Auditing Independence in Multi-Channel Measurement

An Agreement-Curve Diagnostic for Genuine versus Correlated Redundancy

Jeremy C. Jones

HoldingLight LLC — ORCID 0009-0007-2515-3774 — universalcollapse.com

Series: Universal Collapse Theory — Methods Paper • Version v1.0 • 2026-05 • CC BY 4.0

Abstract

Principle. Multi-channel measurement is ubiquitous: multi-detector physics readouts, component-separated reconstructions, ensemble sensors, multi-rater annotation, and replicate measurement all produce families of putatively independent observations of a shared latent quantity. Pairwise agreement statistics (Pearson correlation, intraclass correlation, Cohen's kappa, RMS deviation, Bland–Altman) report whether channels agree on average, but they cannot tell whether the rate at which agreement improves with channel count k matches what would be expected if the channels were genuinely conditionally independent.

Result. We propose an agreement-curve diagnostic in two regimes. In the classification regime, conditional independence with positive Chernoff information implies that pairwise observer disagreement decays exponentially in k at rate C, where C is the per-channel Chernoff information of an independently specified channel model. In the continuous regime, conditional independence implies that the variance of the k-channel mean decays as σ^2/k_eff, while modeled shared structure appears as a floor or covariance term; deviations from the modeled curve (a floor above what is predicted, slower decay, or a different scaling) diagnose shared structure beyond what the channel model accounts for. The protocol's sharp model-conditional falsifier is in both cases the comparison of an observed agreement curve against the scaling law predicted by an independently specified channel model — sharp conditional on that model's adequacy. The classification version is demonstrated on a synthetic binary-symmetric-channel case; the continuous version on a synthetic Gaussian-channel case; the application to Planck PR3 component-separated CMB maps is specified for execution against existing analysis infrastructure. The result is operational, not philosophical: it characterizes when pairwise agreement is evidence of independent redundancy, not when independence is metaphysically true.

Keywords: redundancy; multi-channel measurement; Chernoff information; variance reduction; effective sample size; sensor fusion; quantum Darwinism; reproducibility.

Review target

The review target is split into two acceptance levels, paralleling the structure of the S_2 and S_3 Methods Papers in this series.

Specification acceptance

Acceptance of v1.0 of this Methods Paper means the reviewer agrees that the proposed audit structure is well-defined: independently specify a channel model — Chernoff information c_0 in the classification regime, residual-covariance prediction $V^*(k) = w^T\Sigma w + F$ in the continuous regime — from channels not using the agreement data being tested; construct the agreement curve under the appropriate regime; report the effective-sample-size audit k_eff with explicit accounting for known shared upstream structure; compare the observed agreement-curve scaling (fitted exponent in the classification regime, fitted residual-variance curve in the continuous regime) against the model-predicted scaling; test named confound alternatives including correlated pseudo-redundancy, shared additive bias, and block-correlated structure. The structural specification can be accepted independently of whether the diagnostic outperforms standard pairwise agreement statistics on any specific dataset.

Performance acceptance

Acceptance of the protocol's performance — a stronger claim than acceptance of the specification — requires the frozen synthetic demonstrations of §4 and a downstream real-data application that satisfies the dataset prerequisites of §5. v1.0 presents the synthetic demonstrations in both the classification and continuous regimes; real-data execution against the Planck PR3 component-separated CMB maps is the work of a companion empirical paper (Jones 2026d), with the further illustrative role that high cross-product agreement under shared upstream inputs is not independent redundancy unless the channel model and k_eff audit justify that interpretation.

Rejection criterion

The protocol should be revised or rejected if the channel model, k_eff audit, or predicted scaling cannot be specified without circular use of the same agreement data; if controlled synthetic cases fail to distinguish genuine independent redundancy from the named confounds (correlated pseudo-redundancy, shared additive bias, block-correlated structure); if the real-data protocol lacks an independent channel model; or if the continuous-regime formulation cannot detect shared-bias confounds when an independent reference, simulation truth, or independently specified noise model is available. Acceptance does not require that all confounds be disambiguated from pairwise statistics alone; some require external channel-model input or k_eff audit, and the protocol's sharpness is conditional on those external channels being available.

Stack placement

This paper is a Universal Collapse Theory Methods Paper for S_1 applications. Records Across Nature, Life, and Mind (Jones 2026a) defines records as the persistence layer and states the redundancy-implies-consensus signature S_1 as a portable empirical claim. The Structuralization of Empiricism (Jones 2026b) treats S_1 as a stabilization test for empirical convergence under shared records. The Update Integrity Standard (Jones 2026c) requires an independence audit and k_eff reporting before S_1 convergence can support a Level 3 claim. Objectivity from Records (Jones 2026e) supplies the formal binary Chernoff bound. The present paper translates those commitments into a practical multi-channel diagnostic — a bridge between the formal lemma and downstream T16 empirical demonstrations. Formal pair: TN-S_1 (Objectivity from Records). Together they constitute the S_1 formal–protocol pair — the paired Technical Note proves the formal bound; this Methods Paper translates it into a deployable audit protocol. It is a Methods Paper, not a standard.

1. Introduction

Independent measurements of a shared latent quantity are usually expected to agree more closely as more measurements are aggregated. This expectation underlies sensor fusion in engineering (Chair & Varshney 1986), multi-detector analyses in physics, ensemble methods in machine learning, multi-rater annotation in psychology and the social sciences (Cohen 1960; Dawid & Skene 1979), and replicate measurement in metrology. When multiple channels report converging values, the convergence is often cited as evidence that the measurement is robust: shared structure dominates over channel-specific noise.

