VALUE-TWISTED SEQUENTIAL MONTE CARLO FOR TASK-AWARE MISSING-MODALITY CONNECTOME COMPLETION

Anonymous authors
Paper under double-blind review

ABSTRACT

Completing a missing modality in the field of medical imaging data could create a Prior–Utility Dilemma: samples that faithfully follow a conditional generative prior may not help prediction, while samples aggressively optimized for a classifier may no longer represent that prior. Our goal is to make diffusion-generated connectomes more useful downstream while keeping their sampling distribution explicit and auditable. We introduce Value-Twisted Sequential Monte Carlo (VT-SMC), a task-aware inference procedure for missing-connectome completion. To resolve this dilemma, VT-SMC formulates task-aware connectome completion as inference under an explicit utility-tilted generative distribution. It combines a conditional diffusion prior with frozen uncertainty-aware task feedback, then uses corrected particle inference to favor diagnostically useful connectomes without losing distributional accountability. This unified formulation balances anatomical plausibility, downstream utility, and subject-specific uncertainty, while admitting a consistent SMC estimator and a gradient-free variant. Sufficient experiments show that the method manages, rather than eliminates the fidelity–utility trade-off.

1 INTRODUCTION

Multimodal neuroimaging relates brain function, anatomy, and white-matter organization to disease: fMRI describes functional coupling, sMRI anatomy, and dMRI structural connectivity. Yet fewer than 20% of subjects in our ABIDE II cohort have usable dMRI (Di Martino et al., 2017). Requiring complete modalities would discard most usable data. Existing responses fall into three paradigms. Modality avoidance uses available inputs, modality dropout, or distillation, but forgoes an explicit missing connectome (Neverova et al., 2016; Hinton et al., 2015). Generative completion predicts that connectome through regression, adversarial learning, or diffusion (Zhang et al., 2022; Tang et al., 2023; Ho et al., 2020; Rajadhyaksha & Rekik, 2023); it restores the expected input but usually optimizes reconstruction rather than diagnostic utility. Task-guided sampling instead uses classifier gradients or particle reweighting (Dhariwal & Nichol, 2021; Wu et al., 2023).

However, each step toward utility creates a new failure mode. Unguided completion follows the learned conditional prior, yet may allocate samples away from a finite-sample decision boundary. Direct gradient guidance can improve classifier score, yet its shifted transitions need not sample either the base prior or an explicitly defined alternative. Pure Feynman–Kac reweighting and resampling retain a declared target, but cannot actively move prior proposals into a narrow high-value region; under strong tilts, overlap shrinks and weights degenerate. Entropy-aware diffusion guidance has adapted the strength of a single noisy classifier (Zheng et al., 2022a), and ensembles are established uncertainty estimators (Ovadia et al., 2019); however, neither alone defines a cross-seed, label-free entropy target with proposal correction. These limitations form the Prior–Utility Dilemma: preserving the generative law can waste task-relevant probability, while stronger task guide can make that law ambiguous. Moreover, generated dMRI cannot add independent label information beyond its observed inputs, so gains must be attributed to representation and regularization rather than recovered biology.

Figure 1 contrasts three sampling laws: sensible but task-agnostic, task-seeking but unspecified, and task-aware yet auditable. We realize the third with Value-Twisted Sequential Monte Carlo (VT-SMC). A label-free diffusion model defines plausibility; an independent pilot freezes subject

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

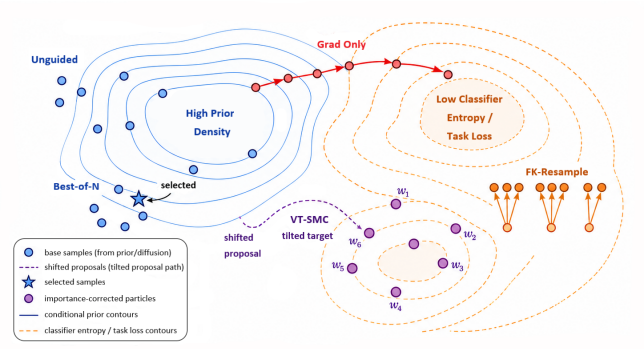

Figure 1: The Prior–Utility Dilemma. Unguided sampling (left) preserves the conditional prior but ignores task preference. Heuristic guidance (middle) pursues utility without correcting its changed law. VT-SMC (right) declares a value-tilted target, uses gradients as proposals, and restores accountability through importance correction and resampling.

ambiguity; an ensemble converts cross-seed predictive entropy into a label-free value; and corrected SMC combines shifted proposals, density-ratio weights, and adaptive resampling. A dimension-independent budget limits shifts, while fresh evaluators prevent the guiding model from judging its own generations. Our contributions can be concluded as follows. (1)we identify the Prior–Utility Dilemma and formulate task-aware completion as inference in an explicit value-tilted distribution, without interpreting synthesis as new subject-specific information. (2)we introduce a label-free ensemble-entropy potential, pilot-frozen confidence, and proposal-corrected SMC for symmetric connectomes. (3)we derive the exact Gaussian correction, a dimension-independent proposal-KL budget, and consistency of the resulting particle estimator. (4)we separate utility from fidelity through closed-form target recovery, independent evaluators, two cohorts, particle diagnostics, and controlled ablations.

2 Related Work

Missing modalities and brain-network prediction. Missing-modality learning spans synthesis, shared representations, modality dropout, and distillation (Neverova et al., 2016; Havaei et al., 2016; Hinton et al., 2015; Azad et al., 2022; Wu et al., 2024); conditional GANs and diffusion models support input-space completion (Chartsias et al., 2018; Dorjsembe et al., 2023). Brain-network predictors range from single-input graph models to multimodal fusion. They motivate the diagnostic baselines detailed in Appendix C.8, but do not account for how task feedback changes a conditional generative law. We therefore report prediction and dMRI fidelity separately.

Diffusion guidance. Classifier and classifier-free guidance steer reverse diffusion with auxiliary gradients or conditional-score differences (Dhariwal & Nichol, 2021; Ho & Salimans, 2022). An unweighted mean shift generally samples from neither the base model nor a declared tilted target. In contrast, our ensemble potential defines the target reweighting and its gradient only proposes efficient moves, so the two roles can be ablated independently.

Sequential Monte Carlo and twisting. SMC approximates sequential targets with weighted particles and resampling (Del Moral et al., 2006; Naesseth et al., 2019); twisted diffusion samplers retain importance correction while steering trajectories (Wu et al., 2023). We add a label-free ensemble-entropy potential, pilot-frozen confidence, and symmetric-subspace densities. Because the potential is learned in the fold receiving generated training data, fold isolation and the causal separation between pilot uncertainty and guided particles are part of the method. ESS, ancestry, independent classifiers, and graph-distribution metrics test collapse and evaluator-specific exploitation (Kawahara et al., 2017; Gretton et al., 2012).

Reliability and structural fidelity. Predictive confidence requires calibration or selective-risk evaluation (Guo et al., 2017; Geifman & El-Yaniv, 2017; Ovadia et al., 2019), while edgewise similarity can miss graph-topological changes (Rubinov & Sporns, 2010). The appendix therefore retains calibration, graph statistics, and independent-evaluator audits.

3 METHOD

The goal of VT-SMC is not to claim recovery of unobserved biological information, but to resolve the Prior–Utility Dilemma for a finite-sample diagnostic model. It treats diffusion as a distribution over sensible missing connectomes, classifier agreement as a label-free preference over that distribution, and sequential Monte Carlo as the mechanism that realizes the preference without hiding the change of sampling law. The method has four parts. Section 3.1 learns a label-free conditional diffusion prior in the symmetric connectome space. Section 3.2 estimates uncertainty from an independent pilot and injects it into the provisional and final classifiers. Section 3.3 converts ensemble agreement into an explicit value tilt and performs proposal-corrected particle inference. Section 3.4 calibrates the amount of guidance, states the guarantees, and closes with the end-to-end workflow. Fold construction and leakage control are specified in Appendix C.1.

3.1 LABEL-FREE DIFFUSION IN THE SYMMETRIC SUBSPACE

The first component provides the anatomical prior that later guidance is allowed to redistribute. Because structural connectome is symmetric, running diffusion on all n^2 matrix entries would duplicate every off-diagonal degree of freedom and make Gaussian densities inconsistent with the actual state space. We instead represent X^d by its upper triangle $v = \mathrm{vech}(X^d) \in \mathbb{R}^K$, where $K = n(n+1)/2$. Noise is added only to these free coordinates and the sampled vector is mirrored once, so every intermediate and terminal state is exactly symmetric. We use the standard variance-preserving forward process

$$q(v_t \mid v_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t} v_0, (1 - \bar{\alpha}_t) I_K). \tag{1}$$

At a randomly sampled step t, a Transformer denoiser receives the noisy connectome tokens, a time embedding, and tokens from observed modalities (Peebles & Xie, 2023). It predicts the injected noise through

$$\mathcal{L}_{\mathrm{diff}} = \mathbb{E}_{t, v_0, \epsilon} \left[\|\epsilon - \epsilon_\theta(v_t, t; X^f, x^s)\|_2^2 \right]. \tag{2}$$

The generator is label-free: diagnosis is never embedded in the denoiser or used to select its checkpoint. On ABIDE II it conditions on fMRI and sMRI; on the second cohort the absent sMRI branch is removed rather than replaced by a zero vector. Joint condition dropout supplies an unconditional prediction, and all downstream samplers use the same classifier-free-guidance kernel

$$\epsilon_{\mathrm{cfg}} = (1 + w)\epsilon_\theta(v_t, t; e_f, e_s) - w\epsilon_\theta(v_t, t; \varnothing, \varnothing), \tag{3}$$

with a fixed scale w. Holding this kernel constant for the pilot, Unguided, and every SMC variant is essential: it makes the diffusion transition a common reference measure, so later differences can be attributed to value twisting rather than to a changed generator.

