MCCI: Unified Theory of Mental Holes–Causality–Choice Architecture

(With Definitions–Criteria–Theorems–Proofs–Verification Protocols, Compatible with WSIG / EBOC / RCA–CID)

Auric (S-series / EBOC)

Version: v1.7 (2025-11-05, Asia/Singapore)

Keywords: Mental holes; Causal diagrams (SCM); Choice architecture (default/framing/order); Bias–noise decomposition; Loss aversion; Reference point; CATE; I-projection (KL/Bregman); WSIG; EBOC; RCA–CID

MSC: 62Cxx; 62Pxx; 68Txx; 91Bxx; 94Axx

Abstract

We construct a theory of "mental holes" verifiable under the triple norm of probability–utility–causality: given a rational baseline strategy and an embedding of observable architecture variables, we define the total deviation functional and its four-dimensional decomposition (bias, noise, causal mismatch, architecture sensitivity), provide identification criteria via backdoor/frontdoor/instrumental variables/discontinuity/difference-in-differences, and specify minimal experimental designs. Under I-projection and Bregman geometry, we prove the "Pythagoras–decoupling" structure and derive realizable estimation–audit pipelines (DQC). In the WSIG dictionary, the I-projection of the rational constraint family is viewed as the "readout norm", and deviations are written as KL/Bregman distances; in EBOC, the pipeline is implemented as "window selection leaf" rules; in RCA–CID, reversible logs guarantee intervention replayability and external audit. We also provide an in-model determination criterion for "loss aversion–love" via the indicator $L = \eta(\lambda - 1)$ (concern weight $\times$ fracture coefficient). Core proofs follow Csiszár's I-projection and Bregman–Pythagoras, Pearl's causal criteria, and modern estimation theory.

1 Notation & Axioms / Conventions (WSIG–EBOC–RCA Unity)

A1 (Measure–Strategy–Readout): The observation triple $(\mathcal{H}, w, \mathcal{D})$ induces windowed readouts; all strategies and distributions on the standard simplex are metrized by Bregman divergence $D_\phi(\cdot\|\cdot)$ and KL; rational baseline given by I-projection on constraint families [**?**].

A2 (Calibration Identity, WSIG card): Under the unified calibration of scattering–information geometry, we adopt the mother scale $\varphi'(E)/\pi = \rho_{\mathrm{rel}}(E) = (2\pi)^{-1} \operatorname{tr} \mathsf{Q}(E)$, where $\mathsf{Q} := -\mathrm{i}\, S^\dagger \partial_E S$ is the Wigner–Smith group delay matrix; as the measure coordinate connecting to this system [**?**].

A3 (Finite-order NPE discipline): All discrete–continuous transformations and windowed integrations uniformly adopt "finite-order Euler–Maclaurin + Poisson" three-term error closure, asserting non-increasing singularity and pole = primary scale.

A4 (RCA–CID reversibility): Implementation and audit are uniformly mapped to Bennett reversible computation and Zeckendorf-encoded logs; guaranteeing reversible replay of interventions and estimation versions [**?**].

2 Model and Baseline Norm

Variables: Context X, action $A \in \mathcal{A}$, outcome Y, unobserved disturbance U; **architecture variables** $C = (F, D, S)$ for presentation framing, default selection, presentation order.

SCM: Directed acyclic graph G and structural equations $V_i := f_i(\mathrm{Pa}(V_i), U_i)$.

Rational baseline: Under identified intervention distribution $P(Y \mid do(A = a), X)$ and utility u, the Bayes–decision optimal strategy

$$\pi^\star(\cdot \mid x) \in \arg\max_{\pi} \mathbb{E}\left[u(Y) \mid do(A \sim \pi(\cdot \mid x)), X = x\right].$$

Actual strategy: $\pi(\cdot \mid x, c)$ may explicitly depend on c.

Divergence: Take KL or general Bregman divergence D_ϕ.