The standard practice for reporting such agreement is well established: Pearson correlation between channel outputs, intraclass correlation (Shrout & Fleiss 1979; McGraw & Wong 1996), Cohen's kappa for categorical agreement (Cohen 1960), root-mean-square deviation between channel pairs, or Bland–Altman analysis (Bland & Altman 1986). These statistics ask whether the channels agree at the level of marginal pairs. Bland and Altman noted explicitly that high correlation between methods does not imply agreement between methods, and that pairwise summaries can mislead about systematic differences. What none of these statistics directly tests is whether the rate at which agreement improves as more channels are aggregated matches what would be expected if the channels were genuinely conditionally independent.

This is more than a technicality. Two distinct generating processes can produce nearly identical pairwise agreement statistics:

> (i) Genuine conditionally independent fragments. Each channel produces a noisy readout of the shared latent, with channel-specific noise that is independent across channels conditional on the latent. As the number of channels k grows, the variance of the aggregate decays at the rate predicted by a noise model — exponentially in k for well-posed classification tasks, as σ^2/k for continuous averaging.
>
> (ii) Correlated pseudo-redundancy. Each channel produces a readout with substantial shared systematic structure — a common processing pipeline, an upstream source of correlated noise, an unmodeled confound — in addition to channel-specific variation. Pairwise agreement statistics may be high, but adding more channels does not buy the predicted reduction: the curve of disagreement-versus-k saturates at a floor set by the residual correlated component.

Pairwise summaries can be matched across these regimes under plausible calibration choices (Bland & Altman 1986 emphasizes this concern for the special case of correlation versus agreement in method comparison). They differ in how disagreement scales as channels are aggregated. This difference is detectable, but it is not detected by the standard pairwise statistics.

The diagnostic proposed here is a comparison of the observed agreement curve against the scaling law predicted by an independently specified channel model. The form of this prediction depends on the outcome type. In classification settings — finite decision outcomes — the prediction is exponential decay of pairwise observer disagreement at the per-channel Chernoff information rate (Chernoff 1952; Cover & Thomas 2006, Ch. 11). In continuous settings — measurements aggregated by averaging — the prediction is variance reduction toward zero at rate σ^2/k_eff under independent noise. Both forms are agreement-curve diagnostics; they differ in what mathematical object the model predicts.

The contribution of this paper is methodological:

(1) An agreement-curve protocol applicable in both classification and continuous regimes, with the same four-step structure in each.

(2) An effective-sample-size audit (in the tradition of Kish 1965 design-effect analysis) that adjusts k for measurable inter-channel correlation, producing k_eff.

(3) A sharp model-conditional falsifier: the comparison of fitted agreement-curve scaling against the model-predicted scaling. When the model is well-specified and the channels are conditionally independent, observed scaling tracks predicted scaling within confidence. When the channels share structure beyond what the model accounts for, the observed curve deviates in characteristic ways: a finite floor in the continuous case, an exponent below C in the classification case, or a saturation k_eff well below the audited level. Sharpness is conditional on the channel model; if the model is wrong, the falsifier is still informative but no longer sharp.

The sharpness comes from comparison against a prediction computed without using inter-channel agreement. This breaks the circularity that affects diagnostics calibrated against the same agreement they are supposed to validate.

The paper is organized as follows. Section 2 states the formal results — the classification (§2.1) and continuous (§2.2) regimes. Section 3 specifies the protocol. Section 4 demonstrates it on synthetic cases in both regimes. Section 5 specifies the application to Planck PR3 component-separated CMB maps. Sections 6–7 discuss applicable domains and limitations. Section 8 provides a reporting template. Section 9 places the protocol in the broader UCT context. Appendix A provides reproduction code.

2. Theory: agreement curves in two regimes

2.1 Classification regime: redundancy implies exponential consensus

Setup. Let $X \in \{0, 1\}$ be a binary latent variable with equal priors. Each channel i produces a fragment Y_i taking values in a measurable space. Conditional on $X = x$, each Y_i has density p_x with respect to a common dominating measure μ.

Assumption A (conditional independence). $Y_1, \ldots, Y_k$ are conditionally independent given X.

Assumption B (discriminability). The pair $\{p_0, p_1\}$ has Chernoff information at least $c_0 > 0$:

$$C(p_0, p_1) := \sup_{0 \le s \le 1} \left[-\log \int p_0(y)^s p_1(y)^{1-s} \, d\mu(y) \right] \ge c_0 > 0.$$

Theorem 1 (Jones 2026e). Under Assumptions A and B with iid fragments, the Bayes-optimal (MAP) error satisfies $P_e(k) \le \tfrac{1}{2} \exp(-kC)$, and pairwise observer disagreement under disjoint k-fragment samples satisfies

$$\Pr[\hat{X}^{(1)}_k \ne \hat{X}^{(2)}_k] \le 2 P_e(k) \le \exp(-kC).$$

Sample complexity: $k \ge (1/c_0) \log(1/\delta)$ suffices to drive pairwise disagreement below $\delta \in (0, 1)$.

Proof. The log-likelihood ratio is a sum of k iid random variables; Chernoff's bound gives the MAP error rate (Chernoff 1952; Cover & Thomas 2006, Ch. 11). The pairwise disagreement bound follows from a union bound over the two independent MAP decisions. Full proof is in Jones 2026e.

Asymptotic tightness. The Chernoff exponent C is the best achievable error exponent for binary hypothesis testing with iid samples (Cover & Thomas 2006, Theorem 11.9.1); it is therefore the rate, not merely a lower bound. Empirical exponents systematically below C indicate violation of one of the assumptions.

