

Consciousness, Closure, and Remainder

A framework for studying stable conscious content under uncertain observation

C. F. Dietz

Version 2.1 | September 2026 | Author-review manuscript

Abstract

Apparent stability can reflect persistence in a system, persistence in its environment, or measurements that conceal relevant changes. Consciousness, Closure, and Remainder (CCR) provides a framework for distinguishing these possibilities when studying conscious content. It separates the occurrence of experience from its organization. Closure stabilizes distinctions, identities, and permitted transitions under stated conditions. Numerical remainder is the lowest mismatch a specified model family can attain or approach. The framework distinguishes the organization a system occupies, the best fit its permitted candidates allow, and the candidate preferred after costs are included. It also treats observation and transition assumptions as parts of the model that can fail. In a controlled two-stream Gaussian study, a joint model estimates both an uncertain target and the occupied state. Across 144 fresh simulated cases, it selected inertia in 0 of 24 no-inertia cases at each measurement setting. It selected real generating inertia in 24 of 24 favorable, 2 of 24 moderate, and 0 of 24 severe cases. Removing target history restored misleading selections. Joint inference improved state recovery without reliably identifying the mechanism across conditions. The contribution is a measurement and diagnostic framework built from established statistical methods, not a universal consciousness measure. A prospective perceptual-organization protocol specifies how model states could be tested against independently assessed experience.

Keywords: conscious content; closure; remainder; hidden-state inference; measurement; persistence; model adequacy; perceptual organization.

1. What makes an organization persist?

A system can appear stable for different reasons. Its environment may remain stable, its own organization may resist change, or measurements may conceal changes that matter. These explanations can coexist. CCR asks how to distinguish them when studying conscious content. A percept can remain recognizable while sensory support changes; a category can apply across examples; a goal can persist despite distraction. The task is to specify and test that organization, not merely find a statistical pattern that appears stable.

Existing consciousness theories address important parts of this task. Global neuronal workspace theory concerns broad availability and access; recurrent processing emphasizes feedback; integrated information theory proposes an intrinsic cause-effect account; and predictive approaches model inference and adjustment under uncertainty (Dehaene and Changeux, 2011; Lamme, 2006; Albantakis et al., 2023; Neacsu et al., 2022). CCR adds an intermediate description of the distinctions maintained, the changes in organization, and the limits of specified model families. It does not replace those theories by renaming their outcomes.

The question is not only which state a system occupies, but why it persists and what could make it change. Answering it requires distinguishing the current organization, the available alternatives, the transition process, and the measurements used to infer them. Perceptual, task, report, and motor organization can occur together without being the same process. A stable neural partition therefore does not by itself identify a conscious percept.

The paper combines conceptual distinctions, formal definitions, and a bounded synthetic study. Its central result is a contrast: modeling a persistent but uncertain target can remove one source of misleading inertia attribution and improve state estimates, yet still leave weak evidence for intrinsic inertia. The following sections explain the distinctions, state their mathematical conditions, and test their recovery from measurements. The prospective protocol then asks what would connect those results to conscious content.

2. The distinctions the framework needs

2.1 Presence and content

C denotes the occurrence of experience considered apart from its particular contents. c denotes that occurrence as organized perception, affect, memory, bodily orientation, or other determinate content. They are not separate substances. The distinction establishes neither an independently existing content-free state nor a universal subject or an ontologically privileged substrate.

CCR treats C as a methodological primitive: its equations do not derive experience from graphs, likelihoods, or computations. Explaining cognitive organization and explaining why experience occurs remain different tasks (Nagel, 1974; Chalmers, 1995). This is a declared starting point, not a characterization every philosophical position must accept. The companion CCC paper argues for the stronger consciousness-first ontology and cosmological proposal (Dietz, 2026a).

CCR instead works within an assumed physical world; its methods do not require accepting that ontology.

M denotes openness as a limit on claims that finite descriptions exhaust reality. m denotes the unresolved horizon of a situated perspective. Neither is treated as an absolute residual in a brain. Numerical remainder R compares a specified target and model family; with a justified interpretation, it may capture one aspect of m. Its success or failure does not decide the ontological status of C or M. CCC's stronger philosophical remainder conjecture is a separate argument, not a consequence of this statistic.

2.2 Closure, states, and grammars

A closure stabilizes distinctions, identity criteria, and permitted transitions so that a specified organization can persist under stated conditions. Following Lawson (2001), it concerns workable distinctions. This is not deductive or topological closure, nor a claim that physical causation excludes nonphysical causes. Biological closure of constraints is a related but distinct account of mutually maintained causal organization (Montevil and Mossio, 2015).

A closure state is a particular organized configuration. A grammar specifies which states, model families, transformations, and transition rules are permitted. A poor state within an adequate grammar differs from a grammar that excludes every adequate state. Delayed switching may explain poor occupancy; it is not a third, mutually exclusive kind of failure. Several problems can coexist.

The closure triad concerns capacity, stability, and consistency. Capacity asks which distinctions can be sustained. Stability asks how organization persists under stated conditions; variational stability is the local approach that ranks candidates by fit and cost. Consistency asks which relations must remain compatible within a system or across agents. This distinguishes the broader notion of stability used in CCC from the particular variational model used here. The three concerns are analytical standpoints, not a proven set of necessary and sufficient conditions.

Nested domains concern enabling relations, not a derivation of presence. Biological, sensorimotor, perceptual, cognitive, phenomenal, and intersubjective identities may persist in different senses. An application must identify its domain before assigning psychological meaning to a fitted structure.

2.3 Claim types

Table 1. Claims and their required forms of support.

Claim type	Example	Appropriate evaluation
Starting point	Experience occurs	Declared methodological posture
Conceptual	Presence differs analytically from content organization	Coherence, objections, and explanatory usefulness
Formal	A nested model family cannot have larger optimal fit error than its subfamily	Proof under explicit assumptions
Synthetic	A fitted pipeline recovers a known simulated state	Reproducible recovery and failure tests
Empirical bridge	A specified state metric predicts independently assessed perceptual organization	Prospective behavioral or neural evidence
Ontological interpretation	Presence is fundamental	A separate philosophical argument

Table 1 separates the support each kind of claim needs. A mathematical proof cannot validate a link to experience, and a successful simulation cannot establish an ontology. A quantity requires independent content validation before it can serve as evidence about conscious organization.

3. Representation and observation are part of the model

3.1 How a candidate is represented

Let X denote the observed degrees of freedom. A candidate $K = (\Pi, E)$ comprises a partition Π and an interaction graph E , and specifies a nonempty family $\mathcal{Q}(K)$ of probability or dynamical models. Network partitions and time-dependent communities have established precedents (Mucha et al., 2010; Bassett and Sporns, 2017). They serve here as a proxy, not as an assertion that conscious content literally is a covariance block.