3 Definitions: Deviation Functional and Four-Dimensional Decomposition of Mental Holes

Definition 3.1 (Total Deviation–Repeated Review Unification)**.** *For each context $X = x$, fix a* ***baseline presentation*** *c_0; let the r-th review's strategy be $\pi^{(r)}(\cdot \mid x, c_0)$. Define*

$$\mathcal{L} := \mathbb{E}_X \, \mathbb{E}_r\big[D_\phi\big(\pi^\star(\cdot \mid X) \| \pi^{(r)}(\cdot \mid X, c_0)\big)\big].$$

(Architecture sensitivity is separately measured by AS *and its regularization term $\mathcal{R}_{\mathrm{AS}}$; see Theorem* **??***.)*

Definition 3.2 (Same-Case Repetition and Four Components)**.** *For the same case x, repeat reviews $A^{(r)} \sim \pi^{(r)}(\cdot \mid x, c)$. Here $\pi^{(r)}(\cdot \mid x, c)$ denotes the action distribution of the r-th review (or reviewer); its Bregman centroid*

$$\bar{\pi}_\phi(\cdot \mid x, c) := (\nabla\phi)^{-1}\big(\mathbb{E}_r[\nabla\phi(\pi^{(r)}(\cdot \mid x, c))]\big).$$

Define

$$\begin{aligned}
\mathrm{Bias}(x) &:= D_\phi\big(\pi^\star(\cdot \mid x) \| \bar{\pi}_\phi(\cdot \mid x, c)\big),\\
\mathrm{Noise}(x) &:= \mathbb{E}_r\big[D_\phi\big(\bar{\pi}_\phi(\cdot \mid x, c) \| \pi^{(r)}(\cdot \mid x, c)\big)\big],\\
\mathrm{CM}(x) &:= \Big(\mathbb{E}[u(Y) \mid A \sim \bar{\pi}_\phi(\cdot \mid x, c), X = x]\\
&\qquad - \mathbb{E}[u(Y) \mid do(A \sim \bar{\pi}_\phi(\cdot \mid x, c)), X = x]\Big)^2 \geq 0,\\
\mathrm{AS}(x) &:= \sup_{c, c'} D_\phi\big(\pi(\cdot \mid x, c) \| \pi(\cdot \mid x, c')\big).
\end{aligned}$$

Definition 3.3 (Strength Indicator)**.** *Given weights $\omega \succ 0$, define*

$$\mathrm{Defect} := \mathbb{E}_X\Big[\omega_b \,\mathrm{Bias}(X) + \omega_n \,\mathrm{Noise}(X) + \omega_c \,\mathrm{CM}(X) + \omega_a \,\mathrm{AS}(X)\Big].$$

Note: The four terms here correspond one-to-one with $\mathcal{B}, \mathcal{N}, \mathcal{C}, \mathcal{R}_{\mathrm{AS}}$ in Section **??***, where $\mathcal{C} = \mathbb{E}_X[\mathrm{CM}(X)]$ and $\mathcal{R}_{\mathrm{AS}}$ is the penalty functional for* AS*.*

4 Causal Embedding and Identification Criteria

Architecture embedding: Incorporate C as parent or co-parent of A into G: $C \to A \to Y$; allow C to alter information presentation and observation channels but not the structural equations of potential outcomes $Y(a)$.

Backdoor criterion: If there exists $Z \subset X$ blocking all backdoor paths from A to Y, then $P(y \mid do(a)) = \sum_z P(y \mid a, z)P(z)$ [**?**].

Frontdoor/IV/RD/DiD: For unobserved confounding, use frontdoor variables, qualified instruments (relevance, exclusion, monotonicity), regression discontinuity, and modern multi-period DiD (including staggered treatment timing and continuous intensity) respectively [**?**, **?**].

