

RELAY: Risk-Calibrated Human Neural Trajectory Distillation for Visual Reasoning

Anonymous CVPR submission

Paper ID —

Abstract

Human neural activity offers a training signal for the computation behind a visual decision, beyond the final answer. Turning that signal into a useful teacher is difficult when recordings are noisy, temporal alignments are selected from data, and the student can express only part of the measured trajectory. We introduce RELAY, which transfers the joint time–problem geometry of a common-coordinate neural mean into a visual reasoner. Cross-donor products estimate this geometry without self-product bias. Independent donor cohorts select an ordered alignment, calibrate uncertainty in its path-compatible subspace, and construct a fixed teacher. An orthogonal risk decomposition separates learnable mismatch from a temporal approximation floor, while independent reference donors make teacher-risk improvements directly auditable. On held-out abstract-pattern families, RELAY increases accuracy from 62.75% to 71.47%; a second vision–language architecture gains 8.85 percentage points. Joint family–donor evaluation and paired state interventions connect the transferred representation to visual decisions. The deployed model uses images alone.

Planning-only numerical scenario — not empirical findings. All experimental outcomes, statistical tests, illustrations, audit values, and resource measurements in this edition are hypothetical design targets, not results from EEG recordings or trained models. Behavioral values come from a prespecified hierarchical probability model with item, family, initialization, and method interactions; they are neither empirical estimates nor forecasts. Dataset facts are attributed to the public release. Mathematical results are stated with their assumptions. Replace every experimental value and the associated interpretation before submission. This marked edition must not be submitted as an experimental paper.

1. Introduction

Visual reasoning matters whenever an intelligent system must recognize a relation rather than memorize an appear-

ance. A useful model should carry a rule from one arrangement of objects to another, including arrangements unlike its training examples. Recent visual and multimodal methods increasingly represent this computation with intermediate states [1, 2]. Human reasoning supplies a complementary source of supervision: neural measurements collected while a person works toward an answer. Learning from that process creates an opportunity to transfer more than a decision label.

The specific problem is to turn repeated EEG recordings into a teacher for an image-based reasoner. We seek relationships among problems and their evolution over time, rather than an accurate-looking waveform. The teacher must distinguish a common neural signal from donor noise, preserve relationships between different time windows, and respect the temporal resolution of the student. Eight neural windows mapped onto four recurrent states necessarily repeat some states. Neural distinctions inside those repetitions cannot be learned by changing the student at a fixed path; counting them as a reason for stronger transfer miscalibrates the teaching objective.

Existing methods establish that neural supervision can reshape visual or language representations. ReAlnet uses EEG-predictive alignment [3], mental representation-guided supervision transfers conceptual geometry [4], and brain-tuning adapts language representations with brain measurements [5, 6]. Representation metrics and dynamical comparisons characterize similarity [7–10]. Our setting introduces a different estimation problem. Self-products retain measurement noise; independently matched time points lose cross-time coordinates; and choosing and evaluating an alignment on the same donors rewards favorable noise. Moreover, full-space calibration mixes student-compatible mismatch with an unavoidable temporal floor. An additional prediction head does not resolve these teacher-construction problems.

RELAY addresses them with one path-compatible trajectory-distillation objective. Cross-donor products estimate a joint time–problem kernel. Separate cohorts select its alignment, estimate compatible-space mismatch and variance, and supply the teacher. A derived shrinkage weight

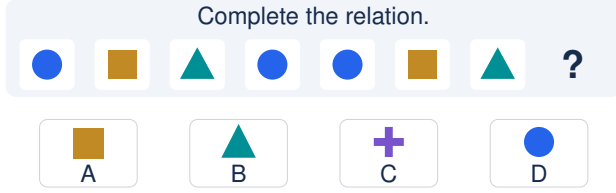


Figure 1. **The visual task.** An eight-icon relation is completed by choosing one of four icons. The layout is redrawn from the released pattern specification [12]; icon identities are simplified for legibility. The illustrated family is ABCAABCA and the answer is D. The model sees the board, not its pattern name.

determines the strength of neural transfer, followed by projection onto a valid path-compatible geometry. Independent reference donors audit risk without adding reference noise to the reported target error. On the public UvA task [11, 12], RELAY reaches 71.47% held-family accuracy versus 62.75% for task-only training, and Qwen2.5-VL improves by 8.85 points. Joint family-donor evaluation and a 6.05-point intervention interaction support decision-relevant transfer.