3.2 PILOT-FROZEN CONFIDENCE AND UNCERTAINTY-AWARE CLASSIFICATION

One sample cannot reveal subject-specific ambiguity in the conditional prior. Before task guidance, we therefore draw $G_{\mathrm{pilot}} = 16$ unguided samples. Their mean $\bar{X}_{\mathrm{p}}$ estimates the missing connectome, while dispersion provides node uncertainty u, edge confidence C^e, and four standardized subject statistics s. A Generative Confidence Estimator (GCE) combines s with quality-tier embedding e_r to produce confidence $c \in [0, 1]$, while C^e attenuates unreliable edges:

$$\begin{aligned} A^d &= \mathcal{T}_k(\bar{X}_{\mathrm{p}}) \odot C^e, \\ c &= \sigma(W_2 \,\mathrm{ReLU}\,(W_1[e_r; s] + b_1) + b_2)\,. \end{aligned} \tag{4}$$

where $\mathcal{T}_k$ is fitted on the training fold; observed-dMRI subjects use zero pilot dispersion. We freeze $\mathcal{U}_i = (u_i, C_i^e, s_i, c_i)$ and reuse it in classifier training and SMC. Recomputing uncertainty from guided particles would let entropy reduction increase apparent reliability and couple each proposal to the particle population. Freezing prevents this feedback, preserves particlewise proposals, and makes inference variants comparable. The provisional ensemble $C_0^{(1:M)}$ and final classifier C_1 share fMRI/dMRI graph encoders and cross-atlas attention. Node uncertainty augments the structural branch, and c_i controls its contribution to the fused prediction:

$$a_i = \sigma\Big(\mathrm{MLP}([h_i^f; h_i^d; c_i])\Big),\ h_i^{\mathrm{gate}} = a_i h_i^d + (1 - a_i) h_i^f,\ \widehat{y}_i = \mathrm{Linear}\Big(\mathrm{MLP}([h_i^f; h_i^d; h_i^{\mathrm{gate}}])\Big). \tag{5}$$

Each differently seeded $C_0^{(m)}$ is trained on pilot-completed subjects and then frozen, providing a smoother value signal than one decision surface. After SMC, C_1 is trained from scratch for final prediction, so the guiding model does not evaluate its own samples. Both stages retain the pilot context and use the objectives detailed in Appendix C.

3.3 Value-Twisted Sequential Monte Carlo

The central question is how to favor diagnostically useful connectomes without pretending that an unreasoned classifier-gradient trajectory is still a sample from the diffusion model. VT-SMC resolves this by separating what distribution is desired from *how particles reach it*. The diffusion prior supplies credibility, while a nonnegative value function expresses preference. Gradients are used only to construct efficient proposals; exact importance weights correct the resulting change of transition law.

A label-free value function. For particle x, each frozen provisional classifier f_{ϕ_m} predicts from x, the observed fMRI, and the subject's fixed pilot context $\mathcal{U}_i$. We define value loss as the mean predictive entropy

$$\ell_i(x) = \frac{1}{M}\sum_{m=1}^{M} H\Big(\operatorname{softmax} f_{\phi_m}(x, X_i^f; \mathcal{U}_i)\Big), \tag{6}$$

where $H(p) = -\sum_k p_k \log p_k$. This choice has three practical advantages. It requires no subject label during generation, is available under the same rule for training and test subjects, and becomes less sensitive to one classifier's local geometry through seed ensembling. It does not assume that confident predictions are correct; independent evaluators and fidelity constraints are therefore necessary. We initialize λ so that $\lambda \operatorname{sd}(\ell) \approx 0.5$ on training-pilot samples and select it only on validation data. Since $0 \le \ell \le \log K_y$, the resulting potential is bounded.

For observed context $o = (X^f, x^s)$ (or fMRI alone when sMRI is absent), this preference defines the terminal distribution that VT-SMC approximates:

$$\pi_\lambda(x_0 \mid o) \;\propto\; p_\theta(x_0 \mid o)\exp\{-\lambda \ell(x_0; o)\}. \tag{7}$$

Thus the task score changes a declared target law, rather than silently changing the diffusion model.

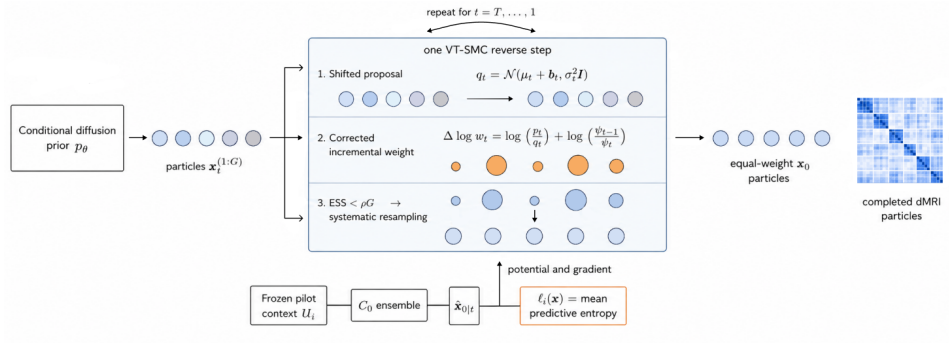

Figure 2: Overview of VT-SMC. At each reverse step, particles are proposed using a confidence-scaled entropy gradient obtained from the frozen pilot context and the provisional classifier ensemble. The Gaussian proposal shift is corrected by the exact density ratio, while adjacent twist potentials define the incremental weight. Systematic resampling is triggered by the effective sample size, and a terminal resampling produces equal-weight completed connectomes.

Twisting the reverse path. Directly applying the terminal tilt only after sampling would waste most particles in low-value regions. Instead, Figure 2 introduces intermediate potentials along the reverse path. At noisy steps the potential is weak because the clean estimate is unreliable; it strengthens as the signal-to-noise ratio increases:

$$\psi_t(v_t) = \exp\{-\lambda_t \ell(\widehat{v}_{0|t}(v_t))\}, \qquad \lambda_t = \lambda\sqrt{\bar{\alpha}_t}\ \ (0 < t < T), \quad \lambda_0 = \lambda, \quad \lambda_T = 0. \tag{8}$$

The endpoints are set explicitly: $\lambda_T = 0$ initializes particles from the diffusion prior and $\lambda_0 = \lambda$ recovers the desired terminal tilt. The schedule determines particle weights, not subject-specific guidance strength; confidence affects only proposal efficiency through Section 3.4.

A guided proposal with exact correction. At a guided step, the entropy gradient of the current clean estimate points toward a lower-loss region. We shift the base DDPM mean by b_t while keeping its covariance unchanged:

$$\begin{aligned} p_{\theta,t}(v_{t-1} \mid v_t) &= \mathcal{N}(\widetilde{\mu}_t, \widetilde{\sigma}_t^2 I), \\ q_t(v_{t-1} \mid v_t) &= \mathcal{N}(\widetilde{\mu}_t + b_t, \widetilde{\sigma}_t^2 I), \qquad b_t = \kappa_t \Delta_t. \end{aligned} \tag{9}$$

This shift makes useful states easier to reach, but it also changes the distribution. If $z_t \sim \mathcal{N}(0, I)$ and $v_{t-1} = \widetilde{\mu}_t + b_t + \widetilde{\sigma}_t z_t$, the equal-covariance Gaussians give the exact correction

$$\log \frac{p_{\theta,t}}{q_t} = -\frac{\|b_t\|_2^2}{2\widetilde{\sigma}_t^2} - \frac{\langle z_t, b_t \rangle}{\widetilde{\sigma}_t}. \tag{10}$$

computed in the K free symmetric coordinates. This term is the key distinction from gradient-only guidance, the mean shift silently defines an unknown sampling law if without it. With it, proposal design can improve efficiency without changing the declared target.

Weighting and resampling. Each particle combines the proposal correction with the change between adjacent potentials:

$$\delta \log w_t^{(g)} = \log \frac{p_{\theta,t}}{q_t} - \lambda_{t-1} \ell(\widehat{v}_{0|t-1}^{(g)}) + \lambda_t \ell(\widehat{v}_{0|t}^{(g)}). \tag{11}$$

Across a full path, the two potential terms telescope to $\exp\{-\lambda \ell(v_0)\}$, while Eq. 10 cancels the proposal shift. We normalize weights in float64 with log-sum-exp and compute $\mathrm{ESS} = (\sum_g (W^{(g)})^2)^{-1}$. When $\mathrm{ESS} < \rho G$, systematic resampling reallocates computation toward high-weight trajectories; a final resample returns equal-weight connectomes expected by the downstream loader. Clean estimates are reused between the potential and the next proposal, and gradients are evaluated only every k steps. The ablations isolate the two mechanisms: Grad-Only keeps the shift but discards weighting, FK-Resample keeps weighting but sets $b_t = 0$, and full VT-SMC combines both. Appendix B gives tensor-level pseudocode and endpoint handling.

3.4 Calibrated Guidance and Guarantees

Calibrating the proposal shift. The entropy-gradient norm varies with dimension, subject, and diffusion step. We normalize it in free-coordinate space and scale the shift by the reverse-process standard deviation. For $\widehat{g}_t = \nabla \ell / \|\nabla \ell\|_2$, a guided step uses

$$b_t = -\tau c^p \widetilde{\sigma}_t \widehat{g}_t, \qquad \Delta_t = b_t / \kappa_t, \tag{12}$$

where τ is the trust budget, c^p reduces shifts for uncertain subjects, and κ_t is the DDPM mean coefficient. Confidence controls proposal efficiency, not the target. At the terminal Dirac step, $b_0 = \Delta_0 = 0$ preserves absolute continuity.

Because the proposal and base transition share covariance, this normalization fixes both the log-ratio variance and transition KL:

Proposition 1 (Step calibration and KL budget)**.** *For each non-degenerate guided transition under Eq. 12,*

$$\frac{\|b_t\|_2}{\widetilde{\sigma}_t} = \tau c^p, \qquad \mathrm{Var}\left[\log \frac{p_{\theta,t}}{q_t} \,\middle|\, b_t\right] = \tau^2 c^{2p}, \qquad D_{\mathrm{KL}}(q_t \| p_{\theta,t}) = \frac{\tau^2 c^{2p}}{2}.$$

Thus the shift-induced path KL over guided steps $\mathcal{G}$ is $\frac{1}{2} \sum_{t \in \mathcal{G}} \tau^2 c^{2p}$, independent of connectome dimension K.