An application must state its observation map: how the modeled organization connects to what can be measured. It must also declare its clock, permitted transformations, and identity criteria. Sensor channels, source estimates, cortical parcels, and behavioral features need not support the same candidate families. A representation can hide the relation the analysis aims to recover.

A fitted state's interpretation should survive specified identity-preserving transformations and fail under specified identity-destroying ones. Resolve equivalent candidates before calculating uncertainty. Giving one candidate several names can change an entropy over labels without adding any distinct organization.

3.2 What the model fits, and what the system occupies

The target is the structure against which fit is evaluated. The occupied organization is the structure attributed to the system at a given time. A target is not automatically the correct percept, and disagreement with it is not automatically error. Several percepts can be compatible with an ambiguous stimulus.

Evidence about target and occupancy must be justified separately. Inferring both from the same fitted pattern can make their agreement circular. The synthetic study uses two distinguishable streams: U measures a hidden target H , and Y measures a hidden occupied organization Z . An observation model specifies the measurements each hidden state can produce. Conditional independence of the streams is a design assumption, not something established by splitting one neural recording into arbitrary electrode groups.

Hidden-state probabilities express uncertainty within a fitted model, not across all possible observation maps, causal structures, or grammars. The model parameters have uncertainty too, even when reported as point estimates. Identifiability asks whether distinct explanations can be distinguished from the available observations. Good classification of states can coexist with poorly identified parameters or an incorrect explanation of transitions.

3.3 Report and conscious content

Reports provide evidence about experience while also involving task and motor organization. Pitts et al. (2014) separated awareness from task relevance and report-related effects. Kapoor et al. (2022) decoded percept-related transitions from macaque prefrontal activity in a no-report paradigm. These findings motivate controls and converging evidence. They validate neither CCR's quantities nor a blanket dismissal of prefrontal activity as report artifact.

Assign candidate variables a proposed role before interpreting them: perceptual, task, access, report, or motor. Delayed reports, changes in task relevance, and independently validated physiological markers can test whether a content interpretation generalizes. No-report markers are also inferential and can have confounds. The aim is converging evidence, not an infallible measure.

4. What the quantities measure

Table 2. A guide to the quantities defined in this section.

Question	Quantity or distinction
How well can this candidate family represent the target?	Remainder R: the lowest mismatch it can attain or approach.
What is the best fit among the permitted candidates?	The grammar's raw mismatch floor.
Which candidate is preferred once declared costs count?	The score-optimal candidate.
Which organization does the system actually occupy?	Observed or inferred occupancy.
How much worse is its fit than the best permitted fit?	Raw occupancy excess.

4.1 Remainder: the best fit within a candidate family

The first question is how well a candidate family can represent a fixed target. For target law p and nonempty family $Q(K)$, define

$$R(K | p) = \inf_{q \in Q(K)} D_{\text{KL}}(p || q). \quad (1)$$

The infimum is the lowest mismatch the family can attain or approach; a minimizing member need not exist. Divergences and their differences must be finite wherever the following comparisons require them. Relative entropy and constrained approximation are established tools (Kullback and Leibler, 1951). The direction of comparison matters: p is fixed while q varies. Results for another direction or family do not transfer merely because both are called information projections (Csiszar, 1975).

R depends on the target variables, representation, model family, grammar, and access conditions. It is nonnegative, but a finite constraint does not force a positive value. A restrictive family can contain p exactly or approach it arbitrarily closely. Neither case decides the philosophical claim that reality is not exhausted by a finite articulation.

R measures the best-achievable population fit, not a momentary prediction error or a causal force. A transition law may use a score involving R , but that law is an additional hypothesis. The definition of a diagnostic quantity does not explain what causes a system to change.

4.2 Best fit, preferred candidate, and actual occupancy

The next question is what the entire permitted candidate set can fit. For the finite, nonempty operative set $\mathcal{K}(\Omega, \mathcal{B})$, with $\mathcal{B}$ denoting declared admissibility or resource restrictions, define

$$R_{\text{floor}}(p) = \min_{K \in \mathcal{K}(\Omega, \mathcal{B})} R(K | p). \quad (2)$$

This floor describes what the permitted set can fit. It does not identify the state the system occupies.

Best fit need not determine preference when costs matter. A static score can add interaction and partition costs:

$$S_0(K | p) = R(K | p) + \lambda C_E(E) + \gamma C_\Pi(\Pi). \quad (3)$$

The fit-optimal candidate K_{fit} minimizes R . The score-optimal candidate K_{score} minimizes S_0 . They need not be the same. Statistical penalties and hypothesized biological resource costs also differ. Description-length and Bayesian criteria motivate statistical model choice; they do not directly measure an organism's energetic or computational budget (Rissanen, 1978; Schwarz, 1978).

Actual occupancy is a separate question. For occupancy probabilities $\pi(K)$ over the same operative set, define its average mismatch and excess above the best permitted fit:

$$R_{\text{occ}} = \sum_K \pi(K) R(K | p), \quad A_{\text{occ}} = R_{\text{occ}} - R_{\text{floor}}. \quad (4)$$

The excess compares actual or inferred occupancy with the best raw fit available. It measures a difference in fit, not the cause of that difference.

Score regret makes the corresponding comparison after the declared costs are included:

$$G_{\text{occ}} = \sum_K \pi(K) S_0(K | p) - \min_L S_0(L | p). \quad (5)$$

Score regret asks whether occupancy is worse under the full objective, including cost. A state can have positive raw excess and no score regret.

Both excesses are nonnegative by definition, but answer different questions. A system may accept a worse raw fit at a lower cost, so positive occupancy excess need not mean failure to occupy the score-optimal state. If occupancy is observed, π is a point mass. If hidden, π requires a justified observation model; it cannot simply be replaced with a static score distribution.

4.3 When the target is uncertain too

Uncertainty about the target changes how excess is averaged. Given conditional target laws p_h and joint target/occupancy posterior $\rho(h, K)$, define the posterior expected raw excess as

$$\mathbb{E}[A | D] = \sum_{h, K} \rho(h, K) [R(K | p_h) - R_{\text{floor}}(p_h)]. \quad (6)$$

This diagnostic is a proposed definition, not an established estimator for arbitrary neural recordings. Target and occupancy may depend on one another, so multiplying their marginal probabilities does not generally replace their joint posterior. There is another distinction: finding an optimum for each target and then averaging is not the same as averaging targets first and then optimizing. The operations assume different information is available to the decision maker.

The joint study also uses the simpler binary mismatch probability $P(H \neq Z | D)$. It is not a KL remainder and is not labeled R. Recovering it does not recover an absolute grammar floor or separate the causes of mismatch.

4.4 Why one residual cannot establish grammar adequacy

A grammar can fail in its target representation, occupied-state family, transition rules, or observation assumptions. If one permitted candidate contains the target, the raw floor is zero even when a complexity penalty discourages selecting it. A high-floor trigger therefore requires explicit restrictions on the operative set.