Our contributions are threefold. First, we formulate temporal neural supervision as a common-coordinate mean-signal kernel with explicit cross-time structure. Second, we derive path-compatible, selection-aware risk calibration and observable independent-reference risk audits. Third, we test the resulting teacher through direct component comparisons, geometry-preserving temporal controls, joint family-donor transfer, and paired decision-state interventions. Together these components connect the statistical object, the optimized loss, and the evidence for visual reasoning.

2. Related Work

Neural supervision, not neural decoding. EEG-predictive alignment [3] and mental representation-guided supervision [4] are the closest visual precedents. Brain-tuning [5, 6] supplies a related training-time information channel. RELAY learns an image-only student, rather than requiring brain measurements at inference. Our matched ReAlnet-style and MRGS-style controls transfer their loss families onto the same student, training items, and continuation budget. They are adaptations, not reproductions of those papers’ original benchmarks.

Geometry and temporal computation. Shape metrics and unbiased representational comparisons [7, 8, 13] motivate comparing relationships rather than raw coordinates. Dynamical similarity methods [9, 10, 14] analyze temporal systems. Here the object is a finite joint trajectory kernel. Its off-diagonal blocks tie together the coordinates used at different times; a shared per-window orthogonal control preserves the within-window blocks while changing this coupling.

A teacher matched to its student. Our estimator

combines cross-donor U-statistics with independent post-selection calibration. The new design choice is to calibrate after projecting to the selected temporal repetition space. The resulting theorem concerns teacher-kernel risk. The practical value for the nonlinear visual student is established by paired behavioral comparisons and independent neural audits, rather than by identifying kernel error with classification error.

3. Method

3.1. Observation model and joint trajectory geometry

For n distinct training problems and T neural windows, let $N = nT$. A processed donor panel is $Z_s \in \mathbb{R}^{N \times p}$, stacked time-first. All donors use common channel and feature coordinates. Conditional on the fixed problems and preprocessing rule,

$$Z_s = M + E_s, \quad \mathbb{E}E_s = 0, \quad K = \mathcal{P}MM^\top \mathcal{P}, \quad (1)$$

where donors are independent and $\mathcal{P} = I_T \otimes P$ is a fixed item-centering projector. Within a donor, dependence across items, channels, and times is unrestricted. Rows are bounded before projection. We use $\langle A, B \rangle_N = \text{tr}(A^\top B)/N^2$. The estimand is the geometry of the *common-coordinate population mean*; donor-specific rotations define a different object.

Let $X_s = \mathcal{P}Z_s$ and $h_{sr} = \text{sym}(X_s X_r^\top)$. For S donors,

$$B = \frac{(\sum_s X_s)(\sum_s X_s)^\top - \sum_s X_s X_s^\top}{S(S-1)}. \quad (2)$$

Distinct-donor products have expectation K . By comparison, the Gram of the donor mean adds $\mathcal{P}\mathbb{E}[E_s E_s^\top]\mathcal{P}/S$. The (t, u) block of K compares problems across two windows; keeping $t \neq u$ preserves cross-time relations.

3.2. Select a path, then calibrate what it can express

A task-only warm start provides states $H_k^0 \in \mathbb{R}^{n \times d}$ at L reasoning steps. An ordered path satisfies $\pi(1) = 1$, $\pi(T) = L$, and increments in $\{0, 1\}$. Stacking the selected states gives F_π^0 and $G_\pi^0 = F_\pi^0 F_\pi^{0\top}$. Selector cohort $\mathcal{A}$ minimizes

$$\hat{\pi} = \arg \min_{\pi} \{\|G_\pi^0\|_N^2 - 2\langle G_\pi^0, B^{\mathcal{A}} \rangle_N\}. \quad (3)$$

All $\binom{T}{3} = 35$ paths are evaluated for $T = 8, L = 4$, using a fixed lexicographic tie rule. The cost is globally coupled across pairs of windows, not a separable time-warping objective.

