algebra is elementary and old; the floor statement has its proper home in the forecast- and expert-combination literature, where Clemen and Winkler [6] showed that positive dependence among sources sharply limits achievable precision, with the limiting value of additional sources far below the independent case (also [7, 8, 9]; Broomell and Budescu [10] decompose *why* judges correlate — shared cues — an analysis in the same spirit as the cause-keying above). In the voting register the same structure is regime-dependent: the Condorcet jury theorem survives correlation under explicit bounds and can fail beyond them [11, 12, 13, 14], and Pivato's large-society condition — average covariance vanishing asymptotically [15] — is exactly what (1) violates: τ^2 does not vanish.

What the deployment setting adds to this older material is the record-layer reading. The floor is not an analysis pitfall awaiting a careful analyst; it is a *storage defect*. Its determining object (the coupling), its key (the cause entity), and its magnitude (τ^2) are all present at write time and absent from the entry one instruction later. The entries agree with one another because they share one draw of ξ , not because they share the truth; a million of them carry one error, copied a million times.

Proposition 2 (One independent judge). *Under (1), let V be a further estimate of $\bar{X}$ with centered error independent of the η_i and of ξ , of variance ω^2 . The minimum-variance unbiased linear combination of $\bar{W}$ and V has error variance*

$$\left(\frac{1}{\sigma^2/n + \tau^2} + \frac{1}{\omega^2}\right)^{-1} \xrightarrow{n \rightarrow \infty} \left(\frac{1}{\tau^2} + \frac{1}{\omega^2}\right)^{-1} \leq \min(\tau^2, \omega^2).$$

In particular, whenever $\omega^2 \leq \tau^2$ the independent judge contributes at least as much precision as an unbounded number of instances of the instrument, and the combination beats the floor by at least a factor of two.

Proof. Inverse-variance weighting of two independent unbiased estimators; the limit is Proposition 1. $\square$

In the record register the proposition says two things. The combination weight is a function of τ^2 and ω^2 , so it is computable from a record that carries the coupling and its cause key and not from one that carries $W_1, \dots, W_n$ and V as $n+1$ points; and the hypothesis on V — error independent of ξ — is a statement about cause keys: V counts as a further estimate only if its cause entity is not the deployed model's. An estimate produced downstream of the model's own outputs shares the key and is, to that extent, another W_i . A point-valued record cannot tell the two apart. A cause-keyed one can identify the first case — same key, same draw, not a further estimate — and can only flag the second: a distinct key is necessary for independence and not sufficient, since distinct models trained on overlapping records share error by a mechanism the key does not see (§8; [3, Pred. 10]). What the key buys is the identification; what it does not buy is a certificate, which is Q2 of §9.

4 The floor propagates, and marginals do not carry it

The core's second-order correction [1, Rem. 3.5] quantifies the collapse discrepancy at a curved consumer: $\frac{1}{2} \text{tr}(\nabla^2 f(m) \Sigma) + o(\varepsilon^2)$ in the concentration family, with Σ the covariance the collapse discarded. The following shows that the shared cause requires no modification of that machinery — it enters through Σ alone — and that its contribution survives every $n \rightarrow \infty$ limit that kills idiosyncratic error. We use the trace term in the core's sense: it is the leading term of the correction in the concentration family, exact for quadratic g and an estimate otherwise.

Proposition 3 (Floor propagation). *Under (1), the joint covariance of $(W_1, \dots, W_n)$ is*

$$\Sigma_n = \Sigma_X + \sigma^2 I_n + \tau^2 J_n, \quad J_n = \mathbf{1}\mathbf{1}^\top,$$

with the τ^2 off-diagonal wholly error-borne. Let $\sup_i \text{Var}(X_i) < \infty$, let $\mathbb{E}W_i = m_0$ for all i (so $\bar{m} = m_0$), and let the consumer be $f_n(w) = g(\bar{w})$ with g twice continuously differentiable and g'' bounded on $\mathbb{R}$. Then the leading term of the second-order correction of [1, Rem. 3.5] at f_n is

$$\frac{1}{2} \text{tr}(\nabla^2 f_n \Sigma_n) = \frac{1}{2} g''(m_0) \left[\frac{\mathbf{1}^\top \Sigma_X \mathbf{1}}{n^2} + \frac{\sigma^2}{n} + \tau^2 \right],$$