This makes τ interpretable and testable from log-ratio traces. The initialization $\tau \approx (\bar{c}^p \sqrt{N_g})^{-1}$ gives an order-one path budget. The budget controls correction difficulty; Eq. 10 ensures target correctness.

Target and support guarantees. SMC corrects the guided shifts through importance weights and resampling. If the potentials are bounded and positive, and $p_{\theta,t}$ is absolutely continuous with respect to q_t on non-degenerate transitions, the weighted estimate converges to the tilted target as the particle count grows:

Proposition 2 (SMC consistency)**.** *For any bounded test function, the weighted estimate from Eqs. 8–11 converges in probability to its expectation under Eq. 7 as $G \to \infty$. The normalizing-constant estimator is also consistent.*

When $q_t = p_{\theta,t}$, as in FK-Resample, task feedback only reweights or resamples states from the base reverse process:

Corollary 1 (Support-preserving endpoint)**.** *The terminal tilted measure in FK-Resample remains within the support of the diffusion prior.*

Together, normalization limits each shift, importance correction preserves the target, and resampling reallocates particles. These guarantees concern the learned prior, not biological truth or clinical benefit. Proofs appear in Appendix A; independent-evaluator and pseudo-missing audits appear in Section 4.

Cost and diagnostics. Inference costs $O(GTC_d + GMTC_c)$ for denoiser cost C_d and classifier cost C_c, with backward computation only at N_g guided steps. Correction adds $O(GT)$ scalar work without another denoiser call; memory is $O(GK)$. Stored weights, ESS, resampling events, potential and ratio traces, $\log \widehat{Z}$, and ancestry expose excessive tilting, particle collapse, or ratio errors. The complete diffusion–pilot–C_0–SMC–C_1 workflow is in Appendix B.

4 Experiments

4.1 Experimental Setup and Implementation

We evaluate ABIDE II (Di Martino et al., 2017) and an independent SCHZ/CTRL cohort (subjects with fMRI and structural connectomes under Lausanne129) (Vohryzek et al., 2020). The latter has no sMRI, its fMRI-conditioned model is retrained within each fold rather than transferred from ABIDE II. Dataset transformations, fold isolation, and cohort composition are specified in Appendix C and Table 3. The experiment has two comparative dimensions: classification performance and dMRI/structural-connectome generation quality. Classification metrics are accuracy (ACC), area under the ROC curve (AUC), specificity (SPE), and sensitivity (SEN). In ABIDE II, all four metrics come from the epoch selected by validation AUC. SCHZ/CTRL uses inner-split early stopping and one outer-test evaluation per fold. Reconstruction is assessed with Pearson correlation (PCC) and mean absolute error (MAE) on pseudo-missing subjects, whose observed dMRI/SC is hidden during generation.

All sampler variants share the denoiser checkpoint, classifier-free guidance scale, and initial-noise seeds. We use $T = 1000$, $G_{\text{pilot}} = 16$, $G = 8$, $M = 3$, $\rho = 0.5$, $p = 1$, and $\tau = 0.06$. An independently trained BrainNetCNN (Kawahara et al., 2017) audits evaluator-specific guidance. Optimization and calibration details are recorded in Appendix C and Table 4.

4.2 Results

Synthetic exact-target validation. A linear-Gaussian prior with a quadratic potential provides the exact target in Figure 3. Corrected samplers improve as the particle count grows, where Grad-Only tends to becomde flat because additional particles cannot remove transition bias. The separation becomes largest when correction matters most: at $d = 50$, $\lambda = 8$, and $G = 256$, VT-SMC reduces 2-Wasserstein error from 4.572 for Grad-Only to 1.711. The remaining distance to the exact-sample floor is nevertheless substantial, exposing finite-particle loss of prior–target overlap rather than suggesting that correction solves high-dimensional degeneracy. FK-Resample can be competitive under mild tilts, where base-prior particles already cover the target; gradient proposals become beneficial as that overlap narrows.

Classification and connectome generation. Table 1 compares classification and pseudo-missing reconstruction in independent blocks. Classification baselines are SGVM (Li et al., 2025), Brain-OSM (Guo et al., 2026), ContrastPool (Xu et al., 2024), A-GCL (Zhang et al., 2023b), CWBrain-Net (Jeong et al., 2026), LG-GNN (Zhang et al., 2023a), KMGCN (Zeng et al., 2024), NEGAN (Chen et al., 2024), MMGL (Zheng et al., 2022b), MM-GTUNets (Cai et al., 2025), and Cross-GNN (Yang

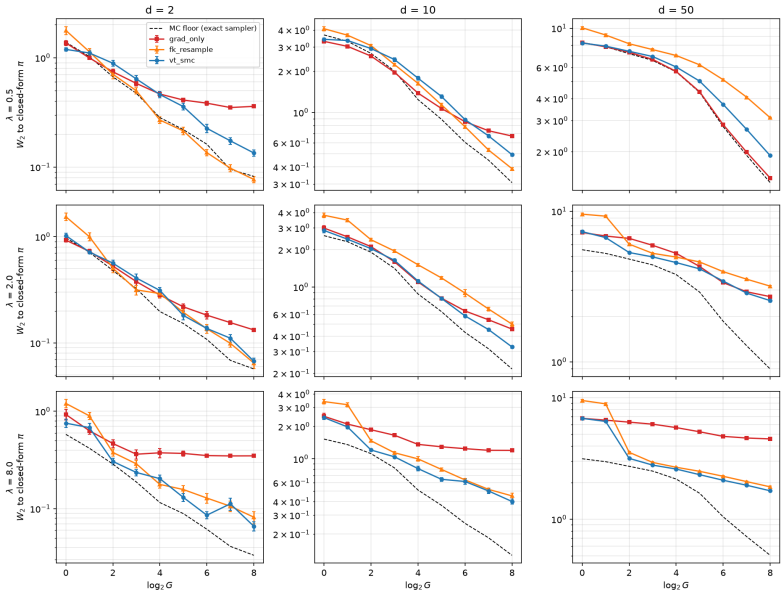

Figure 3: Synthetic test of whether task-guided samplers recover their declared tilted distribution rather than merely attaining a low task loss. We compare uncorrected gradient guidance (Grad-Only), resampling without a guided proposal (FK-Resample), and corrected VT-SMC against a closed-form Gaussian target. Columns vary dimension ($d = 2, 10, 50$), rows vary tilt strength ($\lambda = 0.5, 2, 8$), and the horizontal axis is $\log_2 G$. Curves report mean 2-Wasserstein error over 20 seeds; the dashed line is the finite-sample floor from exact target draws.

Table 1: Two-dimensional comparison of classification performance and dMRI/structural-connectome reconstruction quality. Classification values are fold-aggregated percentages; PCC and MAE use outer-fold pseudo-missing subjects. Implementation details are given in Appendix C.8.

Classification performance					**dMRI/structural-connectome generation quality**		
Method	ACC ↑	AUC ↑	SPE ↑	SEN ↑	Method	PCC ↑	MAE ↓
ABIDE II—single-input prediction					*ABIDE II pseudo-missing*		
SGVM	66.93	71.87	62.09	71.18	DSBGM	0.5179	0.2482
BrainOSM	62.37	63.25	63.04	61.55	MGCN-GAN	0.8020	0.0415
ContrastPool	66.78	67.46	61.77	71.19	DDPM	0.7065	0.1326
A-GCL	62.93	62.49	55.21	69.74	DDIM	0.7378	0.1452
CWBrainNet	56.05	54.86	36.49	73.28	Dif-GSR	0.7938	0.0482
LG-GNN	61.59	59.29	51.90	70.16	VT-SMC (ours)	0.8324	0.0173
KMGCN	65.10	65.85	70.90	58.65			
NEGAN	67.59	67.30	63.27	71.39			
MMGL	70.14	70.11	69.67	70.56			
ABIDE II—multimodal prediction					*SCHZ/CTRL Lausanne129 pseudo-missing*		
MM-GTUNets	69.69	69.27	**75.43**	63.02	DSBGM	0.3190	0.3971
Cross-GNN	66.81	66.80	66.35	67.22	MGCN	0.6955	0.1357
VT-SMC (ours)	**72.82**	**81.01**	75.40	**73.80**	DDPM	0.5882	0.2454
					DDIM	0.5637	0.1975
					Dif-GSR	0.6286	0.1468
					VT-SMC (ours)	0.7272	0.0856
SCHZ/CTRL Lausanne129—fMRI-only adaptation							
VT-SMC (ours)	66.67	74.70	53.13	88.89			

et al., 2024); reconstruction baselines are DSBGM (Tang et al., 2023), MGCN-GAN/MGCN (Zhang et al., 2022), DDPM (Ho et al., 2020), DDIM (Song et al., 2021), and Dif-GSR (Rajadhyaksha & Rekik, 2023).

On ABIDE II, VT-SMC reaches 72.82% ACC and 81.01% AUC, exceeding the listed alternative by 2.68 and 9.14 percentage points, respectively. Its 75.40% specificity is within 0.03 points of

MM-GTUNets, while its 73.80% sensitivity is the highest in the table. Thus the AUC gain is not obtained by simply favoring the positive class: discrimination improves while specificity is preserved. This pattern is consistent with generated structure acting as task-aware regularization rather than merely shifting the operating threshold. On SCHZ/CTRL, the fMRI-only adaptation obtains 74.70% AUC and 88.89% sensitivity, but only 53.13% specificity. This asymmetric operating point and the absence of classification baselines limit the result to evidence of feasibility under a new diagnosis and atlas, not comparative clinical superiority.

For reconstruction, VT-SMC also leads both cohorts on PCC and MAE. On ABIDE II it improves PCC from 0.8020 for the nearest baseline to 0.8324 and reduces MAE from 0.0415 to 0.0173, a 58.3% reduction. On SCHZ/CTRL, PCC rises from 0.6955 to 0.7272 and MAE falls from 0.1357 to 0.0856 relative to MGCN. Agreement between correlation and absolute error argues against a mere rescaling effect, while the second-cohort result indicates that the conditional modeling workflow survives a different atlas and the removal of sMRI. This cross-method advantage can coexist with the within-backbone fidelity cost below: a strong learned prior may outperform other generators even when task tilting moves away from its own reconstruction optimum. Appendix Table 8 and Figure 6 provide topology, selective-risk, and case-level evidence for this distinction.