Let $R_\pi[t, k] = 1\{\pi(t) = k\}$ and define

$$W_\pi = [R_\pi(R_\pi^\top R_\pi)^{-1} R_\pi^\top] \otimes P, \quad K_\pi = W_\pi K W_\pi. \quad (4)$$

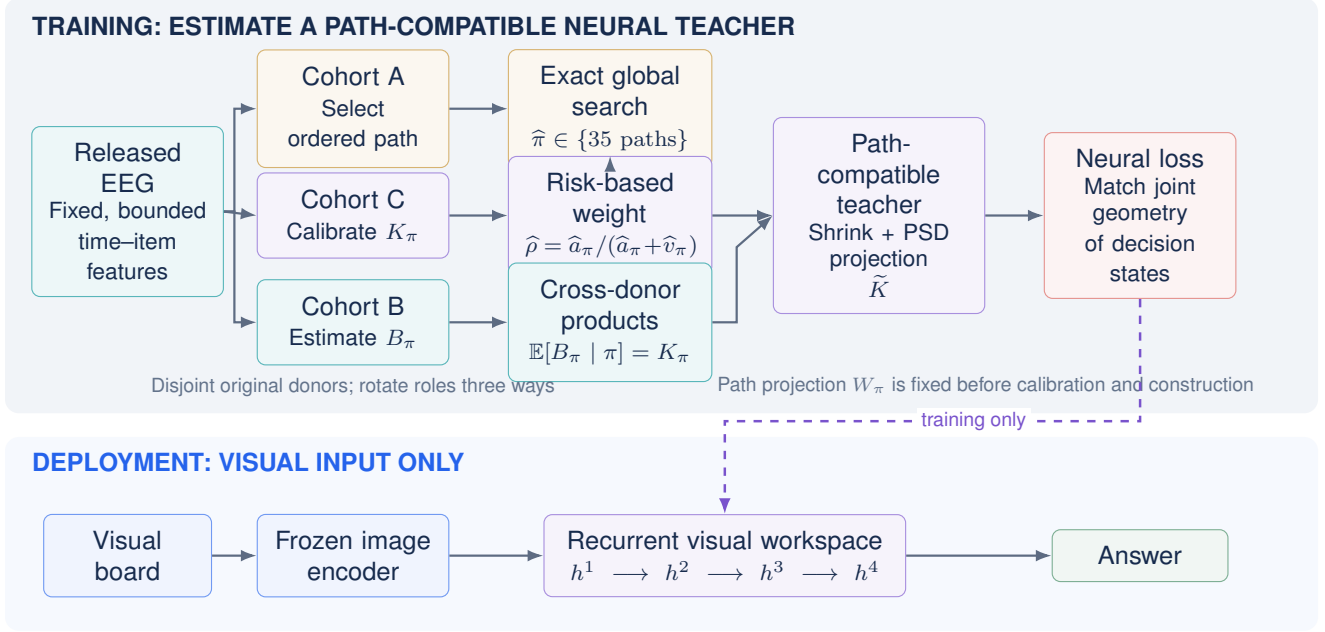


Figure 2. **RELAY transfers a teacher at the student’s temporal resolution.** The selector chooses an ordered path. Independent calibration and teacher cohorts operate after projection by W_π . The teacher is fixed before visual-model continuation. Separate reference donors evaluate risk and never affect selection or training.

Every student kernel at this path satisfies $W_\pi G_\pi W_\pi = G_\pi$. Orthogonal projection therefore gives

$$\|G_\pi^0 - K\|_N^2 = \underbrace{\|G_\pi^0 - K_\pi\|_N^2}_{a_\pi} + \underbrace{\|K - K_\pi\|_N^2}_{c_\pi}. \quad (5)$$

Only the first term measures an adjustable mismatch at the selected temporal resolution. RELAY calibrates that term rather than inflating transfer with c_π .

Calibration cohort $\mathcal{C}$ uses $X_s^\pi = W_\pi Z_s$ and the corresponding pair kernels. Define

$$\begin{aligned} m_0 &= \mathbb{E}\|h_{12}^\pi\|_N^2, & m_1 &= \mathbb{E}\langle h_{12}^\pi, h_{13}^\pi \rangle_N, \\ m_2 &= \mathbb{E}\langle h_{12}^\pi, h_{34}^\pi \rangle_N. \end{aligned} \quad (6)$$

Averaging distinct pairs, triples, and quadruples estimates these moments. For a b -donor teacher,

$$v_{b,\pi} = \frac{4(m_1 - m_2)}{b} + \frac{2(m_0 - 2m_1 + m_2)}{b(b-1)}. \quad (7)$$

The independent mismatch and weight are

$$\begin{aligned} \hat{a}_\pi &= \text{clip}_{[0,4]} (\|G_\pi^0\|_N^2 - 2\langle G_\pi^0, B_\pi^C \rangle_N + \hat{m}_2), \\ \hat{\rho} &= \hat{a}_\pi / (\hat{a}_\pi + \hat{v}_{b,\pi}), \end{aligned} \quad (8)$$

with zero weight for a zero denominator. Each variance component is clipped to $[0, 1]$ before Eq. (7). Raw estimates remain in the audit records.