5 Three Core Theorems and Proofs

Theorem 5.1 (Bregman–Pythagoras Dual Decomposition + Regularization)**.** *For each x, taking expectation over r yields*

$$\mathbb{E}_r\Big[D_\phi\big(\pi^\star \| \pi^{(r)}\big)\Big] = D_\phi\big(\pi^\star \| \bar{\pi}_\phi\big) + \mathbb{E}_r\Big[D_\phi\big(\bar{\pi}_\phi \| \pi^{(r)}\big)\Big].$$

Taking expectation over X, by Definition **??** *we obtain*

$$\mathcal{L} = \underbrace{\mathbb{E}_X\big[D_\phi(\pi^\star \| \bar{\pi}_\phi)\big]}_{\mathcal{B}} + \underbrace{\mathbb{E}_X\big[\mathbb{E}_r D_\phi(\bar{\pi}_\phi \| \pi^{(r)})\big]}_{\mathcal{N}}.$$

Introducing regularization to penalize causal mismatch and architecture sensitivity, define

$$\mathcal{L}_{\text{aug}} := \mathcal{L} + \underbrace{\mathbb{E}_X[\text{CM}(X)]}_{\mathcal{C}} + \underbrace{\Psi_{\text{AS}}}_{\mathcal{R}_{\text{AS}}} \quad \Rightarrow \quad \mathcal{L}_{\text{aug}} = \mathcal{B} + \mathcal{N} + \mathcal{C} + \mathcal{R}_{\text{AS}},$$

where $\mathcal{C}, \mathcal{R}_{\text{AS}} \geq 0$.

Proof. The Bregman three-point identity $D_\phi(x_1 \| x_3) = D_\phi(x_1 \| x_2) + D_\phi(x_2 \| x_3) + \langle x_1 - x_2, \nabla\phi(x_3) - \nabla\phi(x_2)\rangle$, taking $x_1 = \pi^\star, x_2 = \bar{\pi}_\phi, x_3 = \pi^{(r)}$ and conditional expectation over r, using $\bar{\pi}_\phi = (\nabla\phi)^{-1}\mathbb{E}[\nabla\phi(\pi^{(r)})]$ to make the cross term 0 (Bregman centroid first-order condition), yields the first identity and baseline equality; $\text{CM}(X)$ is defined as a nonnegative squared difference by Definition **??**, Ψ_{AS} is the penalty functional for AS; incorporating both as regularization terms gives the augmented $\mathcal{L}_{\text{aug}}$ [**?**]. ☐

Theorem 5.2 (Architecture Equivalence and Architecture Effect)**.** *If two presentations c, c' only affect information channels without altering the structure of $Y(a)$, then*

$$\text{AS}(x) = 0 \iff \pi(\cdot \mid x, c) = \pi(\cdot \mid x, c') \textit{ almost surely.}$$

If $\text{AS}(x) > 0$, there exists an ***architecture effect*** *induced by pure presentation difference $P(a \mid x, c) \neq P(a \mid x, c')$.*

Proof. By positive definiteness of divergence and the definition, immediate. ☐

Theorem 5.3 (In-Model Determination of "Loss Aversion–Love")**.** *Let $s \in \{0, 1\}$, reference point $s^* = 1$, other's welfare weight $\eta \geq 0$, fracture loss coefficient $\lambda > 1$,*

$$U(x, y, s) = u(x) + \eta\, u(y) + v(s - s^*), \qquad v(z) = \begin{cases} \alpha\, z, & z \geq 0, \\ -\lambda\, \beta(-z), & z < 0. \end{cases}$$

where $\beta(\cdot) > 0,\ \beta(0) = 0$. Operationalize "love" as: WTP to reduce separation probability from $\varepsilon \downarrow 0$ to 0 exceeds the baseline implied solely by risk aversion of u. Then under the premise $\lambda > 1$,

$$\textit{Love} \iff \eta > 0, \qquad L := \eta(\lambda - 1) > 0.$$

Proof. At first-order approximation,

$$\text{WTP} \;\sim\; \varepsilon\Big(\eta \cdot \Delta u + (\lambda - 1) \cdot \beta(1)\Big),$$

where Δu represents the marginal difference in other's welfare between $s = 1$ and $s = 0$; if $\eta = 0$, this term vanishes; if $\lambda = 1$, there is no loss aversion correction for separation. Both being positive yields positive WTP excess. ☐

6 Identification and Estimation (DQC: Document–Counter–Causalize–Audit)

D1 Document: Case file contains $(X, C, \mathcal{A}, \text{objective}, \text{constraints})$.