Exact fit to one-time observations does not establish a correct account of transitions. Two symmetric two-state chains can each spend half their time in either state while having stay probabilities of 0.9 and 0.1. Against the latter target, the fixed persistent transition model has conditional KL loss of about 1.757780 nats per update. A refitted family constrained to a stay probability of at least one-half still has minimum loss 0.368064; an unrestricted symmetric family fits exactly. These distinguish a poor fixed parameter, an inadequate restricted family, and an adequate family (Dietz, 2026b).

An expanded grammar should be a specified alternative tested on data not used to fit it. Expansion is not an unlimited response to any mismatch. Failure of one alternative to help can reflect low sensitivity, a weak alternative, or an observation error shared by both models. It does not certify the original grammar.

4.5 What held-out log loss estimates

In an empirical study, the population target is unknown. One useful comparison evaluates a fitted candidate against a fitted reference on held-out observations, meaning data excluded from fitting. The excess log loss is

$$\hat{L}_{\text{rel}}(K) = \frac{1}{n} \sum_{i=1}^n \log \frac{\hat{q}_{\text{ref}}(x_i)}{\hat{q}_K(x_i)}. \quad (7)$$

For fixed fitted models, averaging this quantity over the population gives

$$\mathbb{E}_p[\log \hat{q}_{\text{ref}} - \log \hat{q}_K] = D_{\text{KL}}(p \| \hat{q}_K) - D_{\text{KL}}(p \| \hat{q}_{\text{ref}}). \quad (8)$$

This estimates a difference in mismatch, not automatically the candidate family's infimum. It can be negative even in the population when the reference fits worse. Fitting uncertainty, optimization, choice of reference family, and candidate selection all affect its interpretation.

Estimation error, approximation, and closure-imposed restrictions can each contribute to mismatch, but are not automatically additive or separately identifiable. Temporal dependence also means overlapping windows are not independent subjects. The unit of inference should determine the split: independent sessions, held-out participants or sites, nested selection, and appropriate null simulations. This follows the concerns raised in time-varying connectivity and model-recovery work (Hindriks et al., 2016; Wilson and Collins, 2019).

5. What the definitions establish

5.1 A Gaussian example

Gaussian distributions make the fit calculation explicit. For zero-mean laws with positive-definite covariances,

$$D_{\text{KL}} = \frac{1}{2} \left[\text{tr}(\Sigma_K^{-1} \Sigma) - d + \log \frac{\det \Sigma_K}{\det \Sigma} \right]. \quad (9)$$

This compares the target covariance with a candidate covariance. The next step minimizes that comparison within the permitted block structure.

When a candidate requires independent blocks, its best covariance preserves the target covariance within each block. Substituting that optimum gives

$$R(\Pi \mid \Sigma) = \frac{1}{2} \left[\sum_{B \in \Pi} \log \det \Sigma_{BB} - \log \det \Sigma \right]. \quad (10)$$

The result has an established interpretation. For unrestricted within-block densities, the best factorized approximation is the product of the block marginals. Its divergence is total correlation across blocks, or mutual information for two blocks (Watanabe, 1960). CCR uses this measure rather than proposing a new one.

5.2 The same mismatch can require different remedies

Consider four variables with pairings $A=(12)(34)$, $B=(13)(24)$, and $C=(14)(23)$. Each target has unit variances, correlation 0.6 within its pairs, and zero other covariance. These are the companion CCC examples (Dietz, 2026a, Appendix A). Admit only A and B to the operative grammar.

Table 3. Population diagnostic separation in a restricted grammar.

Target and occupied candidate	R_{occ}	R_{floor}	A_{occ}
A, occupied A	0.000000	0.000000	0.000000
A, occupied B	0.446287	0.000000	0.446287
C, occupied A	0.446287	0.446287	0.000000

The two positive occupied residuals are equal, but the available remedies differ. In the second row, switching to another permitted candidate removes the mismatch. In the third, exact fit needs a pairing the grammar excludes. Neither row explains why the current state is occupied. The separation follows from the declared families and is also available to a conventional analyst using them. All values are population divergences in nats.

5.3 Supported structure and changes of representation

The eight-variable calibration has two planted groups of four, within-group correlation 0.65, and cross-group correlation 0.05. Among 35 balanced partitions, the planted remainder is 0.0023034881 and the worst mixed value is 0.9681911611. When every pair instead has correlation 0.25, all balanced partitions give 0.1976563683. At illustrative effective sample size 200, the normalized BIC-style score is 0.2672193564 for the planted partition and 0.4768485630 for the unrestricted one-block covariance model (Dietz, 2026e).

These population calculations recover a supplied structural distinction and preserve ambiguity when the target is symmetric. They do not establish recovery from noisy neural data. With increasing sample size, a decreasing per-observation BIC penalty can eventually favor a full model over a slightly misspecified structured model. That statistical behavior says nothing about biological resource costs disappearing as an analyst collects data.

An invertible measurable change of coordinates T preserves KL remainder when both the target and every candidate family are transformed with it. Applying the old matrix restrictions in the new coordinates usually asks a different question. In the four-variable 45-degree rotation example, leaving the restrictions unchanged raises the floor from zero to 0.247836 nats. Transforming the family consistently preserves zero (Dietz, 2026b). Channel loss and added noise are different operations, not invertible re-expression.

5.4 Four conditional guarantees

Four results summarize useful checks. First, enlarging a family gives $R(Q_1 | p) \geq R(Q_2 | p)$ when $Q_1 \subseteq Q_2$, for the same target. Second, forced factorization has nonnegative cost if it restricts the family and the difference is finite. Third, a unique static winner survives score perturbations bounded below half its gap to the next candidate. Fourth, positive inertia favors remaining in place at a two-state static tie when self-distance is zero and inverse temperature is positive.

Appendix A states the conditions and proofs. They concern the mathematical comparison, not phenomenal unity or biological maintenance. In particular, adding models for the same target is not the same as observing new variables and changing the target. Hiding variables can conceal dependence and reduce measured mismatch.

6. From static scores to dynamics

6.1 A proposed transition law

A ranking alone does not specify movement. Let K_t be the occupied candidate, L a possible destination, and $d(L, K_t)$ a nonnegative transition distance with zero self-distance. Define a minimal transition energy:

$$J(L | K_t, p_t) = S_0(L | p_t) + \eta d(L, K_t). \quad (11)$$

The transition rule turns those energies into probabilities. For $\beta > 0$ and $\eta \geq 0$,

$$P(K_{t+1} = L | K_t, p_t) = \frac{\exp[-\beta J(L | K_t, p_t)]}{\sum_H \exp[-\beta J(H | K_t, p_t)]}. \quad (12)$$

This rule specifies sensitivity to scores and a cost of changing state. It leaves the evolution of p_t , any synaptic implementation, and the physical duration of an update unspecified. Distance units matter: rescaling distance can be offset by rescaling its coefficient η . The two must therefore be fixed or identified together.