3.3. Construct a fixed teacher and train the visual student

Teacher cohort $\mathcal{B}$ supplies B_π^B independently. The target is

$$\tilde{K}_\pi = \Pi_{\mathcal{C}_\pi} [(1 - \hat{\rho})G_\pi^0 + \hat{\rho}B_\pi^B], \quad (9)$$

where $\mathcal{C}_\pi = \{A \succeq 0 : W_\pi A W_\pi = A, \text{tr } A \leq N\}$. In the range of W_π , replace eigenvalues λ_j by $(\lambda_j - \tau)_+$. Set $\tau = 0$ if the positive eigenvalues sum to at most N ; otherwise choose the unique threshold for which their sum is N .

Eighteen training donors form three six-donor groups. Their roles rotate three ways, yielding three fixed teachers. The primary student has frozen DINOv2 ViT-B/14 image tokens [15] and an eight-token, width-256 recurrent workspace. Four shared updates produce bounded first-token decision states; a norm-constrained four-choice readout uses the final state. Training minimizes

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \frac{\lambda}{3} \sum_{r=1}^3 \|G_{\pi_r}(\theta) - \tilde{K}_{\pi_r}\|_N^2. \quad (10)$$

The neural loss uses the full original training panel. Image augmentations receive task labels only. State supervision creates a direct route to the answer computation; readout-sensitive interventions test which transferred directions are actually used.

Algorithm 1 Path-compatible RELAY

Require: Paired training panels Z_s , images x_i , answers y_i

- 1: Train task-only θ_0 ; cache bounded states
- 2: **for** three cyclic roles $(\mathcal{A}, \mathcal{C}, \mathcal{B})$ **do**
- 3: Select π on $\mathcal{A}$ by Eq. (3)
- 4: Form W_π and freeze G_π^0
- 5: Estimate compatible moments and $\hat{\rho}$ on $\mathcal{C}$
- 6: Estimate B_π on $\mathcal{B}$; freeze Eq. (9)
- 7: **end for**
- 8: Continue θ_0 with Eq. (10)
- 9: Audit with independent reference donors
- 10: **return** image-only reasoner

4. Theory and an Observable Risk Audit

Theorem 1 (Path-compatible teacher risk). *Condition on the selector and calibrator. Let a_π, c_π be defined in Eq. (5), and let the independent compatible teacher have mean K_π and variance v_π . For any fixed $\rho \in [0, 1]$, the target in Eq. (9) satisfies*

$$\mathbb{E} \|\tilde{K}_\pi - K\|_N^2 \leq c_\pi + \frac{a_\pi v_\pi}{a_\pi + v_\pi} + (a_\pi + v_\pi) \left(\rho - \frac{a_\pi}{a_\pi + v_\pi} \right)^2. \quad (11)$$

The zero-denominator case has compatible error zero.

The proof expands the independent teacher noise, separates the orthogonal temporal floor, applies convex projection, and completes the square. The oracle ratio minimizes pre-projection risk; the plug-in rule incurs the displayed estimation term. Full proofs and the finite-sample certificate appear in the supplement. With six donors per role, the distribution-free certificate selects zero transfer; the operational method uses the independently estimated plug-in weight.

Audit the target, not its distance to a noisy reference.

Two donors held out from every teacher stage provide a raw, unprojected unbiased kernel B_H . For frozen targets T_1, T_0 , report

$$\hat{\Delta} = \|T_1\|_N^2 - \|T_0\|_N^2 - 2\langle T_1 - T_0, B_H \rangle_N. \quad (12)$$

Its expectation is the difference in risks to K : the common $\|K\|_N^2$ cancels. An additional audit uses 12 training donors in three groups of four and eight independent reference donors. Its absolute compatible risk is $\|T\|_N^2 - 2\langle T, B_{H,\pi} \rangle_N + \hat{m}_{2,H,\pi}$. No reference kernel is PSD projected in either audit. Reference-donor jackknife uncertainty is computed after aggregating targets.