Fidelity–utility and particle diagnostics. Figure 4 jointly reports independent-evaluator utility, fidelity, ESS, and proposal-correction traces for the frozen operating point.

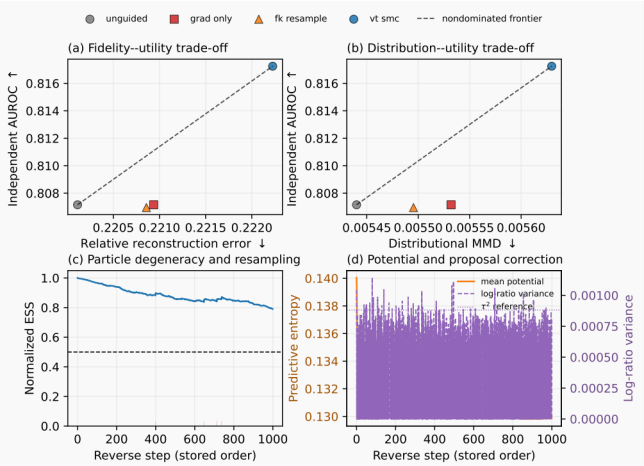

Figure 4: Fidelity–utility and particle diagnostics on ABIDE II. (a,b) Independent-evaluator AUROC versus relative reconstruction error and MMD for 148 pseudo-missing subjects; dashed segments connect nondominated points. (c) Mean normalized ESS over the 1000 stored reverse steps; the dashed line is the resampling threshold $\rho = 0.5$ and rug marks denote resampling events. (d) Mean predictive entropy and proposal log-ratio variance, with the τ^2 diagnostic reference.

VT-SMC raises independent-evaluator AUROC from 0.8072 to 0.8172, while relative error changes only from 0.2201 to 0.2222 and MMD from 0.00544 to 0.00563. The transfer gain is therefore not confined to the ensemble defining the potential, weakening a reward-hacking explanation. At the same time, the nonzero fidelity change confirms the Prior–Utility Dilemma rather than hiding it. Appendix Table 8 sharpens this interpretation: edge correlation is nearly unchanged, whereas AURC improves from 0.2595 to 0.2425. Guidance therefore appears to reorganize task-relevant ranking and abstention more than it globally repairs connectome edges. Appendix Table 5 states the complementary evaluator-transfer criterion.

Normalized ESS remains above the 0.5 resampling trigger and finishes near 0.79, while the proposal log-ratio variance stays bounded. Together, these diagnostics indicate that the calibrated shift preserves sufficient overlap for eight particles and that resampling is corrective rather than continuous. An overstrong or incorrectly scaled proposal would instead appear as ratio inflation followed by ESS collapse. Appendix Tables 6 and 7 predeclare compute and failure accounting; their placeholders support no empirical claim.

Table 2: Component ablations on ABIDE II and SCHZ/CTRL. "Exact ratio" denotes Eq. 10; PCC and MAE are computed on outer-fold pseudo-missing subjects. Classification AUC is higher-is-better, whereas fidelity need not improve under task tilting.

Dataset	Variant	Exact ratio	Resample	M	AUC ↑	PCC ↑	MAE ↓
ABIDE II	Unguided	–	–	–	76.02	0.8645	0.0153
	Grad-Only	no	no	3	76.57	0.8364	0.0164
	FK-Resample	n/a	yes	3	76.78	0.8405	0.0162
	VT-SMC without resampling	yes	no	3	77.31	0.8538	0.0185
	VT-SMC with $p = 0$	yes	yes	3	78.35	0.8394	0.0178
	VT-SMC with $M = 1$	yes	yes	1	78.26	0.8332	0.0182
	Full VT-SMC	yes	yes	3	81.01	0.8324	0.0173
SCHZ-CTRL	Unguided	–	–	–	72.84	0.7364	0.0719
	Grad-Only	no	no	3	73.25	0.7281	0.0845
	FK-Resample	n/a	yes	3	72.98	0.7292	0.0747
	VT-SMC	yes	yes	3	74.70	0.7272	0.0856

Ablations. Table 2 isolates proposal correction, resampling, confidence modulation, and ensemble guidance on both datasets.

On ABIDE II, full VT-SMC raises AUC from 76.02 for Unguided to 81.01. Grad-Only and FK-Resample reach only 76.57 and 76.78: the former moves particles without correcting the induced law, whereas the latter preserves weighting semantics but cannot steer proposals toward narrow value regions. Their small gains for different reasons explain why proposal guidance and correction must be combined. Removing resampling lowers AUC to 77.31, while disabling confidence modulation or using one classifier gives 78.35 and 78.26. Relative to the full model, these drops of 3.70, 2.66, and 2.75 points indicate complementary contributions from particle allocation, subject-specific trust, and cross-seed smoothing rather than one dominant component.

The gain is not free: on ABIDE II, PCC changes from 0.8645 to 0.8324 and MAE from 0.0153 to 0.0173 relative to Unguided. SCHZ/CTRL shows the same direction, with a smaller AUC improvement of 1.86 points and a moderate fidelity decrease. Repetition of this pattern across cohorts supports a controllable fidelity–utility trade-off, while the smaller external-cohort effect is consistent with limited data and fMRI-only retraining. Appendix Figure 5 further shows that recomputing uncertainty worsens selective risk, supporting a frozen pilot that cannot be manipulated by the particles it guides.

5 Conclusion

Missing-connectome completion exposes the Prior–Utility Dilemma: preserving the learned conditional prior does not guarantee utility, while aggressive task guidance can obscure the sampled distribution. We take VT-SMC as an uncertainty-aware, distribution-accountable inference layer between a label-free generator and a downstream predictor. Its frozen pilot prevents confidence feedback, its ensemble entropy defines the task preference, and proposal-corrected SMC turns preference into an explicit tilted distribution. Across the experiments, one pattern is consistent: exact correction recovers the synthetic target, the complete design improves diagnostic discrimination on ABIDE II and after SCHZ/CTRL retraining, and independent evaluation shows that utility gain carries a measurable fidelity cost. Thus VT-SMC does not eliminate the dilemma; it makes the trade-off controllable, diagnosable, and scientifically interpretable. Future work should estimate uncertainty on larger multi-site paired cohorts, complete matched classification and reconstruction baselines on the second dataset, test site-held-out calibration, and adapt particle count and tilt strength to a fixed subject-level compute budget.

AI Use Statement

In this work, generative AI tools were utilized for the following required disclosure tasks: providing critical derivation steps for mathematical proofs, assisting in the writing and organization of proof processes, implementing experimental and methodological code, translating manuscript texts,

designing the overall experimental scheme, and offering optimization feedback on the proposed experimental methods. No generative AI tools were used for other mandatory research subtasks including synthetic dataset generation, theoretical model and conceptual framework construction, formulation of mathematical claims, proposal and refinement of research hypotheses, dataset cleaning and reformatting, qualitative and thematic data analysis, and result interpretation. All AI-assisted mathematical derivations, proofs, code, experimental designs, and translated text were reviewed and verified by the authors. The authors take full responsibility for the final content of this work, including all claims, methods, experiments, code, and text produced with the assistance of generative AI.

REFERENCES

Reza Azad, Nika Khosravi, Mohammad Dehghanmanshadi, Julien Cohen-Adad, and Dorit Merhof. Medical image segmentation on MRI images with missing modalities: A review. *arXiv preprint arXiv:2203.06217*, 2022.

L. Cai, W. Zeng, H. Chen, H. Zhang, Y. Li, Y. Feng, H. Yan, L. Bian, W. T. Siok, and N. Wang. MM-GTUNets: Unified multi-modal graph deep learning for brain disorders prediction. *IEEE Transactions on Medical Imaging*, 44(9):3705–3716, 2025.

Agisilaos Chartsias, Thomas Joyce, Mario Valerio Giuffrida, and Sotirios A. Tsaftaris. Multimodal MR synthesis via modality-invariant latent representation. *IEEE Transactions on Medical Imaging*, 37(3):803–814, 2018.

Y. Chen, J. Yan, M. Jiang, T. Zhang, Z. Zhao, W. Zhao, J. Zheng, D. Yao, R. Zhang, K. M. Kendrick, and X. Jiang. Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6): 7275–7286, 2024.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436, 2006. doi: 10.1111/j.1467-9868.2006.00553.x.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794, 2021.

Adriana Di Martino, David O'Connor, Bosi Chen, Kaat Alaerts, Jeffrey S. Anderson, Michal Assaf, Joshua H. Balsters, Leslie Baxter, Anita Beggiato, Sylvie Bernaerts, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Scientific Data*, 4:170010, 2017. doi: 10.1038/sdata.2017.10.

Zolnamar Dorjsembe, Sodtavilan Odonchimed, and Furen Xiao. Conditional diffusion models for medical image synthesis. *Medical Image Analysis*, 88:102867, 2023.

Randal Douc, Olivier Cappé, and Eric Moulines. Comparison of resampling schemes for particle filtering. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pp. 64–69, 2005. doi: 10.1109/ISPA.2005.195385.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *International Conference on Machine Learning*, pp. 1321–1330, 2017.

G. Guo, G. Wen, L. Liu, R. Song, P. Cao, J. Yang, and O. R. Zaiane. BrainOSM: Outlier screening for multi-view functional brain network analysis. *Computer Methods and Programs in Biomedicine*, 273:109092, 2026.

Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. HeMIS: Hetero-modal image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 469–477, 2016.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

A.-Y. Jeong, D.-W. Heo, and H.-I. Suk. Complex wavelet-based transformer for neurodevelopmental disorder diagnosis via direct modeling of real and imaginary components. *Medical Image Analysis*, 109:103914, 2026.

Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, Jill G. Zwicker, and Ghassan Hamarneh. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

Z. Li, J. Liu, M. Sun, F. Zhang, B. Hu, and Q. Dong. Similarity-guided multi-view functional brain network fusion. *Medical Image Analysis*, 105:103632, 2025.