whose error-borne part $\frac{1}{2} g''(m_0) \tau^2$ is independent of n and, when $g''(m_0) \neq 0$, nonzero. Moreover, decompose the τ^2 term by matrix position: the diagonal contributes $n\tau^2/n^2 \rightarrow 0$; the

off-diagonal contributes $n(n-1)\tau^2/n^2 \rightarrow \tau^2$. The persistent bias is carried asymptotically entirely by the coupling. In particular, when $g''(m_0) \neq 0$, a record layer that stores exact per-cell uncertainties but no dependence — Σ_n 's diagonal without its off-diagonal — predicts a leading-term correction of order $1/n$ where the leading-term correction is $\Theta(1)$. If $n^{-2}\mathbf{1}^\top \Sigma_X \mathbf{1}$ converges to $\bar{v}_X$ (zero for independent X_i), the whole bracket converges to $\bar{v}_X + \tau^2$.

Proof. $\nabla^2 f_n = n^{-2}g''(m_0)J_n$, so $\text{tr}(\nabla^2 f_n \Sigma_n) = n^{-2}g''(m_0)\mathbf{1}^\top \Sigma_n \mathbf{1}$, and $\mathbf{1}^\top \Sigma_n \mathbf{1} = \mathbf{1}^\top \Sigma_X \mathbf{1} + n\sigma^2 + n^2\tau^2$. The decomposition splits $n^2\tau^2$ as $n\tau^2$ (diagonal) plus $n(n-1)\tau^2$ (off-diagonal). The diagonal-only record replaces Σ_n by its diagonal, whose $\mathbf{1}$ -quadratic form is $\text{tr} \Sigma_X + n\sigma^2 + n\tau^2 = O(n)$ under $\sup_i \text{Var}(X_i) < \infty$, giving a correction of order $1/n$. $\square$

Proposition 3 upgrades Proposition 1 from a statement about variance to a statement about *bias at every consumer curved at the mean*, and it sharpens the storage claim past the obvious repair. The obvious repair — annotate each cell with an uncertainty — is the marginal record, and the marginal record fails here not approximately but categorically: as n grows it certifies a vanishing correction while the leading-term correction is fixed at $\frac{1}{2}g''(m_0)\tau^2$. The object that must be carried is the coupling, keyed by its cause. The audited-collapse program of [1, Problem 6] applies as written, with Σ_n in the place of the discarded covariance. The conformance checks of [2, §5] are stated per pipeline; within one database state they extend to a coupling that runs across records by substitution of the state's joint covariance, since affine transparency transports covariances and curvature accounting consumes them, and across states they are undefined for the reason Remark 4 gives.

5 Independence across institutions

Remark 4 (The product default at deployment scale). Let institutions A and B maintain disjoint records, exchanging no data, both consuming outputs of a deployed model M as in (1). Any downstream computation that aggregates across A and B under an independence assumption — risk pooling, meta-analysis, ensemble voting, redundancy engineering — computes the pushforward of $\mu_A \otimes \mu_B$ where the true joint is a coupling with error-borne covariance τ^2 . By the marginal non-identifiability of [1, §2.2], the resulting error is not determined by the institutions' marginals; a bound on it derived from the marginals alone is a bound over the Fréchet class of couplings with those marginals [22], the product coupling included, and so cannot distinguish the shared draw ξ from independent draws of the same variance: it is a statement about what the marginals permit, not about what the deployment did.

This is the product default of [1, Rem. 2.5] — the unexamined choice the measure-relational model exists to make explicit — with the two marginals in two institutions that do not know they share an instrument; it is the core's remark at institutional scope, not a new result. No institution's own record contains the evidence of the coupling: cause-keyed dependence across record boundaries is representable only in a layer that carries causes, which is [1, Problem 2]'s factorization with the model as cause entity. One caveat is owed to the reader. The core's Definition 2.9 fixes one joint law per database state; a coupling across two institutions' states is an object the core does not yet define, and this remark uses it informally, with the cause entity of Problem 2 as its evident key. One historical instance is standardly read in this shape and should be cited only as shape: in the correlated-defaults failure of 2008, a shared pricing

model occupied the role of ξ across institutions whose risk computations did not carry its error as shared [26]. The claim is the shape-analogy, not that any particular formalism would have prevented the event.