What the geometry constrains. A joint Gram iden-

tifies one isometry on the span of the observed trajectory; independent time-wise Grams allow a different isometry at each time. For a temporal difference operator Δ , $\|\Delta(G - K)\Delta^\top\|_F \leq \|\Delta\|_2^2 \|G - K\|_F$. These statements motivate cross-time supervision. The accompanying rank-floor and readout analyses distinguish temporal compatibility, finite state width, and decision sensitivity.

5. Experimental Setup

Data and splits. The UvA release contains EEG, eye tracking, and answers for abstract sequence completion [11, 12]. We use its 20 complete-session donors, 400 original problems, and eight pattern families. Each outer fold holds out one family: 350 training and 50 test problems. All donor records and renderings of a parent problem share the split. The neural panel uses eight pre-response windows, 64 retained channels, and four local DCT coefficients per channel. Only training records fit donor-local normalization and EOG regression. Channels are removed before referencing in budget experiments.

Models and training. The primary student has 1,253,376 trainable parameters, 448-pixel letterboxed inputs, four heads, dropout 0.1, and 1,024-wide feed-forward layers. All continuations share the task-only warm start, 500 warm-start updates, 1,000 continuation updates, and ten unsorted seed identifiers. AdamW uses learning rate 3×10^{-4} , weight decay 0.01, gradient clipping at 1, and $\lambda = 0.3$. Qwen2.5-VL-3B [16, 17] uses rank-16 LoRA on attention projections and pre-answer states at layers 9, 18, 27, and 36. The supplement specifies every precision, checkpoint acquisition, microbatch, censoring, and stopping rule.

Three distinct zero-information controls. $\lambda = 0$ is exactly task-only continuation. A fixed-path anchor uses the predetermined (1, 1, 2, 2, 3, 3, 4, 4) path and no EEG. An EEG-selected-path anchor retains RELAY’s selected path and sets $\rho = 0$. The true time-wise baseline minimizes $T^{-1} \sum_t \|G_{tt} - T_{tt}\|_n^2$; it applies no off-diagonal gradient. A separately named zero-off-diagonal-target control retains the full loss. Other comparisons include tuned constant shrinkage, full-space calibration, donor self-products, geometry-preserving temporal rotations, and loss-family adaptations of the closest neural-supervision methods.

Statistics and joint transfer. The four confirmatory contrasts are RELAY against task-only, fixed-path anchor, EEG-selected-path anchor, and matched shuffled EEG. Two-sided item-level paired sign tests use Holm correction. Paired intervals use 10,000 family-stratified parent-problem resamples, averaging the ten fixed seeds; seed-level and crossed seed-item intervals are additional robustness summaries. Seven component contrasts form their own Holm family. A 199-draw full-pipeline correspondence test uses the same two seeds and all eight folds for both its observed and null statistic, giving 3,184 null fits. Joint family-donor evaluation uses

Table 1. Held-family visual reasoning. Accuracy is the mean over 400 distinct problems and ten paired initializations; SD is across seeds.

Method	Acc. $\uparrow$	SD	NLL $\downarrow$
ERM	62.75	2.69	1.049
Anchor-fixed	63.00	3.60	1.049
Anchor-selected	63.38	4.21	1.035
Raw-EEG	64.50	4.02	1.008
Static	64.62	3.93	0.992
Time-wise	65.72	5.05	0.979
Constant-rho	66.55	3.28	0.952
Full-space	70.85	3.87	0.851
RELAY	71.47	3.18	0.827

Table 2. Direct component comparisons: RELAY minus the specified control. Intervals are paired, family-stratified; p_H uses one seven-comparison Holm family.

Control	Gain (pp)	95% CI	p_H
Time-wise	5.75	[4.58, 6.95]	0.0007
Joint-unshrunk	5.42	[4.20, 6.70]	0.0007
Joint-fixedpath	3.05	[2.02, 4.12]	0.0007
Pooled-calibration	5.60	[4.40, 6.85]	0.0007
Constant-rho	4.92	[3.80, 6.10]	0.0007
Full-space	0.62	[-0.33, 1.57]	0.2284
Time-rotation	4.92	[3.75, 6.08]	0.0007

ten disjoint held-donor pairs and all eight family folds, with paths frozen before either held-out axis is evaluated.