Christian A. Naesseth, Fredrik Lindsten, and Thomas B. Schön. Elements of sequential monte carlo. *Foundations and Trends in Machine Learning*, 12(3):187–306, 2019. doi: 10.1561/2200000074.

Natalia Neverova, Christian Wolf, Graham W. Taylor, and Florian Nebout. ModDrop: Adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1692–1706, 2016.

Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D. Sculley, Sebastian Nowozin, Joshua V. Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023.

Nishant Rajadhyaksha and Islem Rekik. Diffusion-based graph super-resolution with application to connectomics. In *Predictive Intelligence in Medicine*, pp. 96–107. Springer, 2023. doi: 10.1007/978-3-031-46005-0_9.

Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=St1giarCHLP.

Haoteng Tang, Lei Guo, Xiyao Fu, Yalin Wang, Scott Mackin, Olusola Ajilore, Alex D. Leow, Paul M. Thompson, Heng Huang, and Liang Zhan. Signed graph representation learning for functional-to-structural brain network mapping. *Medical Image Analysis*, 83:102674, 2023. doi: 10.1016/j.media.2022.102674.

Jakub Vohryzek, Yasser Aleman-Gomez, Alessandra Griffa, Jeni Raoul, Martine Cleusix, Philipp S. Baumann, Philippe Conus, Kim Do Cuenod, and Patric Hagmann. Structural and functional connectomes from 27 schizophrenic patients and 27 matched healthy adults. Zenodo dataset, 2020. URL https://doi.org/10.5281/zenodo.3758534.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Renjie Wu, Hu Wang, Hsiang-Ting Chen, and Gustavo Carneiro. Deep multimodal learning with missing modality: A survey. *arXiv preprint arXiv:2409.07825*, 2024.

J. Xu, Q. Bian, X. Li, A. Zhang, Y. Ke, M. Qiao, W. Zhang, W. K. J. Sim, and B. Gulyas. Contrastive graph pooling for explainable classification of brain networks. *IEEE Transactions on Medical Imaging*, 43(9):3292–3305, 2024.

Y. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, and T. Ma. Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. *IEEE Transactions on Medical Imaging*, 43(1):108–121, 2024.

X. Zeng, J. Gong, W. Li, and Z. Yang. Knowledge-driven multi-graph convolutional network for brain network analysis and potential biomarker discovery. *Medical Image Analysis*, 99:103368, 2024.

Hao Zhang, Ran Song, Liping Wang, Lin Zhang, Dawei Wang, Cong Wang, and Wei Zhang. Classification of brain disorders in rs-fMRI via local-to-global graph neural networks. *IEEE Transactions on Medical Imaging*, 42(2):444–455, 2023a. doi: 10.1109/TMI.2022.3219260.

Lu Zhang, Li Wang, Dajiang Zhu, and Alzheimer's Disease Neuroimaging Initiative. Predicting brain structural network using functional connectivity. *Medical Image Analysis*, 79:102463, 2022. doi: 10.1016/j.media.2022.102463.

Shengjie Zhang, Xiang Chen, Xin Shen, Bohan Ren, Ziqi Yu, Haibo Yang, Xi Jiang, Dinggang Shen, Yuan Zhou, and Xiao-Yong Zhang. A-GCL: Adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders. *Medical Image Analysis*, 90:102932, 2023b. doi: 10.1016/j.media.2023.102932.

Guangcong Zheng, Shengming Li, Hui Wang, Taiping Yao, Yang Chen, Shouhong Ding, and Xi Li. Entropy-driven sampling and training scheme for conditional diffusion generation. In *Computer Vision–ECCV 2022*, pp. 754–769. Springer, 2022a.

Shuai Zheng, Zhenfeng Zhu, Zhizhe Liu, Zhenyu Guo, Yang Liu, Yuchen Yang, and Yao Zhao. Multi-modal graph learning for disease prediction. *IEEE Transactions on Medical Imaging*, 41(9): 2207–2216, 2022b. doi: 10.1109/TMI.2022.3159264.

APPENDIX

Appendices A–B give derivations and pseudocode; Appendix C records datasets, fold isolation, optimization, baselines, and artifact checks; Appendix D contains extended audits; and Appendices E–F specify figure provenance and limitations.

A PROOFS AND TECHNICAL DETAILS

A.1 WHY COMPLETION ADDS NO CONDITIONAL LABEL INFORMATION

Let $O = (X^f, x^s)$ and let ε denote all diffusion sampling randomness. A label-free generator has $\widehat{X}^d = h_\theta(O, \varepsilon)$ with $\varepsilon \perp y \mid O$. Conditioned on O, the variables form the Markov chain $y \to O \to \widehat{X}^d$, or directly $p(\widehat{X}^d \mid O, y) = p(\widehat{X}^d \mid O)$. Therefore,

$$I\Big(y; \widehat{X}^d \mid X^f, x^s\Big) = 0. \tag{13}$$

This does not imply that adding $\widehat{X}^d$ to a restricted finite-sample classifier has zero empirical effect: the representation and optimization path can change. It does rule out interpreting such a change as recovered subject-specific label information.

Equivalently, conditional mutual information can be written as

$$I(y; \widehat{X}^d \mid O) = \mathbb{E}_{O,y} D_{\mathrm{KL}}\Big(p(\widehat{X}^d \mid O, y) \| p(\widehat{X}^d \mid O)\Big) = 0. \tag{14}$$

The equality assumes a fixed fold-trained generator, conditionally independent sampling noise, and no use of the evaluated subject's label in selecting, guiding, or filtering samples. It would fail for the former train-CE/test-entropy procedure because generated training samples depended on y while generated test samples did not. The revised label-free potential restores the conditional-independence premise and removes that train-test intervention shift.

This result is about information in random variables, not the excess risk of a restricted learning algorithm. A generated representation can alter optimization, regularization, smoothness, and effective architecture even though it is conditionally redundant for a Bayes-optimal predictor. The deterministic, latent, random, and shuffled controls in Section 4 test which of these finite-sample mechanisms explains any observed gain. See also the data-processing discussion in standard information theory (Cover & Thomas, 2006).

A.2 FEYNMAN-KAC SEQUENCE AND TELESCOPING WEIGHTS

For notational simplicity, consider one subject and omit o. Let the reverse path density be

$$p_\theta(v_{0:T}) = p_T(v_T) \prod_{t=1}^{T} p_{\theta,t}(v_{t-1} \mid v_t). \tag{15}$$

For the intermediate twists in Eq. 8, define $\gamma_t(v_{t:T}) = p_\theta(v_{t:T})\psi_t(v_t)$. Particles are initialized from p_T, and the endpoints are explicitly set to $\psi_T = 1$ and $\psi_0(v_0) = \exp(-\lambda \ell(v_0))$. If transition q_t replaces $p_{\theta,t}$, the unnormalized incremental weight is

$$G_t(v_t, v_{t-1}) = \frac{p_{\theta,t}(v_{t-1} \mid v_t)}{q_t(v_{t-1} \mid v_t)} \frac{\psi_{t-1}(v_{t-1})}{\psi_t(v_t)}. \tag{16}$$

Multiplying across reverse steps yields

$$\prod_{t=T}^{1} G_t = \frac{p_\theta(v_{0:T})}{p_T(v_T) \prod_{t=T}^{1} q_t(v_{t-1} \mid v_t)} \frac{\psi_0(v_0)}{\psi_T(v_T)} \tag{17}$$

$$= \frac{p_\theta(v_{0:T}) \exp(-\lambda \ell(v_0))}{p_T(v_T) \prod_{t=T}^{1} q_t(v_{t-1} \mid v_t)}. \tag{18}$$

Thus arbitrary positive intermediate twists affect particle efficiency but telescope out of the terminal target. Taking logarithms of Eq. 16 gives Eq. 11. This also clarifies why the two schedule endpoints must be exact rather than merely numerically close to zero and one.

A.3 Proof of Proposition 1

Equation 12 gives $\|b_t\|_2 = \tau c^p \widetilde{\sigma}_t$ immediately. The random component of Eq. 10 is $-\langle z_t, b_t\rangle/\widetilde{\sigma}_t$. For $z_t \sim \mathcal{N}(0, I_K)$ and fixed b_t, $\langle z_t, b_t\rangle \sim \mathcal{N}(0, \|b_t\|_2^2)$. Its variance after division by $\widetilde{\sigma}_t$ is therefore $\tau^2 c^{2p}$.

The deterministic term of the log ratio is $-\tau^2 c^{2p}/2$. Consequently,

$$\log \frac{p_{\theta,t}}{q_t} \,\Big|\, b_t \sim \mathcal{N}\big(-\tfrac{1}{2}\tau^2 c^{2p}, \tau^2 c^{2p}\big)\,. \tag{19}$$

Equation 19 is a direct implementation diagnostic. A factor-of-two variance error indicates that off-diagonal matrix entries were counted twice; systematic deviation in the mean indicates an inconsistent standard deviation or sign.

KL and path-divergence component. Two Gaussians with equal covariance $\widetilde{\sigma}_t^2 I_K$ and means separated by b_t satisfy

$$D_{\mathrm{KL}}\big(\mathcal{N}(\widetilde{\mu}_t + b_t, \widetilde{\sigma}_t^2 I)\,\|\,\mathcal{N}(\widetilde{\mu}_t, \widetilde{\sigma}_t^2 I)\big) = \frac{\|b_t\|_2^2}{2\widetilde{\sigma}_t^2}. \tag{20}$$

Substitution of Eq. 12 gives the one-step KL stated in Proposition 1. The chain rule sums transition KL terms along the path, and dimension independence follows from fixing the total free-coordinate L_2 norm rather than a per-coordinate or L_∞ displacement.

When confidence is frozen for a subject and gradients are applied on the set $\mathcal{G}$, the shift-induced path divergence is

$$D_{\mathrm{KL}}(q(v_{0:T})\|p_\theta(v_{0:T})) = \frac{|\mathcal{G}|\tau^2 c^{2p}}{2}, \tag{21}$$

provided the numerical Δ cap is inactive. If the cap activates, Eq. 21 becomes an upper bound because it can only reduce $\|b_t\|_2$. By data processing, the divergence between proposal and prior terminal marginals is no greater than the path divergence. This bound describes the proposal before importance correction; the weighted SMC target remains Eq. 7.