Observation 5 (Coupling without a shared datum). The coupling of Remark 4 requires no shared data whatsoever. Take $\text{Cov}(X_A, X_B) = 0$ and disjoint inputs; the entries W_A, W_B remain coupled through ξ . Data-lineage provenance [23, 24] follows the data and finds none; workflow provenance [25] may record M as a processing step but represents no error coupling induced by it. The dependence is invisible to every lineage-shaped audit and representable only cause-keyed.

6 Instances

Four cases, no new proofs; the value is the address.

Many instances of one model. The case the press calls an agentic swarm [5]: n instances, one model, one version, one prompt regime. Each instance’s output is an entry W_i with its own idiosyncratic error and the model’s shared one; instances voting, cross-checking, or converging on a consensus compute $\bar{W}$, and Proposition 1 is their ceiling. Proposition 2 prices the alternative: one estimate whose error is independent of the model’s — a second model trained on a disjoint corpus, an instrument in the world, a reader who has not seen the instances’ outputs — and says when that estimate outweighs all n . The entries the instances write carry neither τ^2 nor the cause key, so a consumer reading them cannot compute how many independent judgments they hold; they hold one.

Model version as the cause entity. The cause table’s rows are (model, version, fine-tune, prompt regime, decoding parameters). A model update is a recalibration event; point-valued practice embeds the old calibration in every derived cell with no update path — the update anomaly of [1, Problem 2], verbatim, at the largest scale it will ever have. The granularity question — how fine the key must be before the factorization loses its economy — is Q1 of §9.

The embedding store. A retrieval layer that encodes text to vectors and re-enters the nearest neighbours into a prompt is the layer [1, Problem 8] names and [2, §6] places, and its correspondence to (1) is by shape: the encoder version is the cause entity, and a re-encode of the store is a recalibration event that point-valued practice performs by overwriting. Whether a scalar τ^2 is the right object for a store whose error is a function of the input, and what the store must carry to represent a cause-keyed coupling among its vectors, is [1, Problem 2]’s question asked of [1, Problem 8]’s relation.

The learner under shared-cause records. What a learner converges to when its training records are model-derived is established in [4]: with labels held, the one-generation artifact is a fixed point ([4, Thm. 4]); with labels re-derived from the previous generation, the learned response contracts toward a constant at a rate set by the instrument ([4, Thm. 5]); with a fixed fraction of retained truth, the mixture has its own fixed point (Loop C, [4, Prop. 7]). Those results are the time axis (k generations) and assume conditionally independent readings. This note is the population axis (n instances) and its model violates that assumption: for records (W_i, Y_i) with (X_i, Y_i) i.i.d. and Y the outcome as in [2, §2], conditionally on ξ the records are i.i.d., so the one-generation target of [2, Prop. 1] is reached, but as $\mathbb{E}[Y \mid W = w, \xi] = h_A(w - \xi)$, with $h_A(w) = \mathbb{E}[Y \mid X + \eta = w]$ that target absent the shared draw and ξ fixed at its realized

draw: the artifact translated by a quantity no record identifies, since every record shares it. ([2, Rem. 3] does not apply: a record process sharing one draw is exchangeable and not ergodic, so its marginal target is not the learner’s.) This is [4, §8]’s observation that a shared offset changes the limit and not only the rate, seen from the point-valued side. The composition of the two axes is open (§9). The learning-side taxonomy of which resulting errors are artifacts is the companion [3].

7 Two domains

Of the many fields the mechanism touches, two are treated here — one because the stakes are highest, one because the mathematics is cleanest — rather than six thinly.

Healthcare. A clinical model imputing fields — risk scores, image reads, structured extractions from notes — deployed across hospital systems (for the categorical outputs among these the shared component is not additive and (1) applies by shape) writes one ξ into records that meta-analyses, registry studies, and pooled trials will later treat as independent readings. Proposition 1 says the pooled estimate converges to the imputation model; Remark 4 says no participating institution’s records contain the evidence; Proposition 3 says the resulting bias at any endpoint curved at the mean — risk models are curved — survives arbitrary pooling and is invisible to per-entry uncertainty annotations. The repair is not less imputation; it is a record that carries the imputation’s cause key and error coupling, so the pooled analysis can be computed against the true joint — or at minimum can price what was assumed.