Heterogeneous and quantum extensions. Heterogeneous fragments admit the additive Bhattacharyya bound; quantum readouts admit the quantum Chernoff bound (Audenaert et al. 2007; Nussbaum & Szkoła 2009). See Jones 2026e for details.

2.2 Continuous regime: variance-reduction agreement curves

Setup. Let X be a continuous latent quantity (or vector field) and $Y_i = X + \varepsilon_i$ where ε_i is a channel-specific perturbation with mean zero and covariance Σ_i. Define the k-channel mean $\bar{Y}_k = (1/k)\,\Sigma_{i=1}^{k} Y_i$ and the residual noise mean $\bar{\varepsilon}_k = \bar{Y}_k - X$.

Independent-channel prediction. If the ε_i are mutually independent with common variance σ^2, then $\mathrm{Var}(\bar{\varepsilon}_k) = \sigma^2/k$. Aggregation drives the variance of the residual to zero at rate 1/k. With $k_{eff} < k$ under correlation, the rate degrades to σ^2/k_{eff}.

Shared-bias prediction. If $\varepsilon_i = \delta + \eta_i$ with δ a per-trial shared component of variance v_δ and η_i independent of variance $\sigma^2 - v_\delta$, then $\mathrm{Var}(\bar{\varepsilon}_k) = (\sigma^2 - v_\delta)/k + v_\delta$. The residual variance saturates at v_δ as $k \to \infty$: this is the operational signature of shared additive bias.

Block-correlated prediction. If channels are partitioned into blocks of size B with identical noise within blocks and independent across blocks, $\mathrm{Var}(\bar{\varepsilon}_k) = \sigma^2 B/k = \sigma^2/k_{eff}$ with $k_{eff} = k/B$. The decay shape is the same $1/k_{eff}$ form; only the rate is degraded.

Heterogeneous channels. The equal-variance shorthand $\sigma^2/k_{eff} + F$ is convenient for explanation but specializes a more general object. For heterogeneous channels with covariance matrix Σ and aggregation weights w (e.g., $w_i = 1/k$ for the unweighted mean), the predicted residual variance is $V^*(k) = w^T \Sigma\, w + F$, where F captures any modeled shared-systematics floor. k_{eff} is then a summary scalar derived from this prediction (e.g., $\sigma^2/V^*(k)$ under the equal-variance reduction) rather than the primary object. Real applications should report the covariance-based prediction directly when channels are heterogeneous.

2.3 The disjoint-subset trap

A natural attempt to mirror the classification protocol uses a disjoint-subset disagreement statistic $D^2(k) = E[(\bar{Y}_A - \bar{Y}_B)^2]$ where A and B are disjoint k-channel subsets. In the classification regime this is sharp. In the continuous regime under additive shared bias, however, the bias cancels exactly:

$$\bar{Y}_A - \bar{Y}_B = (1/k)\Big(\Sigma_{i\in A}\,\eta_i\Big) - (1/k)\Big(\Sigma_{i\in B}\,\eta_i\Big), \qquad E[(\bar{Y}_A - \bar{Y}_B)^2] = 2(\sigma^2 - v_\delta)/k.$$

Disjoint-subset disagreement does not detect additive shared bias in the continuous regime. The continuous diagnostic must therefore use the residual variance $\mathrm{Var}(\bar{Y}_k - X)$ — which requires either a known reference X (e.g., simulation truth) or an independent estimate of the noise level (e.g., from a published noise model). This is one place where the independence audit's external channel-model input is load-bearing: without it, additive shared bias is invisible to within-data agreement statistics.

2.4 Diagnostic comparison

Table 1 summarizes the parallel structure across regimes. The choice of regime is determined by outcome type, not by framework or preference; classification statistics do not generalize directly to continuous data, and continuous statistics do not generalize directly to finite decisions.

Regime	Outcome type	Statistic	Expected independent scaling	Failure signature
Classification	binary / finite decision	$D(k) = \Pr[\text{disagree}]$	$a \exp(-bk)$, $b \approx C^*$	$b < C^*$; saturation below k_{eff}; early floor
Continuous (variance)	scalar / vector / map	$V(k) = \mathrm{Var}(\bar{Y}_k - \hat{X}_{ref})$	$A/k_{eff} + F$	floor above modeled F; slower decay; wrong covariance shape
Continuous (RMS)	scalar / vector / map	$\sqrt{V(k)}$	$\sqrt{(A/k_{eff} + F)}$	same as above on RMS scale

Table 1. Classification, continuous-variance, and continuous-RMS agreement-curve diagnostics. Variance and RMS forms are equivalent up to a square root; reporting one or the other is a presentation choice, not a different diagnostic. Classification and continuous regimes are separate diagnostics with different scaling laws and different failure modes.

Therefore the continuous protocol is a covariance-model audit, not a Chernoff-exponent audit. The two regimes share the principle (compare observed agreement scaling to a model-predicted scaling computed without using inter-channel agreement) but instantiate it on different mathematical objects.

3. The protocol

The protocol consists of four steps, with regime-specific instantiations at Steps 1, 2, and 4. Each step is specified to be auditable independently of the inter-channel agreement statistics that the protocol tests.

3.1 Step 1: Independent channel characterization

Classification: construct a channel model that yields per-channel Chernoff information C without reference to inter-channel agreement. Acceptable inputs: published noise covariance models, validation against ground-truth data, within-channel residual analysis on disjoint subsets, or theoretical channel-capacity calculations. Output: predicted exponent C^*.

Continuous: construct a channel model that yields the per-channel noise variance σ^2 (or covariance, in vector cases) and any expected shared-component structure, again without using inter-channel agreement. Acceptable inputs: instrument noise simulations (e.g., FFP10-style end-to-end simulations for Planck), calibration data, propagation analysis, or validation against ground-truth test signals. Output: predicted variance-reduction curve $V^*(k) = \sigma^2/k_{eff}$ (plus any expected floor F under known shared structure).