A.4 Proposal-Ratio Derivation

Writing $x = \widetilde{\mu} + b + \widetilde{\sigma} z$, the difference of equal-covariance Gaussian log densities is

$$\log p(x) - \log q(x) = -\frac{\|x - \widetilde{\mu}\|_2^2 - \|x - \widetilde{\mu} - b\|_2^2}{2\widetilde{\sigma}^2} \tag{22}$$

$$= -\frac{\|b\|_2^2}{2\widetilde{\sigma}^2} - \frac{\langle z, b\rangle}{\widetilde{\sigma}}. \tag{23}$$

For symmetric matrices B and Z, the free-coordinate norm satisfies

$$\|\operatorname{vech}(B)\|_2^2 = \tfrac{1}{2}\left(\|B\|_F^2 + \|\operatorname{diag}(B)\|_2^2\right), \tag{24}$$

which explains why summing over the full matrix would double-count off-diagonal contributions and invalidate the weight.

The corresponding inner-product identity is

$$\langle \operatorname{vech}(Z), \operatorname{vech}(B)\rangle = \tfrac{1}{2}\left(\langle Z, B\rangle_F + \langle \operatorname{diag}(Z), \operatorname{diag}(B)\rangle\right). \tag{25}$$

Both identities are tested numerically. Sampling is performed directly in free coordinates and mirrored once; no sampled v_{t-1} is subsequently clipped or symmetrized. Such a post-transition projection would change the proposal law and invalidate the Gaussian ratio even if the resulting matrix appeared plausible.

A.5 Proof Sketch of Proposition 2

Define path targets $\gamma_t(v_{t:T}) = p_\theta(v_{t:T} \mid o)\psi_t(v_t)$. Equation 11 is the Radon-Nikodym correction for replacing the base transition by q_t, multiplied by the ratio of adjacent twists. These potential

ratios telescope, leaving the terminal factor $\exp(-\lambda\ell(v_0))$. Standard Feynman-Kac particle-system consistency then applies under bounded positive potentials and absolute continuity (Del Moral et al., 2006; Naesseth et al., 2019). At $t = 0$, the DDPM transition is Dirac; the implementation forces $b_0 = 0$, so $p_0 = q_0$ and defines the proposal ratio as zero rather than applying a Gaussian formula with an artificial variance clamp.

More precisely, for any bounded measurable test function φ and normalized terminal weights $W_0^{(g)}$, standard SMC laws of large numbers imply

$$\sum_{g=1}^{G} W_0^{(g)} \varphi(v_0^{(g)}) \xrightarrow[G\to\infty]{\mathbb{P}} \int \varphi(v)\pi_\lambda(v)\, dv. \tag{26}$$

Adaptive resampling based on a measurable ESS rule preserves consistency under the same regularity conditions. Systematic resampling has lower computational cost than multinomial resampling, but its ancestry dependence is why we report unique ancestors rather than treating post-resampling copies as independent draws (Douc et al., 2005).

A.6 Pilot Context and Per-Particle Factorization

Standard SMC requires a proposal that factors over particles conditional on their ancestors. If uncertainty were recomputed as a standard deviation across the current G particles, the displacement of particle g would be a function of the full set $(v_t^{(1)}, \ldots, v_t^{(G)})$. Perturbing one particle would then change every transition, while only one realization of the joint G-particle system is available. The independent pilot replaces this endogenous statistic with $\mathcal{U}_i$, fixed before SMC. Conditional on $\mathcal{U}_i$,

$$q_t(v_{t-1}^{(1:G)} \mid v_t^{(1:G)}, \mathcal{U}_i) = \prod_{g=1}^{G} q_t(v_{t-1}^{(g)} \mid v_t^{(g)}, \mathcal{U}_i), \tag{27}$$

restoring the particlewise proposal used by Eq. 11. The code verifies this factorization by perturbing particle zero and requiring all other proposal displacements to remain bitwise unchanged.

A.7 Terminal Resampling and the Dirac Step

The legacy classifier loader expects G equal-weight matrices. We therefore compute all weighted summaries and save $\log W$ before an unconditional terminal systematic resample. For any test function, the conditional expectation of the resampled empirical mean equals the pre-resampling weighted estimate up to the standard properties of the chosen resampling scheme. Duplicated ancestors reduce diversity but do not justify recomputing generative uncertainty: the latter remains pilot-derived.

At $t = 0$, the true DDPM posterior variance is zero even if an implementation clamps its logarithm for numerical stability. A nonzero proposal displacement would make two distinct Dirac measures mutually singular. The implementation therefore imposes $b_0 = \Delta_0 = 0$ and sets $\log(p_0/q_0) = 0$ by definition. Only the final twist increment remains. This is checked before any terminal statistics are saved.

B Algorithm

Complete workflow. Taken together, VT-SMC follows a six-stage pipeline. First, the paired training subjects fit the label-free conditional diffusion prior. Second, unguided pilot samples are generated and converted into frozen uncertainty and confidence. Third, M provisional C_0 classifiers learn from pilot-completed training data. Fourth, their ensemble entropy defines the value tilt, and calibrated VT-SMC produces corrected task-aware connectomes without reading subject labels. Fifth, a fresh C_1 is trained on these completions while retaining the original pilot context. Finally, C_1, independent evaluators, reconstruction metrics, and particle diagnostics are evaluated on held-out subjects. This ordering keeps anatomical modeling, task preference, corrected inference, and final evaluation as distinct roles; Appendix C.1 specifies how the sequence is repeated without cross-fold leakage.

Algorithm 1: Value-Twisted SMC for one subject.

1. Input observed context o, frozen pilot context $\mathcal{U}$, frozen $C_0^{(1:M)}$, particle count G, λ, τ, p, and ESS fraction ρ. Initialize $v_T^{(g)} \sim \mathcal{N}(0, I_K)$ and $\log W_T^{(g)} = -\log G$.
2. For $t = T-1, \ldots, 0$, evaluate all particles in one batched denoiser path to obtain $\widehat{v}_{0|t}^{(g)}$ and the base reverse mean. Evaluate Eq. 6 for every particle.
3. Complete the preceding transition's cached weight using Eq. 11. Normalize with log-sum-exp and compute ESS. If ESS $< \rho G$, systematically resample every particle-aligned state and reset weights to $1/G$.
4. On scheduled gradient steps with $t > 0$, detach $\widehat{v}_{0|t}$ from the denoiser, differentiate only Eq. 6, normalize in free-coordinate L_2, and construct b_t with Eq. 12. Otherwise set $b_t = 0$.
5. Draw symmetric free-coordinate noise, apply the shifted transition, compute Eq. 10, and cache the terms needed when $\ell(\widehat{v}_{0|t-1})$ becomes available.
6. At termination, save weighted moments, weights, $\log \widehat{Z}$, traces, and diagnostics. Systematically resample once more to return G equal-weight terminal particles and record the number of unique ancestors.

Algorithm 2: fold-level self-bootstrapping workflow.

1. Split subjects with five-fold stratification on diagnosis and observed-dMRI status. Fit all preprocessing constants on the fold training split.
2. Train the label-free diffusion prior only on paired training subjects. Select its checkpoint on paired validation subjects using the diffusion validation objective, not the downstream test score.
3. Generate G_{pilot} unguided samples for every subject requiring completion. Save full particles, pilot mean and standard deviation, node uncertainty, edge confidence, raw and standardized subject statistics, fold metadata, and the CFG scale.
4. Train $M = 3$ provisional C_0 classifiers on training subjects using pilot content and pilot uncertainty. Store model seeds, stage identifiers, and statistic-normalization constants.
5. Calibrate λ from training pilot potentials, select λ and τ on validation data subject to fidelity constraints, and generate terminal VT-SMC particles for training and test subjects without reading their labels.
6. Train C_1 from scratch on guided training content with the same pilot uncertainty. Freeze its checkpoint using validation metrics, evaluate once on held-out subjects, and repeat for all outer folds and reporting seeds.

C Dataset and Implementation Details

C.1 Problem Setup and Fold-Safe Workflow

Let $\mathcal{D} = \{(X_i^f, x_i^s, X_i^d, y_i, r_i)\}_{i=1}^N$, where r_i indicates observed dMRI. In outer fold k, normalizers and classifiers use only $\mathcal{I}_k^{\text{tr}}$, while diffusion uses the paired subset $\mathcal{P}_k^{\text{tr}} = \{i \in \mathcal{I}_k^{\text{tr}} : r_i = 1\}$. The diffusion–pilot–$C_0$–SMC–$C_1$ sequence is completed independently in every fold, and all artifacts carry fold identifiers. Consequently, held-out labels and dMRI cannot enter preprocessing, generation, potential calibration, or model selection. ABIDE II folds stratify jointly by diagnosis and dMRI availability. On SCHZ/CTRL, where all subjects have paired fMRI and SC, an outer test fold remains untouched while an inner split supplies early stopping and calibration. For both datasets, pseudo-missing scoring reveals the held-out structural connectome only after generation.

C.2 Cohort Construction and Representations

ABIDE II aggregates de-identified multimodal data from multiple institutions (Di Martino et al., 2017). The present preprocessing yields 802 subjects with fMRI and sMRI, of whom 148 also have usable dMRI and 654 do not.

Functional connectivity is represented under CC200 as a 200×200 correlation matrix. Structural connectivity is an AAL116 streamline-count matrix. The sMRI input is a regional morphological feature vector. Count-valued dMRI entries use the same signed-log or log transform in diffusion training, pilot statistics, pseudo-missing evaluation, and downstream loading. Transform inversion, diagonal handling, sparsification, and self-loop rules will be listed from the final configuration rather than inferred from intermediate arrays.

Table 3: ABIDE II cohort after preprocessing. Counts distinguish subjects with observed dMRI from the operational missing-dMRI cohort; age is in years and sites are unique acquisition sites.

Group	Total	Observed dMRI	Missing dMRI
All subjects	802	148	654
ASD	386	106	280
Average age	14.10	10.79	14.85
Sites	14	4	11

Table 3 shows the marked modality imbalance motivating pseudo-missing evaluation: only 148 of 802 subjects have dMRI. The paired subset is never pooled across outer folds when fitting the generator or preprocessing constants.