Redundancy engineering. Redundancy is a purchase of independence: k sensors, k engines, k perception stacks, engineered so that failure probabilities multiply. A shared model across the k channels silently refunds the purchase. Model (1) is additive on values and does not apply to failure indicators as written; the analogue is one line. If a shared cause fails every channel with probability $q > 0$, the joint failure probability is at least q for every k , where the engineering case for k was p^k : a floor no increase in k touches, for the same reason Proposition 1’s floor is n -independent. The contrast with healthcare is instructive: there the coupling corrupts an *inference*; here it voids a *design guarantee*, and the certificate that would price the loss is the object Q2 asks for, carried in the record.

8 Related work

Algorithmic monoculture. The deployment phenomenon has an established literature. Kleinberg and Raghavan [17] formalized monoculture and showed common algorithms across decision-makers can lower welfare even when individually more accurate; Bommasani et al. [18] broadened the frame from same-system to *shared components* — training data, foundation models — and introduced outcome homogenization; Creel and Hellman [19] supplied the normative register. The phenomenon is now empirical: Kim, Garg, Peng, and Garg [20] evaluate over 350 language models and find substantial error correlation — on one leaderboard dataset, about 60% agreement on the same wrong answer when two models both err — driven partly by shared providers and architectures, with more accurate models *more* correlated even across providers. The cause-keyed coupling of §2 is anticipated conceptually by [17, 18], and its mechanism is present in the component-sharing frame of [18]; the learning-side reading of the

same finding — a shared record and a shared target — is the companion’s ([3, Pred. 10–11]), which also names the competing item-difficulty account. The aggregate finding is consistent with both readings and is evidence for neither over the other. What none of these works occupy is the storage claim: the coupling as a first-class object of the record layer, keyed by cause, with the floor (Proposition 1) computable at write time and the propagated bias (Proposition 3) computable at every consumer. This is the same engagement shape as the core’s with MCDB: same neighborhood, different object — welfare analysis of decisions there, the coupling as a stored object here.

Model collapse. The literature on training with synthetic data — the phenomenon, the accumulate regime, and its strong form — is engaged and separated from the family’s results in [4, §6]; this note adds nothing to collapse dynamics and takes that separation as given.

Aggregation under dependence. Engaged in §3; the trail runs [7, 8, 9, 6, 10] for combination and [11, 12, 13, 14, 15, 16] for voting.

9 Open problems

Three problems. The first two are open; the third is closed in the case of conditionally independent readings and open in this note’s own.

Q1 (minimal cause schema). Version and weights are too coarse a key if decoding temperature and prompt regime move the error distribution; too fine a key destroys the factorization’s economy. This is the lossless-vs-no-finite-factorization dichotomy of [1, Problem 2], asked concretely of deployed models.

Q2 (calibration certificates for models). The conditional cell for a model-derived value requires an error model *for the model* — [1, Problem 4]’s elicitation gap, one level up. What is the analogue of an instrument’s calibration certificate for a deployed model? A held-out evaluation suite graded right/wrong is not one: [3, §3] shows that binary grading discards the model’s law before scoring, so such a suite records no dispersion and cannot certify τ^2 in the instrument’s sense. What remains is the constructive half: what a suite would have to record, and at what granularity of cause key, to qualify. The redundancy domain (§7) shows the certificate is not academic: it is the object that would make a voided design guarantee computable, which is what a design review can act on.

Q3 (iteration across generations). For conditionally independent readings the question is answered in [4]: Loop A ([4, Thm. 4]), Loop B ([4, Thm. 5]), and the conforming memory ([4, Thm. 8]), with the separation from the collapse literature made there. This note’s model is the case those results exclude, as [4, §8] states from the conforming side: under a shared draw the readings are not conditionally independent and the coupling changes the limit, not only the rate. What remains open is the composition of the two axes: n instances of generation k writing the entries of generation $k+1$, where the shared component ξ is itself re-learned. The composition, when written, belongs in this paper or below it.

Acknowledgments

Drafted with assistance from Anthropic Claude Fable 5.1. All claims, errors, and rulings are the author’s.