Three kinds of uncertainty must remain distinct: ambiguity in static scores, uncertainty about current occupancy given measurements, and the long-run occupancy distribution of a dynamical model. The static distribution proportional to $\exp[-\beta S_0(K)]$ is not generally the stationary distribution of this inertial chain. Appendix A shows why.

CCC 5.4, Appendix C, illustrates a different continuous-time transition law whose stationary distribution is score-based (Dietz, 2026a). That result does not apply to the discrete-time inertial rule here. Both examples distinguish ranking a candidate from specifying the process that occupies it.

6.2 The equivalent two-state model

For two states with unit cross-state distance, define the score advantage $x_t = S_0(A | p_t) - S_0(B | p_t)$. Let $s_t = -1$ for currently occupied A and $+1$ for B. The transition log odds become

$$\text{logit}P(K_{t+1} = B) = \beta x_t + \beta \eta s_t. \quad (13)$$

This is a constrained logistic state-transition model. The same inputs and restrictions produce the same predictions in a conventional implementation. Input-conditioned transitions and persistent latent-state models are established methods (Bengio and Frasconi, 1994; Fox et al., 2011), not new equations claimed by CCR.

If $x_t = 0$ throughout, the transition probabilities identify only the product $\beta\eta$, not the individual coefficients. Varying x can help distinguish sensitivity from transition cost when score units, distance, and observation assumptions are independently fixed. Recovery in a favorable simulation does not remove this identification limit in a different design.

A switch also need not follow a narrowing score gap. At fixed equal scores, stochastic transitions can occur while the gap and static entropy stay unchanged. A predicted pre-switch trajectory therefore requires a further model of evidence or adaptation. It does not follow from normalizing transition probabilities alone.

6.3 Why persistent targets generate persistent responses

A persistent target can make responses persist even without intrinsic inertia. To show this, let $h_t \in \{-1, +1\}$ be a stationary symmetric target chain with switch probability q . Let $z_t \in \{-1, +1\}$ be conditionally independent across time given the target history, with

$$P(z_t = +1 | h_t) = \text{sigmoid}(ah_t). \quad (14)$$

There is no intrinsic dependence on z_{t-1} . Writing $\alpha = \tanh(a/2)$, conditional expectation gives $\mathbb{E}[z_t | h_t] = \alpha h_t$, and therefore

$$\mathbb{E}[z_t z_{t-k}] = \alpha^2 (1 - 2q)^k, \quad k \geq 1. \quad (15)$$

The responses are serially related because the targets are, not because an occupancy-inertia term generates them. For positive lag-one correlation r_1 , the best symmetric first-order summary has stay probability $(1 + r_1)/2$ and positive logistic persistence coefficient $\log[(1 + r_1)/(1 - r_1)]$. This summarizes one-step behavior; it does not make the response process itself first-order Markov.

Equation (15) explains the mechanism behind the simulations under its stated assumptions. It is not a new general hidden-Markov theorem. A model omitting target history can use a persistence term to improve prediction without recovering an intrinsic occupancy mechanism. The next question is whether noisy measurements allow the hidden target and state to be inferred without giving their labels to the analyst.

7. Joint target and occupancy study

7.1 What was simulated and what was observed

The synthetic study tests that question in a restricted setting. H is an A/B target organization and Z an A/B occupied organization. A pairs channels (1,2) and (3,4); B pairs (1,3) and (2,4). Source covariances have unit diagonal, within-pair correlation 0.6, and zero other covariance. H switches with probability 0.06 per update. After initialization, Z follows

$$\text{logit}P(Z_t = B \mid H_t, Z_{t-1}) = 2(2H_t - 1) + b(2Z_{t-1} - 1). \quad (16)$$

The labels are A=0 and B=1. The no-inertia generator has $b=0$ and the real-inertia generator $b=1.5$. Each sequence starts with an independent fair occupied state. U and Y are separate four-channel Gaussian measurement packets produced by H and Z. Readout noise is independent Gaussian noise, the streams are conditionally independent, and the vectors within each packet are independent. These are assumptions of the synthetic design.

Table 4. Joint access conditions in the simulation.

Setting	Vectors per packet	Added noise variance	Observed paired correlation
Favorable	32	0.25	0.48
Moderate	8	1.00	0.30
Severe	4	3.00	0.15

The condition names describe relative measurement quality in this simulation. Packet size and noise change together, so they do not isolate either factor's effect or define EEG quality thresholds. An update has no established duration in milliseconds.

Each case has four training sequences of 256 updates, eight validation sequences of 128 updates, and four test sequences of 512 updates. The paired cohort regenerates 144 cases from the earlier study; the fresh cohort supplies 144 independent cases. Each has 24 replications per

mechanism and measurement setting, and they are reported separately. An additional 48 paired moderate-condition analyses remove target history. True H and Z labels are withheld from fitting and selection, then used to assess recovery (Dietz, 2026c, 2026d).

7.2 What the models learn

The earlier approach computes a target-score difference from each U packet, normalizes it using training data, and inserts that estimate into a hidden occupied-state model. This is the plug-in approach: it uses one estimated score rather than retaining the target's uncertainty and history jointly. M0 uses the score alone; M1 adds a nonnegative persistence coefficient. Both learn Y emissions without the true occupied-state labels.

The joint model instead keeps probabilities over all four possibilities, $(H, Z) \in \{A, B\}^2$. It models target history and the occupied state's response together:

$$P(h, z \mid h', z') = Q_{h'h} \text{Bern}(z; \sigma[a(2h - 1) + b(2z' - 1)]). \quad (17)$$

Q contains two learned target-transition rows. U depends on H, Y depends on Z, and both streams' block-Gaussian covariances are learned. J0 allows a persistent uncertain target but fixes occupancy inertia $b=0$. J1 estimates nonnegative b under the same observation assumptions. The models estimate initial target probabilities while retaining the design's uniform initial occupancy.

Equation (17) uses a categorical target sign. This is an observation-model extension, not end-to-end estimation of the continuous CCR remainder landscape. Its b is a logit coefficient, not automatically the transition-cost parameter η in Equation (12). Connecting them requires the appropriate score and distance conventions.

Training uses expectation-maximization to maximize the joint likelihood of U and Y, with an exact four-state forward-backward E-step and two initializations. Smoothing estimates past states using later observations; it is used only on completed training sequences. Filtering estimates the current state from observations available so far. Validation and test therefore assimilate current U, score current Y, and only then assimilate Y. No future observations enter these state estimates. Parameters are point estimates, not a full Bayesian posterior over parameters.

The resulting model is a coupled four-state hidden Markov model. Multiple hidden variables are established machinery (Ghahramani and Jordan, 1997); this is neither a new filtering algorithm nor an independent-chain factorial HMM. A conventional representation with the same four joint states provides the numerical equivalence control.