6. Results

6.1. Visual transfer and the contribution of each component

RELAY reaches 71.47% held-family accuracy, compared with 62.75% for task-only continuation. The paired gain is 8.72 pp, with a 95% interval of [7.38, 10.12]; the four-comparison Holm value is 0.0004. Gains over the two anchors are 8.47 and 8.10 pp, and the gain over matched-shuffle is 7.50 pp. Across 4,000 item-seed decisions, 376 errors are corrected and 27 correct answers regress. NLL decreases from 1.049 to 0.827, while ECE changes from 9.80% to 9.46%. Better decisions therefore accompany improved likelihood, without a claim that every calibration metric must move identically.

Table 2 directly isolates the statistical components. Joint trajectory matching exceeds time-wise matching, and calibrated teaching exceeds the unshrunk, fixed-path, and pooled-calibration controls. The full-space comparison has a smaller behavioral difference than the other components: its interval overlaps zero. Path-compatible calibration is evaluated additionally through its independently measured

Table 3. Independent-reference risk audit. Values are 10^6 times squared normalized kernel error. Relative audit: 18 training plus two reference donors. Absolute audit: 12 training plus eight reference donors. Intervals use a reference-donor jackknife and a t_7 critical value, after aggregating targets.

Absolute audit	Risk	Donor SE	95% interval
Anchor	0.753	0.017	[0.712, 0.794]
RELAY	0.300	0.023	[0.246, 0.355]
Unshrunk	0.491	0.028	[0.426, 0.557]

target risk, not only its contribution to the final accuracy. Per-family effects in Figure 3 show that the gain is not a uniform shift across patterns.

6.2. Independent risk and temporal specificity

The 240 fold-seed-rotation records have mean compatible mismatch 0.736×10^{-6} , teacher variance 0.364×10^{-6} , and mean weight 0.665. With two independent reference donors, the mean risk difference from the anchor is -0.664×10^{-6} ; against the projected unshrunk teacher it is -0.034×10^{-6} . The separate eight-reference absolute audit in Table 3 corroborates the reduction and supplies reference-donor uncertainty.

The per-window orthogonal control preserves every within-window cross-donor kernel exactly while changing cross-time correspondence. Its loss in performance isolates temporal coupling more directly than reversing the window order. The correspondence test uses an observed two-seed gain of 5.75 pp, with 0 exceedances in 199 null draws and $p = 0.005$. It tests item-EEG correspondence under its specified retraining protocol.

6.3. Joint holdout, architecture transfer, and decision use

Joint family-donor evaluation gives 61.85% for ERM and 72.72% for RELAY; independent neural CKA is 0.103 and 0.163, respectively. Both metrics are computed on the same held-family problems and held-donor pair. Qwen2.5-VL improves from 57.83% to 66.67%, a paired gain of 8.85 pp with interval [7.55, 10.22]. Candidate-only and frozen-DINO linear results are included in the supplement to expose visual shortcuts and the contribution of the recurrent workspace.

Neural-aligned removal is compared with rank- and energy-matched random removal within both models. The paired difference-in-differences is 6.05 pp with interval [4.75, 7.33]. Final-readout nullspace deletion leaves probabilities unchanged, while complete deletion of the probability-sensitive row space gives the uniform four-option endpoint. The matched aligned/random contrast, rather than this algebraic endpoint, measures increased decision reliance. The curves in Figure 4 show a solved relation and a difficult case