D2 Counter-framing: Apply two or more C to the same case (gain/loss framing, default switching, order shuffling), compute

$$\widehat{\mathrm{AS}}(x) = \max_{c,c'} D_\phi\big(\hat{\pi}(\cdot \mid x, c), \hat{\pi}(\cdot \mid x, c')\big),$$

flag as "architecture sensitive" if above threshold.

D3 Causalization: Draw DAG and identify via backdoor/frontdoor/IV/RD/DiD criteria; prioritize small-scale randomization for randomizable cases. Estimate ATE $= \mathbb{E}[Y(1) - Y(0)]$, $\mathrm{CATE}(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x]$. For observational data, use IPW/DR/TMLE and causal forests; perform Γ-sensitivity analysis for unobserved confounding [?].

D4 Audit (Noise audit): Same-case multi-evaluation estimates Noise and aggregates

$$\widehat{\mathrm{Defect}} = \omega_b \widehat{\mathcal{B}} + \omega_n \widehat{\mathcal{N}} + \omega_c \widehat{\mathcal{C}} + \omega_a \widehat{\mathrm{AS}}.$$

Distinguish "level noise/pattern noise/occasion noise" in reports and provide "decision hygiene" protocols (independent judgment, aggregation, multi-source evidence) [?].

7 Identification Criteria and Minimal Experimental Design (Quick Reference)

Backdoor: Select Z blocking all paths with arrows into A, use $\sum_z P(y \mid a, z)P(z)$ [?].

Frontdoor: When complete mediator M exists and $A \to M$ has no backdoor, $M \to Y$ is backdoor-adjustable, $P(y \mid do(a))$ is identifiable [?].

Instrumental Variables (IV): Z relevant to A, independent of $Y(a)$, affects Y only through A; under monotonicity identifies LATE [?].

Regression Discontinuity (RD): Continuity assumption at threshold guarantees local average causal effect identification [?].

Multi-period DiD: Under staggered treatment and heterogeneous effects, use Callaway–Sant'Anna / Sun–Abraham families and extensions to continuous treatment intensity [?].

8 Estimators and Error Discipline (Non-Asymptotic Implementation)

IPW / DR: Utilize double robustness of propensity score and outcome regression; report small-sample corrections and trimming robustness [?].

TMLE: Two-step substitution estimation respecting efficiency influence function of target functional, easy to integrate with ML; provide influence function standard errors [?].

Causal forests / Generalized random forests: Estimate CATE and uncertainty, handle cluster errors [?].

Sensitivity analysis: Rosenbaum Γ bounds, marginal sensitivity model and its sharper variants [?].

NPE error budget: For all discrete–continuous transformations, report three parts: aliasing, boundary layer (Bernoulli), and tail with total bounds.

9 Isomorphic Connection with WSIG / EBOC / RCA–CID

WSIG (I-projection = Born readout): The I-projection $q^\star = \arg\min_{q\in\mathcal{Q}} \mathrm{KL}(p\|q)$ on rational constraint family $\mathcal{Q}$ is the "norm readout"; total deviation $\mathcal{L} = \mathrm{KL}(q^\star\|p_\pi)$ is the readout–strategy relative deviation; Bregman–Pythagoras gives the additive "bias + noise" structure [?].

EBOC (static block): Case files and randomized designs are window selection rules on static block measures, not altering global measure; time is viewed as leaf reading of blocks, with order induced by selection rules.

RCA–CID (reversible log): Embed DQC pipeline in reversible cellular automata; all intervention–estimation versions recorded in CID logs encoded in Zeckendorf normal form, and Bennett reversible embedding guarantees replayability and external audit [?].