7.3 How selection differs from evaluation

Validation evaluates $\log p(Y_t \mid U_{1:t}, Y_{1:t-1})$, the prediction of current Y after current U and earlier observations are available. Better prediction of U alone therefore does not count toward an inertia selection. Across eight validation sequences, mean J1-minus-J0 gain must exceed 0.005 nats per update and its one-sided 95% t lower bound must exceed zero. This retains the earlier screening rule. It is not a universal test with a guaranteed 5% false-positive rate.

The threshold was not changed after inspecting results. Test sequences evaluate prediction after selection. Training maximizes joint likelihood, whereas validation evaluates conditional Y prediction; the objectives are distinct by design. Accuracy, Brier score, calibration, coefficient recovery, and selection decisions are reported separately because they answer different questions (Wilson and Collins, 2019).

8. Results

8.1 Joint inference avoids spurious selections but misses real inertia under weaker measurements

Table 5. Fresh-cohort inertia selections under the unchanged rule.

Generating mechanism	Setting	Plug-in M1 selected	Joint J1 selected
No intrinsic inertia	Favorable	12/24	0/24
No intrinsic inertia	Moderate	21/24	0/24
No intrinsic inertia	Severe	4/24	0/24
Real inertia	Favorable	24/24	24/24
Real inertia	Moderate	24/24	2/24
Real inertia	Severe	5/24	0/24

The table retains primary archived results. Longer baseline refitting changes the severe no-inertia plug-in count from 4/24 to 3/24; all joint decisions stay unchanged. In the paired cohort, joint selections are 0/24 at each no-inertia setting and 24/24, 5/24, and 1/24 with real inertia. The paired cases are reanalyses, not additional independent confirmation (Dietz, 2026d).

Under the tested assumptions, joint inference avoids the identified misleading selections. This does not establish a zero future error rate. A count of 0/24 has a Wilson 95% upper bound of about 13.8%; 24/24 has a lower bound of about 86.2%. Real inertia is not reliably detected under the moderate and severe measurement conditions.

The plug-in selections need not be false positives for prediction. A persistence term can genuinely help a model that omits relevant target history. The mistake is to interpret that predictive improvement as evidence of intrinsic occupancy inertia when the generating coefficient is zero.

8.2 State estimates improve even when added inertia is not selected

Table 6. Mean fresh real-inertia recovery, evaluating fitted J1 in every case.

Setting	Target accuracy	Occupied-state accuracy	Previous M1 occupancy accuracy
Favorable	99.9%	99.9%	99.9%
Moderate	93.0%	92.3%	89.4%
Severe	75.0%	74.7%	66.7%

Table 6 evaluates fitted J1 in every case, even when validation selects J0. Once selection is taken into account, moderate real-inertia occupancy accuracy is 92.1% for the joint approach versus 89.4% for the plug-in approach. Mean conditional test gain is 0.0382 nats per update. In the fitted J1 analysis, target and occupancy Brier scores average 0.0533 and 0.0580, with ten-bin calibration errors of 0.0185 and 0.0159 (Dietz, 2026d).

The target posterior uses current U and Y as well as past observations. It is not an independent measurement from U alone. Because the model connects target and occupancy, a readout of Z can inform H. Improved recovery is conditional on that structure; it does not establish that environmental and internal causes have been separated in a real experiment.

8.3 Recovering a coefficient does not ensure model selection

Table 7. Fitted J1 persistence coefficient and incremental test gain in the fresh cohort.

Generator and setting	Mean fitted b	b RMSE	J1 minus J0 gain
No inertia, favorable	0.058	0.101	-0.00015
No inertia, moderate	0.194	0.280	-0.00014
No inertia, severe	0.584	1.018	-0.00011
Real inertia, favorable	1.482	0.129	0.05749
Real inertia, moderate	1.466	0.359	0.00369
Real inertia, severe	2.236	2.390	-0.00033

True b is zero or 1.5. Gains are conditional predictive nats per update. Fitted b is constrained nonnegative, so the no-inertia means need not be exactly zero in finite samples. RMSE measures root-mean-square estimation error across the 24 replications; it is not a confidence interval for an individual fit.

In the moderate real-inertia condition, mean fitted b is 1.466, close to the generating value of 1.5. Yet mean held-out gain from adding inertia is only 0.00369 nats per update, below the unchanged 0.005 cutoff. The Monte Carlo 95% interval for that mean gain is approximately

0.00256 to 0.00482. Recovering a coefficient and passing a predictive screening rule are different achievements. This result warrants neither a post hoc threshold change nor a conclusion that inertia is absent. A parameter-based decision would require separate calibration.

The severe real-inertia fits are much less stable: b ranges from zero to about 8.69 and mean held-out benefit is slightly negative. The parameter's plausible name does not make its value a precise measurement of a mechanism (Dietz, 2026d).

8.4 Removing target history restores misleading selections

The target-history ablation is a control that removes the target's dependence on its past while still estimating both hidden variables. It uses the same 48 paired moderate cases. In no-inertia cases, the control selects persistence in 18/24, compared with 22/24 for the plug-in model and 0/24 for the joint model retaining history. Adding persistence gives the control a mean independent-test gain of 0.02311 nats per update, with Monte Carlo 95% interval 0.02010 to 0.02612.

Adding another hidden variable is therefore not enough: its temporal structure matters. This supports Equation (15)'s explanation of regular responses to a slowly changing target without intrinsic inertia. Target history is one demonstrated confounding route, not the only possible latent explanation in brain recordings.

8.5 State recovery does not recover every diagnostic quantity

The joint study assumes the correct A/B target grammar. It does not test discovery of a missing target, establish grammar adequacy, or recover an absolute KL remainder floor from its binary mismatch probability.

The earlier study shows why these tasks differ. Under moderate measurements, its true observed-distribution floor was zero, but the packet-level estimate averaged 0.2069 nats. True occupancy excess averaged about 0.0070; its estimate averaged 0.0434 and correlated only about 0.12 with the true trajectory. Occupied-state accuracy was nevertheless 89.2%. These results concern the particular plug-in estimator, not every possible remainder estimator (Dietz, 2026c).

In that study, rotating the coordinates while retaining the old block constraints reduced moderate state accuracy from 89.2% to 48.1%. Restoring the representation restored performance. The missing-C expansion test did not detect this identification failure. A particular expansion's failure to help does not certify the model's adequacy.

8.6 Numerical stability and remaining uncertainty

Numerical warnings prompted longer refitting with an independently implemented parameter-update solver. All 222 selected-start refits converged, and no joint-model selection or occupied-state classification changed. Some coefficients nevertheless moved along nearly flat likelihood regions. Stable predictions therefore do not guarantee precise parameters or a global optimum. Appendix B.2 retains the counts, checks, and sensitivity procedure.