If no independent channel model is available, the protocol can only test the form of the agreement curve (exponential or 1/k) but not its rate or its predicted floor. The full diagnostic requires the model output.

3.2 Step 2: Agreement-curve construction

Classification: compute D(k) = Pr[$\hat{X}^{(1)}\_k \neq \hat{X}^{(2)}\_k$] from disjoint-subset MAP decisions, varying k from 1 to the available channel count, with empirical confidence intervals.

Continuous: compute V(k) = Var($\bar{Y}\_k - \hat{X}\_ref$), where $\hat{X}\_ref$ is a reference signal estimate obtained independently of the data being tested — e.g., simulation truth, leave-one-pipeline-out prediction, or a published noise-model expectation. Vary k from 1 to the available channel count and report V(k) with confidence intervals across spatial / temporal / sample units.

Caution. Comparing the k-channel mean to the full-channel mean produces a curve that decreases by construction as k → all channels and should not be treated as a test of independence. The reference must be independent of the channels being aggregated.

Reference uncertainty. If $\hat{X}\_ref$ is noisy rather than simulation truth, its uncertainty contributes to V(k) and must be incorporated into the predicted V*(k) — otherwise residual variance from the reference itself may be incorrectly attributed to channel dependence. Leave-one-pipeline-out and other within-data references are particularly susceptible to this; their reference-noise contribution should be modeled or estimated independently.

Confidence intervals. Subset samples at different k values may overlap, so empirical confidence intervals on D(k) or V(k) cannot be treated as independent-binomial intervals. Report intervals using bootstrap over independent units, block bootstrap, simulation ensembles, or analytic covariance where available. This is especially important for small channel counts (e.g., the Planck full-mission k = 4 case) where most inferential power comes from spatial or temporal structure within each channel-comparison rather than from channel-count degrees of freedom.

3.3 Step 3: Effective-sample-size audit

For correlated channels, define and report k_eff. The simplest correction, valid for approximately uniform pairwise correlation $\bar{r}$ (a special case of Kish 1965 design-effect analysis), is

$$k\_eff = k / [1 + (k - 1)\,\bar{r}],$$

where $\bar{r}$ refers to residual inter-channel correlation after the shared latent signal and any modeled shared structure have been removed or conditioned on. Raw output correlation is not an independence audit: genuine channels measuring the same X should be correlated at the output level by construction, and a high raw $\bar{r}$ between Y_i and Y_j is consistent with both genuine independent fragments of X and correlated pseudo-redundancy. The audit must therefore work with residuals against an independently specified reference or noise model, not with output-level correlations.

More careful audits use variance-component decompositions (within-channel versus shared-bias variance), explicit covariance modeling, block-model assumptions, or summable-mixing bounds. The k_eff audit provides a second model-conditional falsifier: if the agreement curve saturates at a higher floor than the audit predicts, residual correlated structure beyond what the audit captures is present. This is the operational diagnostic for the audit missing a confound.

k_eff may be estimated from covariance structure, design effects, simulations, or independent residual models. If it is fit from the same agreement curve being tested, the result is exploratory and cannot serve as the independent falsifier — the audit becomes circular. Reports should specify the source of k_eff and label exploratory estimates as such.

3.4 Step 4: Falsifier comparison

Classification: fit $D(k) \approx a \exp(-bk)$ over the range where the form holds and compare b to the predicted Chernoff information C*. $b \approx C^*$ (within confidence): consistent with conditional independence at the predicted discriminability. $b \ll C^*$: independence violated, or model overstates discriminability. Saturation earlier than k_{eff} predicts: residual correlated structure beyond the audit.

Continuous: compare V(k) to the predicted variance-reduction curve $V^*(k) = \sigma^2/k_{eff} + F$, where F is any predicted shared-structure floor. $V(k) \approx V^*(k)$: consistent with the model. V(k) decays as σ^2/k_{eff} but with k_{eff} smaller than audited: additional unmodeled correlation. V(k) saturates at a higher floor than predicted: shared structure beyond what the model accounts for. V(k) decays faster than predicted: rare; suggests the model overstates noise.

Both the b-versus-C* comparison and the V(k)-versus-V*(k) comparison are forms of the same diagnostic principle: the observed agreement curve is compared to a scaling law predicted by an independently specified channel model. The independence audit, not the bound itself, is what does the work in any real domain. The bound or the noise model says what to expect under independence; the audit says how independent the channels actually are.

4. Synthetic demonstrations

4.1 Classification regime: binary symmetric channel with three correlation structures

Three regimes are simulated:

(A) iid binary symmetric channel (BSC). Each channel produces a noisy bit with crossover probability $p = 0.40$, conditionally independent across channels. Predicted Chernoff information: $C^* = -\log[2\sqrt{(p(1-p))}] \approx 0.0204$.

(B) Shared-bias channel. Each channel produces a noisy bit with a shared per-trial bias term δ perturbing the crossover probability of every channel in that trial. Marginal per-channel error rate matches regime (A); shared bias creates correlated structure that aggregation cannot remove.

(C) Block-correlated channel. Channels are grouped into blocks of size $B = 4$ with identical noise within blocks and independent across blocks. Effective channel count: $k_{eff} = k/B$.

Result. Regime (A): D(k) tracks predicted $\exp(-kC^*)$. $b \approx C^*$, $k_{eff} \approx k$. Status: confirmed. Regime (B): D(k) decays at rate $b < C^*$ and saturates at a finite floor; the b-versus-C* comparison detects the violation. Status: pseudo-redundancy via shared bias. Regime (C): D(k) decays at rate approximately C^*/B and saturates consistent with $k_{eff} = k/B$; both diagnostics flag it. Status: independence rejected with effective channel count k/B.