Calibration alignment: In scenarios requiring convergence with energy spectrum calibration, cite $\varphi'/\pi = \rho_{\mathrm{rel}} = (2\pi)^{-1}\,\mathrm{tr}\,\mathsf{Q}$ as universal coordinate; group delay–bandwidth resource constraints become global budget for DQC [?].

10 Experimental Blueprint and Reproducibility Checklist

A/B (default effect): Randomize $D \in \{\text{opt-in}, \text{opt-out}\}$; test ΔATE and $\widehat{\mathrm{AS}}$.

Dual-framing review: Same notification presented in gain/loss versions; estimate CATE with TMLE [?].

Noise audit: Same-case multi-evaluation; distinguish level/occasion/pattern noise and report post-reduction magnitude and stability [?].

"Love" indicator: Construct insurance-type choice with small-probability separation on voluntary sample; estimate $\widehat{L} = \hat{\eta}(\hat{\lambda} - 1)$ and link with satisfaction/reciprocity secondary endpoints.

Governance and fairness: Report CATE, $\widehat{\mathrm{AS}}$ for key subgroups; set "architecture fairness" thresholds and notification norms.

11 Further Properties and Corollaries

Corollary 11.1 (Backdoor Adjustment ⇒ Causal Mismatch Term Vanishes)**.** *If there exists Z satisfying the backdoor criterion, and when computing* CM *full adjustment is performed on Z, then $\mathcal{C} = 0$ [?].*

Corollary 11.2 (KL Special Case Centroid)**.** *When $D_\phi = \mathrm{KL}$ and the first argument is on the simplex, $\bar{\pi}_\phi$ is a geometric-mean-type centroid, ensuring the cross term in Theorem **??** vanishes [?].*

Corollary 11.3 (Sufficiency of Decision Hygiene)**.** *Independent judgment and de-echo-chamber aggregation on the Bregman platform are equivalent to minimizing $\mathbb{E}_r[D_\phi(\bar{\pi}_\phi\|\pi^{(r)})]$, thus directly reducing $\mathcal{N}$ [?].*

Corollary 11.4 (Group Delay Budget)**.** *In systems calibrated with $\mathrm{tr}\,\mathsf{Q}$, total complexity of windowed evaluation is constrained by group delay–bandwidth product upper bound, serving as resource budget for DQC [?].*

12 Proof Details (Selected)

(I) Bregman–Pythagoras: Banerjee et al.'s general treatment of Bregman three-point identity and clustering centroid, combined with Csiszár I-projection geometry, gives the first-order condition $\bar{\pi}_\phi = (\nabla\phi)^{-1}\mathbb{E}[\nabla\phi(\pi^{(r)})]$, hence cross term is 0 [?].

(II) Causal identification: Pearl's backdoor/frontdoor; Angrist–Imbens–Rubin IV and LATE; Hahn–Todd–van der Klaauw RD; Callaway–Sant'Anna (and subsequent extensions) multi-period and continuous treatment DiD [?].

(III) Estimation theory: Bang–Robins DR; van der Laan–Rubin TMLE; Athey–Wager causal and generalized random forests; Rosenbaum and recent sensitivity reviews [?].

(IV) WSIG calibration: Wigner–Smith group delay and Birman–Kreĭn formula provide equivalent coordinates of calibration–phase–spectrum, used as measure coordinate converging with this theory [?].

(V) RCA–CID reversibility: Bennett's logical reversibility and Zeckendorf theorem guarantee reversible replay and unique factorization of logs, enabling external audit of intervention–estimation versions [?].

13 Implementation Blueprint (Engineering Minimal Set)

1. **Diagramming and criteria:** Each online decision flow first draws DAG and marks backdoor sets/available instruments/possible thresholds and temporal staggering.

2. **Online DQC:** Case file template + dual-framing questionnaire + small-scale randomization; automated IPW/DR/TMLE/causal forests; accompanied by Rosenbaum Γ report [?].