C.3 SCHZ/CTRL LAUSANNE129 ADAPTATION

The second cohort is the publicly released structural and functional connectome dataset of 27 people with schizophrenia and 27 matched healthy adults (Vohryzek et al., 2020). We use its raw-Pearson fMRI and raw fiber-count structural connectomes under Lausanne129. Because sMRI is absent, the cross-attention context is set to `None`; no zero vector is used as a surrogate modality. Each outer fold contains approximately 43 training and 11 test subjects, with an inner train/validation split for early stopping and calibration. SC is transformed as $\log(1 + x)/m_k$, where m_k is the maximum training-split log count stored in the fold checkpoint, and is inverted to raw counts for final evaluation. FC is not resized, interpolated, clipped, or silently re-symmetrized. The outer-test SC is hidden until pseudo-missing scoring. This experiment evaluates fold-wise retraining under a new diagnosis, atlas, and numerical semantics, not zero-shot transfer from ABIDE II.

C.4 FOLD CONSTRUCTION AND NORMALIZATION

The loader forms a four-level stratification label $2y + r$, ensuring that diagnosis and observed-dMRI status are represented in each fold. The split is shuffled with a fixed seed. Training indices fit fMRI and sMRI preprocessing, graph thresholds, and the mean and standard deviation of the four pilot statistics. These constants are serialized with the classifier checkpoint and compared against the pilot metadata at load time. Site is not currently part of the splitter; site imbalance is therefore reported and addressed through a separate leave-site-out analysis rather than hidden inside nominally stratified folds.

C.5 LEAKAGE CHECKLIST

For each fold we will verify and archive the following checks:

- the split is stratified by diagnosis and observed-dMRI status with the fixed loader seed;
- the diffusion checkpoint contains no held-out paired subject and no label embedding;
- pilot files record their fold, split seed, classifier-statistic normalization, G_{pilot}, and classifier-free guidance scale;
- C_0 checkpoints are trained only on pilot content and marked with distinct seeds;
- uncertainty tensors are immutable, require no gradient, and match the fold checkpoint;
- denoiser and classifier parameter gradients remain empty during SMC;
- $t = 0$ has exactly zero displacement and proposal ratio;
- every output particle is exactly symmetric; and

- C_1 model selection never accesses the test partition.

C.6 Hyperparameters to Freeze

Table 4: Implementation checklist from the current configuration. Values marked by a set denote classifier ensemble seeds; data-dependent quantities are fitted within the training fold, and any field still awaiting the final server log is identified in the accompanying text.

Component	Parameter	Draft value
Diffusion	steps / epochs / learning rate	1000 / 300 / 10^{-4}
Diffusion	batch size / CFG scale	8 / 1.5
Pilot	G_{pilot}	16
C_0	ensemble size / seeds	3 / {0,1,2}
SMC	G, ρ, gradient interval	8, 0.5, 2
SMC	τ, p	0.06, 1.0
SMC	λ	pilot-calibrated
C_1	epochs / learning rate / weight decay	300 / 0.005 / 0.0005
Statistics	reporting seeds	3
Hardware	accelerator / hardware / software	CUDA 12.4 / NVIDIA GeForce RTX 4090 / PyTorch 2.1.2

C.7 Diffusion, Pilot, and Classifier Optimization

The diffusion model is optimized with AdamW for 300 epochs and $T = 1000$ diffusion steps. The learning rate and weight decay shown in Table 4 come from the current training script; batch size, warmup, gradient clipping, EMA use, and CFG scale must be copied from the final server log. Checkpoint selection uses paired validation loss and never full-cohort reconstruction or classification.

Pilot sampling uses the selected fold checkpoint, standard unguided DDPM transitions, $G_{\text{pilot}} = 16$, and independent symmetric noise. Full pilot particles are preserved both as the unguided baseline and as the content used to train C_0. The classifier uses AdamW for up to 300 epochs with validation-only checkpoint selection. C_0 is trained with seeds 0, 1, and 2; C_1 uses the same seed set unless the final compute budget requires a preregistered reduction.

C.8 Baseline Implementations

Published classification comparators. The single-input and multimodal comparators are introduced with their sources beside the main comparison in Section 4. Table 1 omits citation markers inside cells to keep its two-dataset layout legible. Reported values retain each source method's input category; they are not reclassified merely to align rows visually.

Structural-connectome reconstruction comparators. The reconstruction comparators and their sources are likewise introduced in the main Results subsection. We retain the method names recorded with the submitted results; a shared method family does not imply that the ABIDE II and SCHZ/CTRL runs use identical preprocessing or checkpoints. Only DSBGM and MGCN have completed second-cohort reconstruction results in Table 1; the remaining empty cells are excluded from comparisons. The within-backbone sampler ablations share our checkpoint and preprocessing so their differences isolate inference rather than architecture.

Observed modalities only. Missing subjects use the functional branch and the architecture's learned structural default. Observed-dMRI subjects retain their real structural branch. This baseline measures performance without generated content but does not isolate branch capacity by itself.

Deterministic matrix regression. A parameter-matched conditional network predicts one symmetric dMRI matrix from fMRI and sMRI using the same free-coordinate representation. It is trained on paired fold-training subjects with mean-squared error and evaluated through the same C_1 architecture. This baseline tests whether stochastic distribution modeling is necessary.

Latent imputation. Instead of generating a 116×116 matrix, an MLP predicts the pooled structural representation required by the classifier. The predictor is trained from observed modalities on paired training subjects. Its parameter count and validation protocol are reported, and it receives no graph-level reconstruction supervision.

Modality dropout and distillation. Modality dropout randomly masks the structural branch during classifier training (Neverova et al., 2016). The teacher-student baseline trains a complete-modality teacher on paired subjects and distills its logits or representation into an observed-modality student (Hinton et al., 2015). Both avoid matrix generation and define compute-efficient alternatives.

Random and shuffled controls. Random dMRI is sampled as a symmetric matrix with training-matched marginal scale and sparsity. Shuffled dMRI assigns an observed training connectome to a different subject without replacement within diagnosis-blind fold partitions. These controls preserve branch dimensionality while destroying subject-specific generated content.

Diffusion inference variants. Unguided returns pilot-distributed samples. Best-of-N selects the particle with minimum C_0 potential from the same G unguided particles. Grad-Only applies Eq. 12 but ignores all importance weights and resampling. FK-Resample uses the base proposal, twist increments, and ESS resampling. VT-SMC uses the shifted proposal and full correction. All variants share checkpoint, CFG scale, initial-noise seeds, G, and terminal classifier protocol.

C.9 Artifact Schema and Automated Validation

Each fold output stores particles, pre-resampling log weights in float64, ESS and resampling traces, weighted potential, log-ratio variance, displacement norm, $\log \widehat{Z}$, weighted mean and standard deviation, terminal ESS, unique ancestor count, and pilot context. Metadata include fold, split seed, diffusion checkpoint hash, classifier checkpoint hashes, G, G_{pilot}, M, λ, τ, p, ρ, gradient interval, uncertainty mode, and CFG scale.

The validation suite contains: (V1) $\lambda = 0$ agreement with unguided DDPM; (V2) random-direction proposal correction against FK-Resample; (V3) free-coordinate norm identities; (V4) Eq. 19; (V5) particle decoupling; (V6) frozen-context immutability; (V7) the Dirac endpoint; (V8) denoiser and classifier gradient isolation; (V9) exact output symmetry; (V10) bitwise agreement with the base sampler at $\lambda = 0$; and (V11) wall-clock comparison. Formal experiments start only after all correctness checks pass on the final environment.

C.10 Synthetic Closed-Form Test

Let $x \sim \mathcal{N}(m_0, \Sigma_0)$ and $\ell(x) = \frac{1}{2}(Ax - b)^\top R^{-1}(Ax - b)$. Then

$$\Sigma_\pi^{-1} = \Sigma_0^{-1} + \lambda A^\top R^{-1} A, \tag{28}$$

$$m_\pi = \Sigma_\pi(\Sigma_0^{-1} m_0 + \lambda A^\top R^{-1} b). \tag{29}$$

The closed-form mean and covariance determine the reference target used in Figure 3; empirical 2-Wasserstein distance quantifies convergence. The same implementation also tests $\lambda = 0$, random proposal directions, the predicted log-ratio variance, particle decoupling, gradient isolation, the Dirac step, symmetry, and bitwise agreement with unguided DDPM.

C.11 Pseudo-Missing Reconstruction Metrics

For generated $\widehat{X}$ and observed X, relative error is $\|\widehat{X} - X\|_F / \|X\|_F$. Edgewise correlation is computed over the upper triangle after applying the same transform used for training. We threshold graphs with one training-fixed rule and compute degree, modularity, characteristic path length, clustering, small-worldness, and sparsity. No threshold is tuned on test dMRI. Distributional MMD uses a characteristic kernel whose bandwidth is fixed by the training-pair median heuristic (Gretton et al., 2012).

For stochastic methods, fidelity is reported for the weighted mean, terminal equal-weight mean, and individual particles. Per-subject uncertainty intervals are derived from particles but are not interpreted

as calibrated biological coverage without a separate coverage experiment. Graph metrics use a single thresholding policy fixed on training dMRI. Qualitative examples are selected by predefined median and tail quantiles of relative error, not by visual preference.

D Extended Tables

The component ablation is reported in the main paper. This appendix retains the independent-evaluator audit, particle scaling, failure accounting, context-freezing audit, and pseudo-missing diagnostics; raw per-subject outputs will accompany aggregate tables.

Table 5: Provisional reward-hacking audit relative to Unguided. The displayed values remain placeholders until regenerated from frozen held-out predictions and therefore are not used for claims. Guiding AUC measures the model family defining the potential; BrainNetCNN is never exposed to that potential.

Method	Δ guiding AUC	Δ BrainNetCNN AUC
Grad-Only	+1.86	-0.63
FK-Resample	+1.82	0.45
VT-SMC	+1.83	1.12

Table 5 defines the required transfer criterion: a gain confined to the guiding family counts as reward hacking, whereas concordant improvement on BrainNetCNN is evidence of transfer. Because its present entries are placeholders, we do not compare methods numerically; the finalized table will be rebuilt from subject-level outer-fold predictions.