Regime	Pairwise summary	Curve fit	Diagnosis
(A) iid BSC	high agreement	$b \approx C^*$	genuine redundancy
(B) shared bias	may match (A)	$b < C^*$, floor	pseudo-redundancy (shared bias)
(C) block-correlated	may match (A)	$b \approx C^*/B$	$k_{eff} \approx k/B$

Table 2. Classification regime: agreement-curve diagnostics across three correlation structures. Pairwise summaries can be matched or made insufficiently diagnostic under plausible calibration choices; the curve-fit comparison provides the sharper test. Reproduction in Appendix A.

4.2 Continuous regime: Gaussian channel with three correlation structures

The same three correlation structures, in continuous form:

(A) Independent Gaussian channels. $Y_i = \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma^2 = 1$, mutually independent. Predicted: $\mathrm{Var}(\bar{Y}_k) = \sigma^2/k \to 0$.

(B) Shared-bias Gaussian channels. $Y_i = \delta + \eta_i$ with $\delta \sim N(0, v_\delta)$, $v_\delta = 0.04$, and $\eta_i \sim N(0, \sigma^2 - v_\delta)$ independent. Marginal per-channel variance is $\sigma^2 = 1$. Predicted: $\mathrm{Var}(\bar{Y}_k) = (\sigma^2 - v_\delta)/k + v_\delta \to v_\delta = 0.04$.

(C) Block-correlated Gaussian channels. Channels grouped into blocks of size $B = 4$ with identical noise within blocks; independent across blocks. Predicted: $\mathrm{Var}(\bar{Y}_k) = \sigma^2 B/k = \sigma^2/k_{\mathrm{eff}} \to 0$ at rate B/k.

Result. Empirical $\mathrm{Var}(\bar{Y}_k)$ under 50,000-trial simulation matches the three predicted curves to high precision. At $k = 128$: regime (A) → 0.0078 (predicted 0.0078); regime (B) → 0.0473 (predicted 0.0475, with floor $v_\delta = 0.04$); regime (C) → 0.0313 (predicted 0.0313, with $k_{\mathrm{eff}} = 32$). Pairwise correlation between channel outputs cannot reliably distinguish the three regimes when calibrated to common marginal variance, but the variance-reduction comparison does. The shared-bias floor is detected only because the predicted variance is computed from an independent noise model — within-data residuals do not reveal it.

Regime	Predicted $\mathrm{Var}(\bar{Y}_k)$	Long-k limit	Diagnosis
(A) iid Gaussian	σ^2/k	→ 0	genuine redundancy
(B) shared bias	$(\sigma^2 - v_\delta)/k + v_\delta$	→ v_δ	pseudo-redundancy (shared bias)
(C) block-correlated	$\sigma^2/(k/B) = B\sigma^2/k$	→ 0 at rate B/k	$k_{\mathrm{eff}} \approx k/B$

Table 3. Continuous regime: variance-reduction diagnostics across three correlation structures. The shared-bias floor is the operational signature of additive shared structure; it is detectable only against an independent noise-model prediction, not from within-data residuals.

5. Application: Planck PR3 component-separated CMB maps (continuous regime)

The Planck PR3 release provides four CMB temperature reconstructions (Commander, NILC, SEVEM, SMICA) produced by independent component-separation pipelines from shared multi-frequency observations (Planck Collaboration 2020). Half-mission splits provide a within-pipeline redundancy probe; the four full-mission products provide a cross-pipeline probe. Maps are distributed in HEALPix format (Górski et al. 2005).

Existing analyses report cross-product agreement in terms of pairwise correlation (mean off-diagonal $\bar{r} \approx 0.98$ for half-mission splits and $\bar{r} \approx 0.99$ for full-mission products under the PR3 common intensity mask) and a convergence curve defined as RMS deviation of k-map subset means from the full-set mean (Jones 2026d, denoted D(k) in that paper; in the present notation this corresponds to $\sqrt{V(k)}$). These statistics

describe the agreement and its rate of convergence under aggregation, but they do not test whether the convergence rate matches what would be predicted from an independent channel model.

Because Planck pipelines produce continuous maps, the protocol applies in the continuous regime of §2.2, not the classification regime of §2.1. The fitted exponent b should not be compared directly to a Chernoff information C*; that would mix mathematical objects. The continuous-regime diagnostic instead compares observed residual variance against a covariance-model prediction.

Step 1 (channel model). Construct per-pipeline residual covariance from Planck FFP10 end-to-end simulations or pipeline-specific noise products. Treat the per-pixel signal as the latent and per-pipeline residual (instrument noise + foreground residual + pipeline systematics) as the channel-specific perturbation. Aggregate to the predicted variance reduction curve $V^*(k) = \sigma^2(\ell)/k_{eff}(\ell) + F(\ell)$, where the ℓ dependence captures multipole structure under the PR3 common intensity mask; $F(\ell)$ is the predicted floor from known shared upstream structure.

Step 2 (agreement curve). Construct V(k) using a reference $\hat{X}_{ref}$ that is independent of the maps being aggregated. Three options, in increasing order of independence: (i) leave-one-pipeline-out — average k pipelines, compare to the held-out pipeline, average over folds; (ii) FFP10 simulation truth — average k pipelines, compare to the input simulation map; (iii) noise-model expectation — compare observed $\mathrm{Var}(\bar{Y}_k)$ to predicted $V^*(k)$ directly. Subset-versus-full-set RMS, as currently reported in Jones 2026d, is not an independent reference and should not be used as the falsifier statistic. Vary k from 1 to 8 (half-mission set) and 1 to 4 (full-mission set), reporting V(k) per multipole bin or per spatial region with appropriate effective sample-size correction.