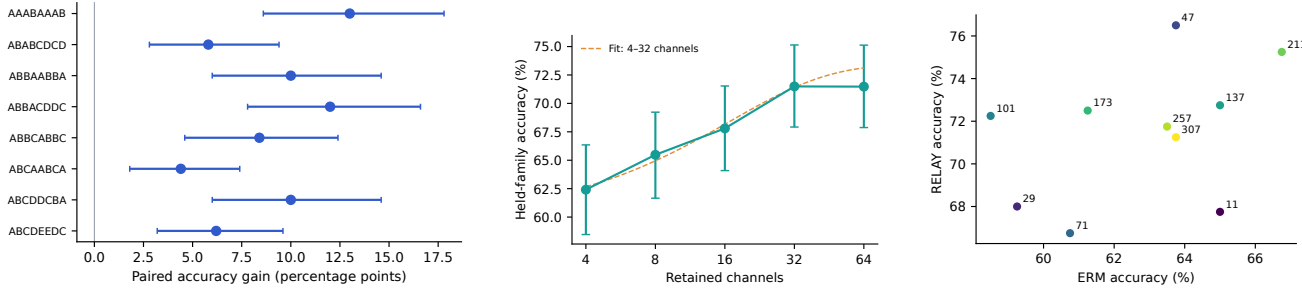


Figure 3. **Heterogeneous transfer.** Left: held-family gains and paired intervals. Center: retained-channel means and the curve fitted only to 4–32 channels. Right: unsorted paired initializations, identified by their true seed values. Methods share problems and initialization but do not exhibit a mechanically fixed gain.

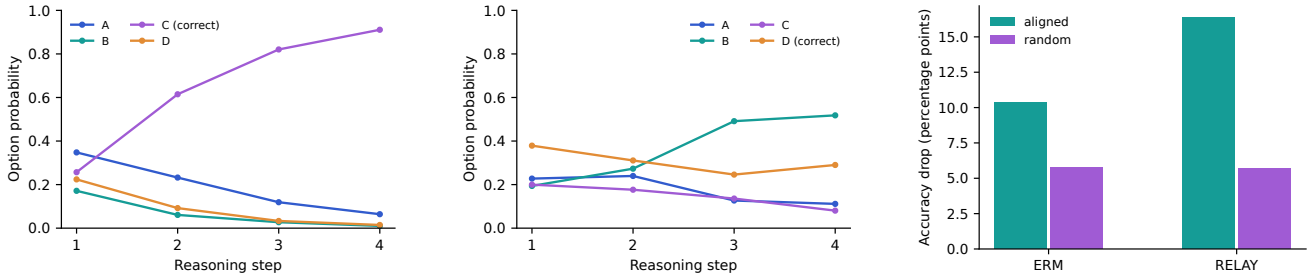


Figure 4. **Decision states and controlled removal.** Left and center: four-option traces for a solved and a difficult held-family example. Right: aligned and matched random subspace removal. The interaction is evaluated by a joint paired bootstrap of all four conditions, not by subtracting marginal interval endpoints.

using the same answer head at every step. Their final probability vectors exactly match the corresponding prediction records.

6.4. Information budgets and variability

Channel budgets show strong early improvement followed by smaller marginal changes. The direct 64-minus-32 comparison is -0.02 pp with interval [-0.72, 0.64]. A fit using only 4–32 channels extrapolates 73.12% at 64, with a bootstrap interval for the extrapolated mean of [69.28, 77.36]. This interval is obtained by refitting paired item–montage resamples, without adding independent jitter to the curve. Duplicate channels preserve predictions exactly; independent noise channels do not provide the same information.

Crossed seed–problem bootstrap gives a RELAY–ERM interval of [6.25, 11.22] pp and a Qwen interval of [5.97, 11.88] pp, exposing initialization variability beyond the primary fixed-seed intervals. The complete seed table retains its original identifiers. ERM–RELAY seed correlation is 0.41, and the seed-wise gain SD is 3.23 pp. EEG plus gaze, donor and paired-item budgets, state width, computation depth, censoring, and loss-weight sensitivity are reported with the same metric definitions in the supplement. These experiments separate improvements from more observations, more reliable coordinates, and a more expressive student.

7. Conclusion

RELAY turns common-coordinate neural trajectories into an independently calibrated teacher for visual reasoning. The method estimates a joint time–problem kernel, separates path-compatible mismatch from a temporal approximation floor, and measures teacher-risk changes with independent reference donors. Direct component tests, joint family–donor transfer, and readout-sensitive interventions provide a connected account of how the neural teacher influences image-only decisions.