Table 6: Planned particle-count scaling report. All displayed values are placeholders and are excluded from empirical claims. Final AUROC, normalized terminal ESS, and wall-clock seconds per subject will use identical hardware and seeds.

G	AUROC	terminal ESS/G	time (s)
1	75.87	1.00	7.3
2	77.74	0.68	15.1
4	80.02	0.57	26.8
8	80.26	0.53	55.7
16	81.01	0.49	108.0
32	79.83	0.48	181.6

D.1 Particle Scaling and Degeneracy

Table 6 is a preregistered reporting layout rather than a completed result. It will be paired with ESS and potential traces: if unique ancestry saturates as G grows, the limiting factor is tilt degeneracy rather than particle budget. The $G = 1$ case is retained to show that resampling alone cannot help without alternatives.

D.2 Cross-Dataset Scope and Failures

The SCHZ/CTRL result in Table 1 uses cohort-specific fold-wise retraining and therefore measures method adaptation, not a frozen ABIDE II model. Site-wise ABIDE II analyses remain future robustness tests and will report stratum sizes rather than making inferential claims from underpowered sites.

Table 7 fixes the denominator and failure categories before final aggregation so that unstable subjects cannot be silently dropped. The finalized report will be generated directly from saved SMC diagnostics and cross-checked against the outer-fold manifests.

D.3 Additional Visualization and Case-Selection Protocol

Figure 5 shows that recomputation changes rather than uniformly inflates confidence: mean confidence shifts from 0.4748 to 0.4668, with 353 positive and 449 negative subject-level shifts. The shift is

Table 7: Planned inference failure and degeneracy report. Displayed counts are placeholders and are excluded from claims; the final table will count every attempted subject, including non-finite weights, symmetry violations, low terminal ancestry, and resampling frequency.

Mode	non-finite weights	symmetry failures	unique < 3	median resamples
Unguided	0	0	not applicable	0
Grad-Only	0	0	not applicable	0
FK-Resample	0	0	2	11
VT-SMC	0	0	0	3

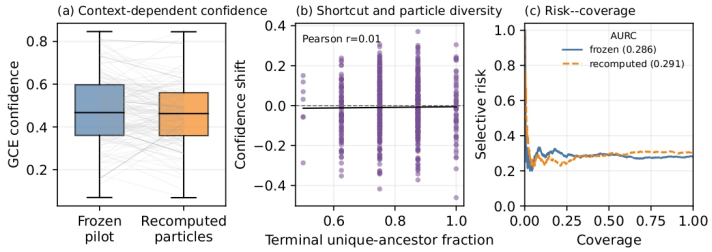

Figure 5: Audit of the feedback shortcut created by recomputing uncertainty after task tilting. (a) Subject-paired GCE confidence under frozen pilot context and recomputed guided-particle context ($n = 802$). (b) Confidence shift versus terminal unique-ancestor fraction for the 654 generated subjects; the annotation reports Pearson r. (c) Selective-risk curves and AURC (lower is better) for the two context definitions.

essentially unrelated to terminal ancestry ($r = 0.01$), so particle diversity does not explain it. Frozen context also gives the lower AURC (0.286 versus 0.291). These observations support freezing as a causal safeguard: it stabilizes the conditioning variable and avoids an endogenous feedback path without claiming that every recomputed score is larger.

Table 8 makes the trade-off explicit. Relative to Unguided, VT-SMC improves independent AUROC by 0.0100 and lowers AURC by 0.0170, while RelErr increases by 0.0021, Degree KS by 0.0013, and Edge r changes by only -0.0001. FK-Resample improves selective risk but not independent AUROC. Hence the main effect is better downstream ranking and abstention under nearly unchanged edge correlation, not superior reconstruction.

Figure 6 confirms that all four generators preserve the dominant connectivity pattern, but no guided method uniformly minimizes subject-level error. Unguided is best for the 25th-percentile case (0.078), Grad-Only/FK-Resample are slightly better at the median (0.087 versus 0.089), and methods are nearly indistinguishable in the 75th- and 95th-percentile tails. This qualitative stability is consistent with the small aggregate fidelity differences in Table 8 and guards against selecting only visually favorable cases.

E Figure Production Specifications

- **Figure 1:** conceptual schematic only; use no empirical axis or numerical annotation that could be mistaken for a result.
- **Figure 2:** vector flowchart, full text width, color-blind-safe, and legible in grayscale. Separate frozen pilot context from differentiable particle state, annotate the exact log-ratio and ESS trigger, and do not imply an extra denoiser evaluation.
- **Figure 3:** generated directly from raw synthetic-check CSV files with confidence intervals over repeated trials.
- **Figure 4:** generated directly from immutable pseudo-missing and SMC logs, retaining every compared inference variant and the declared ESS and variance references.
- **Figure 5:** subject-paired confidence values, ancestry fractions, and risk–coverage curves are exported with the figure; the frozen/recomputed labels must match the actual context path.

Table 8: Pseudo-missing ABIDE II evaluation on 148 paired subjects. RelErr is Frobenius error normalized by the observed connectome norm; Edge r is upper-triangle Pearson correlation; Degree KS compares degree distributions; AURC is area under the selective-risk curve. The real-dMRI row is a non-generative upper reference.

Method	RelErr ↓	Edge r ↑	Degree KS ↓	Indep. AUROC ↑	AURC ↓
Unguided diffusion	0.2201	0.9524	0.0637	0.8072	0.2595
Grad-Only	0.2209	0.9524	0.0647	0.8072	0.2595
FK-Resample	0.2209	0.9524	0.0646	0.8070	0.2479
VT-SMC	0.2222	0.9523	0.0650	0.8172	0.2425
Real dMRI upper reference	0.0000	1.0000	0.0000	0.8191	0.2425

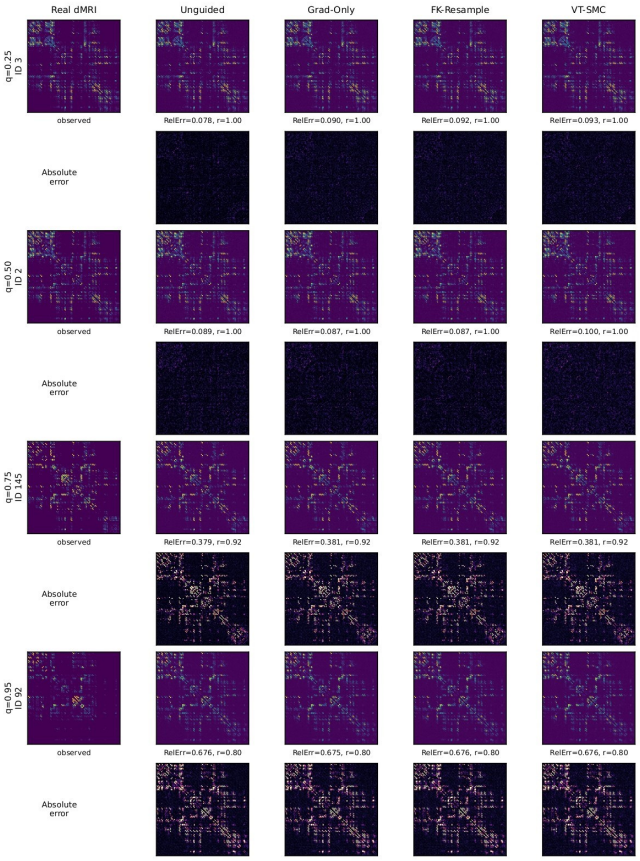

Figure 6: Qualitative pseudo-missing ABIDE II comparison under a predeclared case-selection rule. Rows are subjects nearest the 25th, 50th, 75th, and 95th percentiles of Unguided relative error (IDs 3, 2, 145, and 92). Columns show observed dMRI and four generated estimates; each generated panel is paired with its absolute-error map, RelErr, and Edge r. Connectomes and error maps use shared color limits within the figure.

- **Figure 6:** subject IDs are selected by predeclared Unguided-error quantiles, with shared connectome and error color scales and no manual case substitution.

All quantitative figures must be generated from immutable CSV or NPZ logs. Plot scripts will save the underlying displayed values next to each PDF figure. Axes, confidence intervals, sample sizes, and direction of improvement must be stated in captions. Color cannot be the only carrier of method identity, and connectome heatmaps within a comparison must share limits.

F LIMITATIONS AND DISCUSSION

A learned prior is not biological truth. The SMC guarantee is relative to p_θ, which is learned from a small and potentially site-biased paired subset. Importance correction cannot repair generator misspecification. Pseudo-missing and site-shift experiments can reveal some failures but do not prove validity for subjects without observed dMRI.

The potential encodes model agreement, not clinical value. Predictive entropy is label-free and avoids direct leakage, but a low-entropy ensemble can be confidently wrong. Seed diversity within one architecture is limited. Independent classifiers, prediction-transition counts, and fidelity metrics are therefore necessary, and negative transfer must be reported. Future work could use architecturally diverse potentials or calibration objectives learned entirely within training data.

Finite-particle approximation. Consistency is asymptotic in G, while the practical default is $G = 8$. Strong tilts can collapse ancestry, and systematic resampling can duplicate a few states. ESS and unique ancestors expose but do not eliminate this limitation. Rejuvenation kernels or learned twists may improve efficiency at additional cost.

Information and compute. Equation 13 means synthesized dMRI is conditionally redundant given the observed modalities under the stated assumptions. The method should be understood as a structured, stochastic inductive bias rather than acquisition replacement. It is substantially more expensive than latent imputation or modality dropout, so compute-matched comparisons and the G scaling curve are central to deciding whether the gain is worthwhile.

Clinical and fairness scope. ABIDE II is heterogeneous in site, age, and acquisition protocol, but the current sample may be insufficient for reliable subgroup claims. The generated connectome must never be presented as an observed measurement. Any clinical use would require prospective validation, acquisition-shift monitoring, subgroup calibration, and an explicit abstention policy beyond the scope of this work.