Step 3 (k_eff audit). Compute both the simple uniform-correlation k_eff and a more careful audit that accounts for known shared upstream structure: the four pipelines share input frequency maps, mask choices, and processing assumptions; half-mission splits share time-stream-correlated systematics within each pipeline. Report spatial / multipole effective sample size (sky pixels are spatially correlated after masking, smoothing, and component separation; the effective independent spatial degrees of freedom can be substantially smaller than the pixel count).

Step 4 (falsifier comparison). Compare V(k) to $V^*(k)$. The expected outcome is that V(k) follows $V^*(k)$ within sampling uncertainty if the shared-systematics model $F(\ell)$ in $V^*(k)$ is adequate; otherwise $V(k) > V^*(k)$ at large k, with the excess floor $F_{obs} - F_{model}$ indicating unmodeled shared structure. The empirical question is therefore not whether the four pipelines share structure (they do, by construction) but whether the structure they share matches the independently audited model. A close match supports the audit; an excess floor identifies what the audit missed.

This test is constructive rather than destructive. The four Planck pipelines are not claimed to be conditionally independent in the strict sense: they share input frequency maps and mask choices. The protocol formalizes how much of the inter-pipeline agreement is attributable to genuine algorithmic independence versus shared inputs, and whether the residual shared component matches independent audit. With k = 4 (or k = 8 for half-mission splits), channel-count statistics will be limited; the bulk of the inferential power must come from spatial / multipole structure of the residuals, with appropriate spatial k_eff treatment.

Numerical execution of Steps 1–4 against the existing Planck PR3 analysis (Jones 2026d) is the subject of a companion empirical paper. The present paper specifies the protocol and demonstrates its diagnostic power on the synthetic cases of Section 4.

6. Discussion

6.1 Domains

The protocol applies to any setting with multiple putatively independent channels measuring a shared latent quantity. Direct applications include multi-detector physics experiments (gravitational-wave detectors, multi-frequency astronomical observations, multi-pipeline component separation); sensor fusion (Chair & Varshney 1986); multi-rater human annotation (Cohen 1960; Dawid & Skene 1979); replicate measurement in metrology and reproducibility studies; and Quantum-Darwinism-style experiments measuring environmental redundancy (Zurek 2009). The choice of regime — classification or continuous — is determined by the outcome type, not the framework.

6.2 The independence audit as the load-bearing step

Neither the Chernoff bound nor the variance-reduction law is novel in itself. What this protocol contributes is the pairing of these textbook predictions with an independent channel model and a k_eff audit. The audit converts a descriptive agreement statistic into a falsifiable independence claim. Without independent channel characterization, the protocol can verify the form of the agreement curve (exponential decay, 1/k variance reduction) but not its rate or its predicted floor.

6.3 Comparison to alternative diagnostics

Cohen's kappa, intraclass correlation (Shrout & Fleiss 1979; McGraw & Wong 1996), and Bland–Altman analysis are pairwise measures and share the limitation that high pairwise agreement is consistent with both genuine independence and correlated pseudo-redundancy. Variance-component decomposition (e.g., generalizability theory) is more diagnostic but is typically calibrated to the same data being audited. Rasch and IRT models assume measurement structure that itself encodes independence claims requiring separate validation. Sensor-fusion error analysis (Chair & Varshney 1986) is closest to the present protocol; the contribution here is to make the channel-model-predicted scaling the explicit falsifier and to formalize the k_eff audit. Dawid & Skene's (1979) latent-class approach to observer error is complementary: it estimates per-rater error rates from agreement structure under specified independence assumptions, where the present protocol asks whether such independence assumptions are even compatible with the data. The proposed protocol is most useful when (a) multiple channels exist, (b) an independent channel model can be constructed, and (c) the question is whether agreement is genuine redundancy or shared structure.

7. Limitations

Six limitations are stated explicitly:

(1) Agreement curves do not prove metaphysical independence. The protocol tests whether observed scaling is consistent with a stated independence model. It cannot establish independence as a property of the world; only as a property of a model that has not been falsified by the data.

(2) k_eff is model-relative and audit-relative. Different audit choices (correlation correction, variance-component decomposition, block-model, design effect, mixing-rate bound) yield different k_eff. Sensitivity analysis across plausible audits is essential. A single k_eff number without audit specification is not interpretable.

(3) Continuous measurements require covariance models, not direct Chernoff exponents. Mixing classification-regime b-versus-C* comparison with continuous-regime data is a category error. The continuous regime requires V*(k) from a noise model.

(4) Small channel counts limit fitted-curve inference. With $k \leq 8$ (as in the Planck full and half-mission cases), channel-count scaling alone provides limited statistical power. The bulk of inferential power must come from spatial, temporal, or sample structure within each channel comparison, with appropriate effective-sample-size treatment.

(5) Shared upstream preprocessing can dominate apparent agreement. When channels share input data, masks, calibration steps, or training data, agreement at the output stage may reflect that shared upstream structure rather than the channels themselves. The audit must include the full processing chain, not only the final-stage models.

(6) A fitted curve can reject an independence model but cannot identify the exact confound. A V(k) that saturates above V*(k) tells the analyst there is unmodeled shared structure but does not tell them which structure. Identifying the confound requires further audit (independent residual analysis, simulation injection tests, ablations of pipeline components).

8. Reporting template

The following template specifies the minimum information that should accompany any use of the protocol. The format is modeled on UIS reporting (Jones 2026c).