The conventional flat-HMM filter agreed with the joint filter to approximately 3.13×10^{-15} in probability and 2.85×10^{-14} in conditional log score on the checked trajectories. This is an equivalence control, not an advantage over conventional modeling. Complete warnings, primary results, and sensitivities remain in the research package (Dietz, 2026d).

9. From synthetic organization to conscious content

9.1 What would connect a model state to experience?

H and Z are model labels, not established conscious contents. The simulation identifies a measurement problem and a conditional remedy. It does not establish that clarity is low KL mismatch or that a brain implements the transition equation. Applying CCR to consciousness requires an independently justified mapping between its variables and experience.

Clarity particularly needs that separation. An uninformative target can have low mismatch, and a confidently wrong interpretation can have a large score gap. Clarity, accuracy, confidence, and sensory evidence are different measures. The predicted direction of a metric's relation to content must be registered for a stated task and representation, not assumed as a universal law.

Factorization cost asks how much worse fit becomes when a specified relation is excluded. For a factorized subfamily $\mathcal{Q}_F \subseteq \mathcal{Q}(K)$,

$$\Phi_{cl} = \inf_{q \in \mathcal{Q}_F} D_{KL}(p||q) - R(K | p). \quad (18)$$

The cost is nonnegative when finite, but dependence can also come from shared drivers, duplicated signals, or sensor mixing. It does not establish phenomenal unity. The excluded relation must be independently linked to the content contrast. Φ_{cl} is neither IIT's integrated information nor evidence that an arbitrary system is conscious.

9.2 A prospective perceptual-organization test

The proposed primary application studies visual organization under controlled sensory changes. A task should permit two independently identifiable perceptual organizations, A and B, with evidence favoring each varied on a randomized schedule. Bistability or ambiguous motion are candidates, not already validated apparatus choices. Feasibility work must establish the stimulus, content labels, and observation timing before confirmation.

Stimulus logs and measured delivery quality provide evidence about H, not the participant's experience. EEG or MEG features are candidate evidence about Z. Delayed percept reports assess experiential organization in separate epochs, with response mapping and motor timing controlled. Add report and no-report comparisons only where an independent physiological indicator is adequately validated for the stimulus. An arbitrary split of neural channels does not justify two separable streams.

Training data define the features, observation families, and mapping from latent identities to reports. The primary confirmatory endpoint is participant-level held-out improvement in predicting delayed percept labels from pre-report observations. The comparator includes

prespecified sensory evidence, response history, and neural features. Labels used to define an organization cannot also validate it on the same data. Prospective state estimates must exclude report epochs and future measurements.

The main comparison separates a persistent uncertain-target model from one adding occupancy persistence. A duration-sensitive or common-input alternative should be specified where feasible. A matched conventional HMM is an equivalence control, not a weaker opponent. The substantive question is whether justified closure definitions and observation assumptions support a content mapping that transfers across changes in target history and report demand.

New task-specific recovery and power analyses must determine sample size and the smallest useful predictive gain before confirmation. The synthetic cutoff of 0.005 nats is not a threshold for human research. The study needs participant-level uncertainty, prespecified exclusions, eye and motion covariates, and a held-out confirmation set. A null result must distinguish inadequate sensitivity from evidence that an effect is smaller than a declared bound. No human sample size, registration, or participant result is reported here.

9.3 Further bridge hypotheses

Table 8. Candidate bridges, distinct from completed results.

Domain	Conditional hypothesis	Required challenge
Clarity	Registered clarity ratings relate to specified occupied or posterior-expected mismatch and ambiguity	Separate clarity from accuracy, confidence, and signal quality
Multistability	A registered state/score trajectory predicts held-out switch timing	Control target history, dominance duration, adaptation, eyes, and report
Relational unity	Destroying a specified content-relevant relation increases factorization cost	Rule out common input and recording dependence
Novelty	A specified rule violation transiently disrupts organization before restabilization	Compare with rarity, physical change, and predictive-surprise baselines
Intersubjective coordination	Task-relevant structures transfer across agents during successful reference	Control shared stimulation, timing, and flexible alignment

A stabilization rate must identify its target, for example the negative slope of R_{occ} over a fixed recovery interval. It cannot switch meaning between occupied and score-optimal remainder. A pre-switch entropy rise is likewise an additional hypothesis, not something the transition rule entails.

Cogitate provides a precedent for prospective comparison of divergent predictions, not evidence for CCR (Cogitate Consortium et al., 2025). Its open fMRI resource includes 118 participants viewing suprathreshold stimuli and permits content/task generalization tests

(Khalaf et al., 2026). It does not itself supply a conscious-versus-unconscious contrast. A CCR reanalysis must therefore target a specified aspect of content organization, not generic consciousness detection.

10. Originality and relations to other programs

Constrained KL approximation, total correlation, complexity penalties, graph partitions, latent-state transitions, and structure learning are established tools. CCR's proposed contribution is how they are organized to diagnose content-bearing stability, including a demonstrated route to misinterpretation and a conditional remedy. Equivalent conventional methods can obtain the same diagnoses. The claim is methodological synthesis and interpretation, not a superior algorithm.

Information Closure Theory of Consciousness links non-trivial informational closure to conscious processes, content, and level (Chang et al., 2020). Its system-environment criterion differs from CCR's constrained-family mismatch. Arneth's resonant closure proposes a dynamically self-stabilized informational state as an account of phenomenal presence (Arneth, 2026). CCR infers presence from neither R nor its dynamics. These are differences in explanatory commitments, not empirical refutations.

Rosas et al. (2024, preprint) connect hierarchical emergence to informational and computational processes that form self-contained descriptions. Grouping histories by predictive structure is not automatically the same as partitioning simultaneous neural channels. Relating these constructions requires explicit state spaces and observation assumptions.

Active inference already distinguishes inference, parameter learning, and structure learning. Neacsu et al. (2022) show concept formation improved by model reduction. A changing grammar is therefore not by itself CCR's novelty. Comparison must match quantities and tests rather than portray predictive theories as invariably fixed in structure.

The relationship to mechanistic consciousness theories is complementary but testable. Workspace access, recurrence, and intrinsic causal integration are different proposed explanations. CCR can ask whether those processes maintain independently identified organization and whether its diagnostics add anything useful. It cannot treat every result as compatible and exempt its bridges from failure. IIT, in particular, addresses phenomenal existence and quality, not report alone (Albantakis et al., 2023).

11. Limits, reproducibility, and the interactive companion

The simulation includes controlled measurement difficulties, not realistic EEG or MEG. It assumes predefined identities, zero means, Gaussian packets, independent observations within packets, the correct A/B grammar, and conditionally separable streams. It does not test unknown mixing, correlated readout noise, unmodeled common causes, continuous targets, or cross-participant transfer. Gaussian-emission HMMs have been applied to MEG, but that precedent does not validate this observation design (Baker et al., 2014).