References

- [1] M. P. Laing. Measure-relational algebra: the core. <https://doi.org/10.5281/zenodo.22999911> (all versions); realql.org, 2026.
- [2] M. P. Laing. What a memory must preserve: machine reasoning over uncertain observations. realql.org, September 2026.
- [3] M. P. Laing. Which hallucinations are artifacts? Faithful reporting of an unfaithful record. realql.org, September 2026.
- [4] M. P. Laing. The fixed point of self-improvement: three limits of a learner that trains on its own memory. realql.org, September 2026.
- [5] W. Knight. Why so many AI researchers think the machines could kill everyone. *WIRED*, September 11, 2026.
- [6] R. T. Clemen and R. L. Winkler. Limits for the precision and value of information from dependent sources. *Operations Research* 33, 427–442 (1985).
- [7] J. M. Bates and C. W. J. Granger. The combination of forecasts. *Operational Research Quarterly* 20, 451–468 (1969).
- [8] R. M. Hogarth. A note on aggregating opinions. *Organizational Behavior and Human Performance* 21, 40–46 (1978).
- [9] R. L. Winkler. Combining probability distributions from dependent information sources. *Management Science* 27, 479–488 (1981).
- [10] S. B. Broomell and D. V. Budescu. Why are experts correlated? Decomposing correlations between judges. *Psychometrika* 74, 531–553 (2009).
- [11] K. K. Ladha. The Condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science* 36, 617–634 (1992).
- [12] S. Berg. Condorcet’s jury theorem, dependency among jurors. *Social Choice and Welfare* 10, 87–95 (1993).
- [13] K. K. Ladha. Information pooling through majority-rule voting: Condorcet’s jury theorem with correlated votes. *Journal of Economic Behavior and Organization* 26, 353–372 (1995).
- [14] S. Kaniovski. Aggregation of correlated votes and Condorcet’s jury theorem. *Theory and Decision* 69, 453–468 (2010).
- [15] M. Pivato. Epistemic democracy with correlated voters. *Journal of Mathematical Economics* 72, 51–69 (2017).
- [16] F. Dietrich and K. Spiekermann. Jury theorems. *Stanford Encyclopedia of Philosophy* (2021).
- [17] J. Kleinberg and M. Raghavan. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences* 118(22), e2018340118 (2021).
- [18] R. Bommasani, K. A. Creel, A. Kumar, D. Jurafsky, and P. Liang. Picking on the same person: does algorithmic monoculture lead to outcome homogenization? *NeurIPS* 35, 3663–3678 (2022).
- [19] K. Creel and D. Hellman. The algorithmic leviathan: arbitrariness, fairness, and opportunity in algorithmic decision-making systems. *Canadian Journal of Philosophy* 52, 26–43 (2022).

- [20] E. Kim, A. Garg, K. Peng, and N. Garg. Correlated errors in large language models. In *Proceedings of the 42nd International Conference on Machine Learning*, PMLR 267:30038–30066, 2025. arXiv:2506.07962.
- [21] JCGM 100:2008. *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (GUM).
- [22] R. B. Nelsen. *An Introduction to Copulas*, 2nd ed. Springer (2006).
- [23] P. Buneman, S. Khanna, and W.-C. Tan. Why and where: a characterization of data provenance. In *ICDT 2001*, LNCS 1973, 316–330 (2001).
- [24] J. Cheney, L. Chiticariu, and W.-C. Tan. Provenance in databases: why, how, and where. *Foundations and Trends in Databases* 1(4), 379–474 (2009).
- [25] L. Moreau and P. Missier (eds.). PROV-DM: the PROV data model. W3C Recommendation, April 30, 2013.
- [26] D. MacKenzie and T. Spears. ‘The formula that killed Wall Street’: the Gaussian copula and modelling practices in investment banking. *Social Studies of Science* 44(3), 393–417 (2014).
- [27] M. Favaro and J. Clark. When AI builds itself. Anthropic Institute, June 2026. <https://www.anthropic.com/institute/recursive-self-improvement>

Draft 6, RealQL30, September 21, 2026. Companion essay: “The Leaderboard as Galton’s Fair” (RealQL.org). Nothing in this paper publishes ahead of the core.