Latent quantity	What is shared across channels?
Channel units	What counts as a channel? (instrument, pipeline, rater, run)
Outcome regime	Classification (finite decision) or continuous (real-valued / vector); selects scaling law
Agreement statistic	Pairwise disagreement, variance, RMS, MSE, Mahalanobis, kappa, ICC, etc.
Predicted scaling	$a \exp(-bk)$, $A/k\_eff + F$, $\sqrt{(A/k\_eff + F)}$, $w^T \Sigma w + F$, etc.
Channel-model source	Noise simulations, calibration data, validation set, theoretical bound
Shared-structure model	What floor F or covariance terms are expected from known shared inputs?
Independence audit	Shared instruments, preprocessing, masks, training data, time streams
k_eff method	Uniform residual-correlation, variance-component, block-model, covariance-based, design effect; independent of the agreement curve being tested. $\bar{r}$ must be on residuals, not raw outputs

Effective independent units	Channel k_eff plus spatial / temporal / sample n_eff, with method of estimation for each
Reference signal	Simulation truth, leave-one-out, model expectation; not subset-versus-full-set. Reference uncertainty must be modeled if the reference is itself noisy
Falsifier	What curve mismatch rejects independence? (e.g., $V(k) > V^*(k)$ by $\geq 3\sigma$); model-conditional
Claim level	Method demo / Level 3 empirical application (UIS terminology)
UIS ledger link	Completed entry in independence-audit ledger?

Table 4. Reporting template for the agreement-curve protocol. Every entry should be filled before the diagnostic is treated as informative.

9. Broader context

This protocol was developed within a structural framework on records and constraint-guided collapse (Jones 2026a, 2026b, 2026c, 2026e) as a falsifier for the redundancy-implies-consensus signature S_1 of records-based stabilization. The methods contribution stands independently: an audit for redundancy claims in any domain where multiple channels measure a shared latent. The internal framework is one motivating context; the protocol's correctness and applicability do not depend on adopting it.

Appendix A. Reproduction code

The following Python script reproduces the classification regime (§4.1, three BSC regimes with predicted $\exp(-kC^*)$ overlay) and the continuous regime (§4.2, three Gaussian regimes with predicted variance-reduction curves). Dependencies: NumPy and Matplotlib only. Random seed fixed for reproducibility.

```
import numpy as np
import matplotlib.pyplot as plt

# ---------- §4.1 Classification regime (BSC) ----------
P_BASE = 0.40
SIGMA_DELTA_BSC = 0.05
BLOCK_SIZE = 4
N_TRIALS_BSC = 20000

def C_bsc(p):
    return -np.log(2 * np.sqrt(p * (1 - p)))

C_STAR = C_bsc(P_BASE)

def map_decision(Y, rng):
    half = Y.shape[1] / 2.0
    s = Y.sum(axis=1)
    Xhat = (s > half).astype(int)
    ties = (s == half)
    if ties.any():
        Xhat[ties] = rng.integers(0, 2, size=ties.sum())
```

```
    return Xhat

def disagreement_iid(k, rng):
    X = rng.integers(0, 2, size=N_TRIALS_BSC)
    f1 = (rng.random((N_TRIALS_BSC, k)) < P_BASE).astype(int)
    f2 = (rng.random((N_TRIALS_BSC, k)) < P_BASE).astype(int)
    Y1 = X[:, None] ^ f1; Y2 = X[:, None] ^ f2
    return (map_decision(Y1, rng) != map_decision(Y2, rng)).mean()

def disagreement_shared_bias(k, rng):
    X = rng.integers(0, 2, size=N_TRIALS_BSC)
    delta = rng.normal(0.0, SIGMA_DELTA_BSC, size=N_TRIALS_BSC)
    p_eff = np.clip(P_BASE + delta, 0.01, 0.49)[:, None]
    f1 = (rng.random((N_TRIALS_BSC, k)) < p_eff).astype(int)
    f2 = (rng.random((N_TRIALS_BSC, k)) < p_eff).astype(int)
    Y1 = X[:, None] ^ f1; Y2 = X[:, None] ^ f2
    return (map_decision(Y1, rng) != map_decision(Y2, rng)).mean()

def disagreement_block(k, rng, B=BLOCK_SIZE):
    X = rng.integers(0, 2, size=N_TRIALS_BSC)
    n_b = (k + B - 1) // B
    bf1 = (rng.random((N_TRIALS_BSC, n_b)) < P_BASE).astype(int)
    bf2 = (rng.random((N_TRIALS_BSC, n_b)) < P_BASE).astype(int)
    bi = np.arange(k) // B
    Y1 = X[:, None] ^ bf1[:, bi]; Y2 = X[:, None] ^ bf2[:, bi]
    return (map_decision(Y1, rng) != map_decision(Y2, rng)).mean()

# ---------- §4.2 Continuous regime (Gaussian) ----------
SIGMA_TOTAL = 1.0
V_DELTA = 0.04
SIGMA_INDEP = np.sqrt(SIGMA_TOTAL**2 - V_DELTA)
N_TRIALS_GAUSS = 50000

def var_iid_gauss(k, rng):
    Y = rng.normal(0, SIGMA_TOTAL, size=(N_TRIALS_GAUSS, k))
    return Y.mean(axis=1).var()

def var_shared_gauss(k, rng):
    delta = rng.normal(0, np.sqrt(V_DELTA), size=N_TRIALS_GAUSS)
    eps = rng.normal(0, SIGMA_INDEP, size=(N_TRIALS_GAUSS, k))
    return (delta[:, None] + eps).mean(axis=1).var()

def var_block_gauss(k, rng, B=BLOCK_SIZE):
    n_b = (k + B - 1) // B
    block_eps = rng.normal(0, SIGMA_TOTAL, size=(N_TRIALS_GAUSS, n_b))
    bi = np.arange(k) // B
    return block_eps[:, bi].mean(axis=1).var()

# ---------- run + plot ----------
rng = np.random.default_rng(0)
```

```
ks_bsc = list(range(1, 161, 4))
A = [disagreement_iid(k, rng) for k in ks_bsc]
B = [disagreement_shared_bias(k, rng) for k in ks_bsc]
C = [disagreement_block(k, rng) for k in ks_bsc]

ks_g = [1, 2, 4, 8, 16, 32, 64, 128]
Ag = [var_iid_gauss(k, rng) for k in ks_g]
Bg = [var_shared_gauss(k, rng) for k in ks_g]
Cg = [var_block_gauss(k, rng) for k in ks_g]

pred_bsc = np.exp(-np.array(ks_bsc) * C_STAR)
pred_iid_g = SIGMA_TOTAL**2 / np.array(ks_g)
pred_shared_g = SIGMA_INDEP**2 / np.array(ks_g) + V_DELTA
pred_block_g = SIGMA_TOTAL**2 / np.maximum(np.array(ks_g) / BLOCK_SIZE, 1)

fig, axes = plt.subplots(1, 2, figsize=(13, 5))

ax = axes[0]
ax.semilogy(ks_bsc, A, "o-", label="(A) iid BSC")
ax.semilogy(ks_bsc, B, "s-", label="(B) shared bias")
ax.semilogy(ks_bsc, C, "^-", label="(C) block-correlated")
ax.semilogy(ks_bsc, pred_bsc, "k--", alpha=0.6, label="predicted exp(-kC*)")
ax.set_xlabel("k"); ax.set_ylabel("Pr[disagree]")
ax.set_title("Classification regime"); ax.legend()

ax = axes[1]
ax.loglog(ks_g, Ag, "o-", label="(A) iid Gaussian")
ax.loglog(ks_g, Bg, "s-", label="(B) shared bias")
ax.loglog(ks_g, Cg, "^-", label="(C) block-correlated")
ax.loglog(ks_g, pred_iid_g, "k--", alpha=0.5, label="pred iid")
ax.loglog(ks_g, pred_shared_g, "b--", alpha=0.5, label="pred shared")
ax.loglog(ks_g, pred_block_g, "g--", alpha=0.5, label="pred block")
ax.set_xlabel("k"); ax.set_ylabel("Var(Ybar_k)")
ax.set_title("Continuous regime"); ax.legend()

plt.tight_layout(); plt.show()
```