Identification can fail even with unlimited observations of the same kind. Correlated signal with independent noise can have the same observed law as independent signal with suitably correlated noise. Losing informative channels can likewise make distinct organizations observationally identical. These cases require further assumptions or measurements, not merely more data from the unchanged scheme.

Local protocols were recorded after debugging and before the main replications, not externally preregistered. Fresh and paired cohorts remain separate. Numerical amendments preserve primary results rather than overwrite them. The v0.2-v0.4 research packages retain code, seeds, fitted parameters, emissions, selection outcomes, and uncertainty summaries. This paper presents those results; it reports neither a new human experiment nor independent replication by other researchers.

World, Measurement, Interpretation is an explanatory companion. Replay identifies an archived case, its fitted parameters, and validation decision. The sandbox generates new trajectories with supplied parameters and compares observers. Revealing simulator truth changes the display, not what inference receives. Its mismatch probability is not KL remainder, and the animation adds no empirical evidence.

12. Conclusion

CCR asks what warrants an interpretation of stable conscious content. It requires the target, occupied organization, permitted grammar, observation map, and transition assumptions to be stated before mismatch or persistence is interpreted. This separates the organization being studied from a statistical description that merely appears stable.

The examples show why one residual cannot locate every failure, why exact one-time fit does not validate transitions, and why changing representation requires transforming the model consistently. The joint study shows how an uncertain persistent target can be mistaken for intrinsic occupancy inertia. Accounting for target history addresses that route and improves state estimates, while weaker measurements and the fixed screening rule still limit detection of real inertia.

This is a bounded methodological contribution, not a validated consciousness metric. Its empirical obligation is a content-specific mapping that survives independent measurement, task and report controls, and matched alternatives. Formal clarity, predictive success, and explanation of experience are related achievements, but none substitutes for the others.

Appendix A. Formal results and proofs

A.1 Nested-family monotonicity and factorization cost

For $Q_1 \subseteq Q_2$, every model available to the first optimization is available to the second. Therefore the infimum over Q_1 cannot be lower than the infimum over Q_2 . This proves monotonicity for a fixed p . Subtracting a finite smaller infimum from a finite larger one proves $\Phi_{cl} \geq 0$. The statement does not cover undefined infinity-minus-infinity expressions or targets changed by additional measurements.

A.2 Raw occupancy excess and score regret

For each admissible K , $R(K | p) \geq R_{\text{floor}}(p)$. Multiplying by nonnegative $\pi(K)$, summing, and using $\sum_K \pi(K) = 1$ proves Equation (4)'s excess is nonnegative. Replacing R by S_0 proves the score-regret result. The two inequalities establish different comparisons when cost changes the preferred state.

A.3 Gap robustness

Let K_1 be the unique lowest-score candidate and let ΔS be its gap to the next-lowest score. Suppose every candidate score changes by at most ϵ in absolute value. For every $K \neq K_1$,

$$S'_0(K) - S'_0(K_1) \geq S_0(K) - S_0(K_1) - 2\epsilon \geq \Delta S - 2\epsilon. \quad (A1)$$

If $\Delta S > 2\epsilon$, the original winner remains unique. This is a static robustness statement. It does not force the occupied system to be in that state or prohibit a stochastic transition.

A.4 Inertial persistence and the logistic form

At a static tie between A and B, with zero self-distance, positive cross-distance d , $\eta > 0$, and $\beta > 0$,

$$P(B | A) = \frac{1}{1 + \exp(\beta\eta d)} < \frac{1}{2}. \quad (A2)$$

For unit cross-distance, taking the log probability ratio of B to A under Equation (12) gives Equation (13). At fixed zero score input only the product $\beta\eta$ appears, proving nonidentifiability of the individual parameters in that design.

A.5 Static and stationary distributions differ

For a finite candidate set with fixed scores s_i and symmetric finite distances d_{ij} , define $w_i = \exp(-\beta s_i)$, $a_{ij} = \exp(-\beta\eta d_{ij})$, and $z_i = \sum_j a_{ij} w_j$. Then $P_{ij} = a_{ij} w_j / z_i$. The distribution

$$\mu_i = \frac{w_i z_i}{\sum_k w_k z_k} \quad (A3)$$

satisfies detailed balance, since $\mu_i P_{ij}$ is proportional to $w_i a_{ij} w_j$, symmetric in i and j . When all transitions are positive it is the unique stationary distribution. It equals the static score distribution only under additional conditions, such as constant z_i . This result explains why static-score entropy and long-run occupancy entropy cannot generally be interchanged.

A.6 Persistent-target calculation

For the stationary symmetric target chain, $\mathbb{E}[h_t h_{t-k}] = (1 - 2q)^k$. Conditional independence of the z draws given target history and $\mathbb{E}[z_t | h_t] = \alpha h_t$ give Equation (15) by iterated expectation. Symmetry implies balanced response marginals. If $r_1 = \mathbb{E}[z_t z_{t-1}]$, then $P(z_t = z_{t-1}) = (1 + r_1)/2$. The first-order logistic projection has coefficient $\log[(1 + r_1)/(1 - r_1)]$. With $a=2$ and $q=0.06$, r_1 is positive although the generating intrinsic inertia is zero. The result assumes stationary target sampling and applies after the initialization convention is no longer relevant to the compared pair.

A.7 Factorized projection and representation transport

For arbitrary normalized block densities q_B , where the expressions are finite,

$$D_{\text{KL}}(p \parallel \prod_B q_B) = D_{\text{KL}}(p \parallel \prod_B p_B) + \sum_B D_{\text{KL}}(p_B \parallel q_B). \quad (\text{A4})$$

The second term is minimized at the target marginals. This proves the total-correlation interpretation and the Gaussian specialization. Under a measurable bijection T , transporting both laws preserves the density ratio or the corresponding Radon-Nikodym expression. Transporting every admissible q therefore preserves its family infimum. No corresponding invariance is asserted for information-discarding maps with unchanged target questions.

Appendix B. Reproducible filtering and study records

B.1 Filtering from the observations available so far

Let $\rho_{t-1}(h', z')$ be the joint filter after the previous Y packet. Predict

$$\bar{\rho}_t(h, z) = \sum_{h', z'} \rho_{t-1}(h', z') P(h, z \mid h', z'). \quad (\text{B1})$$

Multiply by $f_U(U_t \mid h)$ and normalize to obtain ρ_t^U . The predictive Y density is $\sum_{h, z} \rho_t^U(h, z) f_Y(Y_t \mid z)$. Its logarithm is the scored quantity. Multiplying by current Y likelihood and normalizing gives the displayed joint filter. H and Z marginals are obtained by summation; mismatch probability is the sum over the two off-diagonal joint states. Future U or Y is never used in this recursion.