3. **Audit and governance:** Report CATE, $\widehat{\mathrm{AS}}$ and $\widehat{\mathrm{Defect}}$ (including $\widehat{\mathcal{B}}, \widehat{\mathcal{N}}, \widehat{\mathcal{C}}, \widehat{\mathrm{AS}}$) for key subgroups; set "architecture fairness" thresholds and review frequency.

4. **RCA–CID:** Use Zeckendorf-log to carry versions; declare reversible replay interface and audit API.

One-Sentence Summary

"Mental holes" are decomposable deviations of strategy relative to rational baseline; via causal criteria and I-projection, they are operationalized into measurable indicators; DQC converts doubt into institutionalized improvement and guarantees auditability and portability within the unified language of WSIG / EBOC / RCA–CID.

References

[1] Csiszár, I. *I-Divergence Geometry of Probability Distributions and Minimization Problems.* Ann. Probab. 1975. `https://projecteuclid.org/journals/annals-of-probability/volume-3/issue-1`

[2] Smith, F.T. *Lifetime Matrix in Collision Theory.* Phys. Rev. 1960. `https://link.aps.org/doi/10.1103/PhysRev.118.349`

[3] Bennett, C.H. *Logical Reversibility of Computation.* IBM J. 1973. `https://users.cs.duke.edu/~reif/courses/complectures/AltModelsComp/Bennett/LogRevComp.pdf`

[4] Pearl, J. *Causal diagrams for empirical research.* Biometrika 1995. `https://fitelson.org/woodward/pearl_95.pdf`

[5] *The Front-door Criterion in the Potential Outcome Framework.* arXiv 2024. `https://www.arxiv.org/pdf/2412.10600`

[6] Banerjee, A., Merugu, S., Dhillon, I., Ghosh, J. *Clustering with Bregman Divergences.* JMLR 2005. https://www.jmlr.org/papers/volume6/banerjee05b/banerjee05b.pdf

[7] Bang, H., Robins, J. *Doubly Robust Estimation in Missing Data and Causal Inference.* 2005. https://www.math.mcgill.ca/dstephens/SISCR2018/Articles/bang_robins_2005.pdf

[8] Kahneman, D., Sibony, O., Sunstein, C. *Noise: A Flaw in Human Judgment.* 2021. https://en.wikipedia.org/wiki/Noise:_A_Flaw_in_Human_Judgment

[9] Angrist, J., Imbens, G., Rubin, D. *Identification of Causal Effects Using Instrumental Variables.* JASA 1996. https://www.math.mcgill.ca/dstephens/AngristIV1996-JASA-Combined.pdf

[10] *Regression Discontinuity Designs: A Guide to Practice.* NBER Working Paper. https://www.nber.org/system/files/working_papers/w13039/w13039.pdf

[11] Callaway, B., Sant'Anna, P. *Difference-in-Differences with multiple time periods.* 2021. https://file-lianxh.oss-cn-shenzhen.aliyuncs.com/Refs/2025-08-Yang/Callaway_2021_Difference-in-Differences_with_multiple_time_periods.pdf

[12] *Targeted Maximum Likelihood Learning.* 2006. https://www.degruyterbrill.com/document/doi/10.2202/1557-4679.1043/html

[13] Athey, S., Tibshirani, J., Wager, S. *Generalized random forests.* Ann. Stat. 2019. https://projecteuclid.org/journals/annals-of-statistics/volume-47/issue-2/Generalized-random-forests/10.1214/18-AOS1709.full

[14] Rosenbaum, P. *An Introduction to Sensitivity Analysis for Unobserved Confounding.* PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC3800481/

[15] *Daniel Kahneman Says Noise Is Wrecking Your Judgment.* Barron's 2021. https://www.barrons.com/articles/economist-daniel-kahneman-says-noise-is-wrecking-your-judgment-51622228892

[16] Pearl, J. *Causality.* 2009. https://archive.illc.uva.nl/cil/uploaded_files/inlineitem/Pearl_2009_Causality.pdf

[17] *An R Package for Targeted Maximum Likelihood Estimation.* J. Stat. Softw. 2011. https://www.jstatsoft.org/article/view/v051i13/653