References

Audenaert, K. M. R., Calsamiglia, J., Muñoz-Tapia, R., Bagan, E., Masanes, L., Acin, A., & Verstraete, F. (2007). Discriminating states: The quantum Chernoff bound. Physical Review Letters, 98(16), 160501. https://doi.org/10.1103/PhysRevLett.98.160501

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet, 327(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Chair, Z., & Varshney, P. K. (1986). Optimal data fusion in multiple sensor detection systems. IEEE Transactions on Aerospace and Electronic Systems, AES-22(1), 98–101. https://doi.org/10.1109/TAES.1986.310699

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. The Annals of Mathematical Statistics, 23(4), 493–507. https://doi.org/10.1214/aoms/1177729330

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

Cover, T. M., & Thomas, J. A. (2006). Elements of Information Theory (2nd ed.). Wiley.

Dawid, A. P., & Skene, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM algorithm. Journal of the Royal Statistical Society: Series C (Applied Statistics), 28(1), 20–28. https://doi.org/10.2307/2346806

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. The Astrophysical Journal, 622(2), 759–771. https://doi.org/10.1086/427976

Jones, J. C. (2026a). Records Across Nature, Life, and Mind. HoldingLight LLC. https://doi.org/10.17605/OSF.IO/7H6DY

Jones, J. C. (2026b). The Structuralization of Empiricism. HoldingLight LLC. https://doi.org/10.17605/OSF.IO/J4GZ9

Jones, J. C. (2026c). Update Integrity Standard (UIS v1.0). HoldingLight LLC. https://doi.org/10.17605/OSF.IO/DWM29

Jones, J. C. (2026d). Redundant Record Consensus in Planck PR3 CMB Component Maps [Forthcoming companion empirical paper]. HoldingLight LLC.

Jones, J. C. (2026e). Objectivity from Records: An Exponential Consensus Bound (UCT Technical Note v1.0). HoldingLight LLC. https://doi.org/10.17605/OSF.IO/6M7N3

Kish, L. (1965). Survey Sampling. Wiley.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Nussbaum, M., & Szkoła, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. The Annals of Statistics, 37(2), 1040–1057. https://doi.org/10.1214/08-AOS593

Planck Collaboration (2020). Planck 2018 results. IV. Diffuse component separation. Astronomy & Astrophysics, 641, A4. https://doi.org/10.1051/0004-6361/201833881

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Zurek, W. H. (2009). Quantum Darwinism. Nature Physics, 5, 181–188. https://doi.org/10.1038/nphys1202

This paper is part of the Universal Collapse Theory library. For a reading guide and full architecture, visit universalcollapse.com/roadmap.

AI Disclosure

AI tools were used to assist with manuscript preparation. The underlying theory, arguments, and interpretive claims are the author's own, and the author takes full responsibility for the content.

Citation

Jones, J. C. (2026). Auditing Independence in Multi-Channel Measurement: An Agreement-Curve Diagnostic for Genuine versus Correlated Redundancy (UCT Methods Paper v1.0). HoldingLight LLC.