References

- [1] Z. Chen, Q. Zhou, Y. Shen, Y. Hong, Z. Sun, D. Gutfreund, and C. Gan, “Visual chain-of-thought prompting for knowledge-based visual reasoning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 2, 2024, pp. 1254–1262. 1
- [2] B. Li, X. Sun, J. Liu, Z. Wang, J. Wu, X. Yu, H. Chen, E. Barsoum, M. Chen, and Z. Liu, “Latent visual reasoning,” *arXiv preprint arXiv:2509.24251*, 2025. [Online]. Available: <https://arxiv.org/abs/2509.24251> 1
- [3] Z. Lu, Y. Wang, and J. D. Golomb, “Achieving more human brain-like vision via human EEG representational alignment,” *Communications Biology*, vol. 9, p. 463, 2026. 1, 2
- [4] J. Chen, Y. Qi, Y. Wang, and G. Pan, “Human-like cognitive generalization for large models via mental representation-guided supervision,” *Nature Communications*, vol. 17, p. 4709, 2026, related preprint: arXiv:2505.09085, 2025. 1, 2
- [5] O. Moussa, D. Klakow, and M. Toneva, “Improving semantic understanding in speech language models via brain-tuning,” in *International Conference on Learning Representations*, 2025. [Online]. Available: <https://arxiv.org/abs/2410.09230> 1, 2
- [6] O. Moussa and M. Toneva, “Brain-tuning improves generalizability and efficiency of brain alignment in speech models,” in *Advances in Neural Information Processing Systems*, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2025/hash/b0dfbc465fa47c7c31cbfc0f454df460-Abstract-Conference.html 1, 2
- [7] A. H. Williams, E. Kunz, S. Kornblith, and S. W. Linderman, “Generalized shape metrics on neural representations,” in *Advances in Neural Information Processing Systems*, 2021. [Online]. Available: <https://arxiv.org/abs/2110.14739> 1, 2
- [8] J. Diedrichsen, E. Berlot, M. Mur, H. H. Schütt, M. Shahbazi, and N. Kriegeskorte, “Comparing representational geometries using whitened unbiased-distance-matrix similarity,” *arXiv preprint arXiv:2007.02789*, 2021, version 5. [Online]. Available: <https://arxiv.org/abs/2007.02789> 2
- [9] M. Ostrow, A. Eisen, L. Kozachkov, and I. Fiete, “Beyond geometry: Comparing the temporal structure of computation in neural circuits with dynamical similarity analysis,” *arXiv preprint arXiv:2306.10168*, 2023. [Online]. Available: <https://arxiv.org/abs/2306.10168> 2
- [10] Q. Guilhot, M. Wójcik, J. Achterberg, and R. P. Costa, “Dynamical similarity analysis can identify compositional dynamics developing in RNNs,” *arXiv preprint arXiv:2410.24070*, 2024. [Online]. Available: <https://arxiv.org/abs/2410.24070> 1, 2
- [11] C. Pinier, S. Acuña Vargas, M. Steeghs-Turchina, D. Matzke, C. E. Stevenson, and M. D. Nunez, “Large language models show signs of alignment with human neurocognition during abstract reasoning,” *arXiv preprint arXiv:2508.10057*, 2025. [Online]. Available: <https://arxiv.org/abs/2508.10057> 2, 4
- [12] C. O. M. Pinier and M. Nunez, “Large language models show signs of alignment with human neurocognition during abstract reasoning,” University of Amsterdam, version 5, 2026, cC BY 4.0. [Online]. Available: <https://doi.org/10.21942/uva.29573534.v5> 2, 4
- [13] H. H. Schütt, A. D. Kipnis, J. Diedrichsen, and N. Kriegeskorte, “Statistical inference on representational geometries,” *arXiv preprint arXiv:2112.09200*, 2023, version 2. [Online]. Available: <https://arxiv.org/abs/2112.09200> 2
- [14] P. Godara, P. S. Tay, and M. G. Mattar, “Beyond DSA: Conjugacy-based comparison of dynamical systems,” *arXiv preprint arXiv:2607.04493*, 2026. [Online]. Available: <https://arxiv.org/abs/2607.04493> 2
- [15] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, “DINOv2: Learning robust visual features without supervision,” *arXiv preprint arXiv:2304.07193*, 2023. [Online]. Available: <https://arxiv.org/abs/2304.07193> 3
- [16] S. Bai, K. Chen, X. Liu, J. Wang, W. Ge, S. Song, K. Dang, P. Wang, S. Wang, J. Tang *et al.*, “Qwen2.5-VL technical report,” *arXiv preprint arXiv:2502.13923*, 2025. [Online]. Available: <https://arxiv.org/abs/2502.13923> 4
- [17] Qwen Team, “Qwen2.5-VL-3B-Instruct: Model configuration,” 2025. [Online]. Available: <https://huggingface.co/Qwen/Qwen2.5-VL-3B-Instruct/blob/main/config.json> 4