The four-variable Gaussian likelihood uses each packet's count and second-moment sufficient statistic. The full v0.4 code specifies covariance projection, regularization, bounds, initialization, convergence checks, and random-stream conventions. The retained environment record lists Python 3.13.5, NumPy 2.3.5, SciPy 1.17.0, and Numba 0.65.1. Those versions document the archived run, not a requirement that later releases reproduce bit-identical fitted values.

B.2 Numerical checks and sensitivity

Among 576 primary selected-start joint fits, two reached the 600-iteration cap. Generic logistic M-step termination warnings occurred across 92 primary cases, although no selected-fit EM likelihood decreases were recorded. An independently coded solver for the small convex M-step was checked on 1,000 weighted problems. All flagged primary cases and 19 flagged ablation cases were then rerun, both joint models, with a 2,400-iteration cap. All 222 selected-start refits converged, and no joint selection or occupied-state classification changed.

B.3 Study records

Table B1. Research materials underlying this review manuscript.

Record	Reusable content	Status
Contribution Comparison v0.2	Population examples, kernel equivalence, observed-state recovery	Completed synthetic and analytic work
Measurement Stress Test v0.3	Noisy-score baseline, observation failures, coordinate controls	Completed synthetic work with preserved correction
Joint Target and Occupied-State Test v0.4	Coupled latent model, fresh and paired results, ablation, numerical sensitivity	Completed synthetic work, not externally preregistered
Shared specification and module v0.6	Shared definitions, replay/sandbox scope, software provenance	Drafting and explanatory implementation

Files are retained in the author-held project research packages; they are not embedded in this DOCX. A permanent public repository or archival identifier has not been established by the manuscript. No unavailable archive is cited as though it were already public. Replicate JSONs, code hashes, protocol records, and amendments remain the primary numerical record; the prose tables are selected summaries.

Author-review and availability note

Version 2.1 is an author-review manuscript. It revises the exposition of Version 2.0 while retaining its formal results, synthetic data, and prospective human protocol. The companion reference is updated to CCC 5.4. AI assistance was used in editorial revision and document preparation, following the earlier AI-assisted drafting and computational development. Final author verification and independent review remain necessary before external submission.

References

- Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., et al. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology*, 19(10), e1011465. doi:10.1371/journal.pcbi.1011465.
- Arneth, B. (2026). Resonant closure: consciousness as a dynamically self-stabilized informational state. *Frontiers in Human Neuroscience*, 20, 1742084. doi:10.3389/fnhum.2026.1742084.
- Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Probert Smith, P. J., and Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *eLife*, 3, e01867. doi:10.7554/eLife.01867.
- Bassett, D. S., and Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, 20, 353-364. doi:10.1038/nn.4502.
- Bengio, Y., and Frasconi, P. (1994). An Input Output HMM Architecture. *Advances in Neural Information Processing Systems 7*, 427-434. Conference proceedings.

- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-219.
- Chang, A. Y. C., Biehl, M., Yu, Y., and Kanai, R. (2020). Information Closure Theory of Consciousness. *Frontiers in Psychology*, 11, 1504. doi:10.3389/fpsyg.2020.01504.
- Cogitate Consortium et al. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature*, 642, 133-142. doi:10.1038/s41586-025-08888-1.
- Csiszar, I. (1975). I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1), 146-158. doi:10.1214/aop/1176996454.
- Dehaene, S., and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200-227. doi:10.1016/j.neuron.2011.03.018.
- Dietz, C. F. (2026a). *Consciousness Closure and the Cosmos*. Version 5.4. Companion author-review manuscript, September 2026.
- Dietz, C. F. (2026b). *CCC/CCR Contribution Comparison*. Research brief and reproducible research package v0.2. Project working materials, 7 September 2026.
- Dietz, C. F. (2026c). *CCC/CCR Measurement Stress Test*. Research report and reproducible research package v0.3. Project working materials, 7 September 2026.
- Dietz, C. F. (2026d). *CCC/CCR Joint Target and Occupied-State Test*. Research report and reproducible research package v0.4. Project working materials, 7 September 2026.
- Dietz, C. F. (2026e). *CCC/CCR Evidence and Originality Audit*. Working register and reproducible research package v0.1. Project working materials, 7 September 2026.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, 5(2A), 1020-1056. doi:10.1214/10-AOAS395.
- Ghahramani, Z., and Jordan, M. I. (1997). Factorial Hidden Markov Models. *Machine Learning*, 29, 245-273. doi:10.1023/A:1007425814087.
- Hindriks, R., Adhikari, M. H., Murayama, Y., Ganzetti, M., Mantini, D., Logothetis, N. K., and Deco, G. (2016). Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? *NeuroImage*, 127, 242-256. doi:10.1016/j.neuroimage.2015.11.055. Corrigendum: *NeuroImage*, 132, 115. doi:10.1016/j.neuroimage.2016.02.007.
- Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T. I., and Logothetis, N. K. (2022). Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports. *Nature Communications*, 13, 1535. doi:10.1038/s41467-022-28897-2.
- Khalaf, A., Richter, D., Vidal, Y., Gorska-Klimowska, U., Hirschhorn, R., et al. (2026). An open-access multi-site fMRI dataset for investigating conscious visual perception. *Scientific Data*, 13, 779. doi:10.1038/s41597-026-07377-y.
- Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86. doi:10.1214/aoms/1177729694.
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10(11), 494-501. doi:10.1016/j.tics.2006.09.001.
- Lawson, H. (2001). *Closure: A Story of Everything*. Routledge.
- Montevil, M., and Mossio, M. (2015). Biological organisation as closure of constraints. *Journal of Theoretical Biology*, 372, 179-191. doi:10.1016/j.jtbi.2015.02.029.
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980), 876-878. doi:10.1126/science.1184819.

- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435-450.
doi:10.2307/2183914.
- Neacsu, V., Mirza, M. B., Adams, R. A., and Friston, K. J. (2022). Structure learning enhances concept formation in synthetic Active Inference agents. *PLOS ONE*, 17(11), e0277199.
doi:10.1371/journal.pone.0277199.
- Pitts, M. A., Metzler, S., and Hillyard, S. A. (2014). Isolating neural correlates of conscious perception from neural correlates of reporting one's perception. *Frontiers in Psychology*, 5, 1078.
doi:10.3389/fpsyg.2014.01078.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5), 465-471.
doi:10.1016/0005-1098(78)90005-5.
- Rosas, F. E., Geiger, B. C., Luppi, A. I., Seth, A. K., Polani, D., Gastpar, M., and Mediano, P. A. M. (2024). *Software in the natural world: A computational approach to hierarchical emergence*. arXiv:2402.09090v2. Preprint, version of 5 June 2024. doi:10.48550/arXiv.2402.09090.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
doi:10.1214/aos/1176344136.
- Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4(1), 66-82. doi:10.1147/rd.41.0066.
- Wilson, R. C., and Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, 8, e49547. doi:10.7554/eLife.49547.