# RETINAPLEX: A LARGE-SCALE BIOLOGICALLY CONSTRAINED RETINAL MODEL FOR PARALLEL VISUAL CODING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Visual perception depends on sensory populations that represent ongoing scenes while retaining information about recent changes. How retinal diversity organizes these complementary signals for downstream computation remains incompletely understood. Here we introduce RetinaPlex, a biologically constrained retinal model comprising approximately ten million modeled retinal units and supporting 4K visual input. The model combines spatial organization constrained by human photoreceptor statistics and informed by primate retinal connectivity with physiologically inspired temporal dynamics. Across controlled stimuli and natural-video experiments, sustained and transient bipolar populations support low-capacity readout of current scene state and recent visual history from a single population snapshot, with retinal parameters held fixed. Distinct histories remain distinguishable after input sequences converge to identical continuing visual input. Model interventions show that retained-response differences shape early state–history geometry, whereas adaptation timescales modulate the relative contribution of history across pathways over time. Reanalysis of public primate retinal recordings reveals a concordant association between independently measured temporal response phenotypes and relative current-versus-recent coding bias. Comparisons with frame-like and event-like interfaces further characterize access to both information sources. Together, these findings establish RetinaPlex as a platform for investigating retinal population coding and suggest biologically grounded principles for sensory representations that jointly preserve scene state and recent change in NeuroAI and neuromorphic vision.

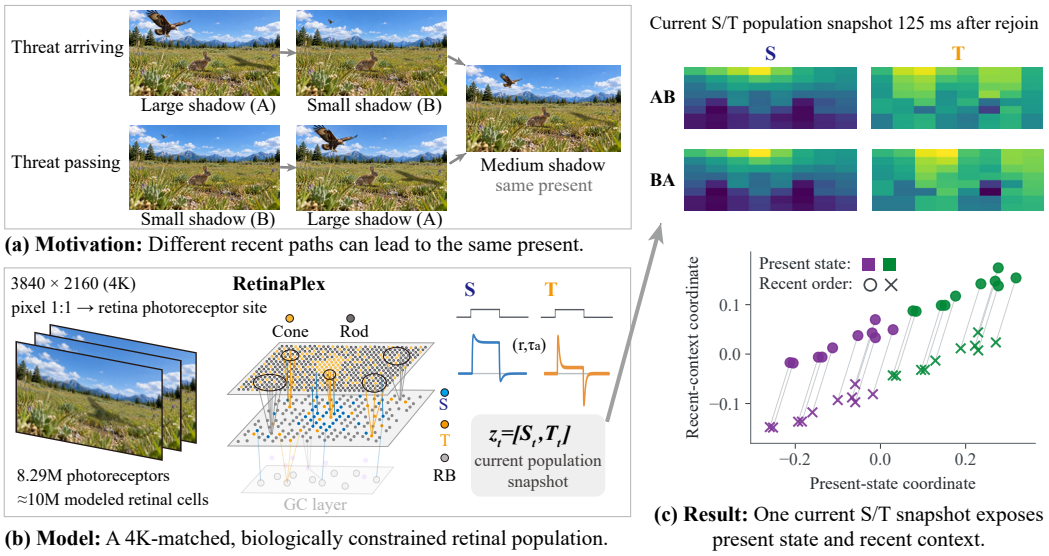
## 1 INTRODUCTION

A fundamental problem in sensory coding is how to transform continuous input into representations that support perception and action. The computational usefulness of a representation depends not only on the information it preserves, but also on how that information is organized for downstream access (Rigotti et al., 2013). In a changing environment, understanding the present can require information about the recent past: identical current scenes can follow different events, or the same events occurring in different orders. Both neural populations and artificial recurrent models can retain preceding inputs within their evolving states (Nikolić et al., 2009; Gu et al., 2020). This raises a question shared by neuroscience and artificial intelligence: how do diverse response dynamics organize information about the present and recent past within a single population state, and how does this organization shape accessibility to downstream computation?

The retina provides a biologically grounded setting for investigating this question. Parallel pathways transform photoreceptor signals through spatial pooling, response polarity, and diverse temporal sensitivities. Sustained and transient responses are prominent features of retinal processing, with differences already evident in bipolar cells and further shaped by retinal circuit interactions (Awatramani & Slaughter, 2000; Franke et al., 2017). Recordings from primate retinal ganglion cells also demonstrate that responses to a new image depend on the image that preceded it (Krüppel et al., 2023b). These observations connect retinal response diversity to visual processing over time. Understanding its computational significance requires explaining how ongoing visual information



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.




**(a) Motivation:** Different recent paths can lead to the same present.

**(b) Model:** A 4K-matched, biologically constrained retinal population.

**(c) Result:** One current S/T snapshot exposes present state and recent context.

Figure 1: **One sensory present, different internal histories.** **(a)** Conceptual illustration. The same current visual state can be reached through different recent paths and carry different contextual meanings; dynamic vision therefore requires recent history to be retained in the current state. **(b)** RetinaPlex maps continuous 4K video pixel by pixel onto a biologically constrained photoreceptor array. Sustained (S) bipolar cells use single-cone input, transient (T) bipolar cells use multicone convergence, and rod bipolar (RB) cells provide a parallel rod pathway; downstream extensions including amacrine cells and ganglion cells (GC) are also retained. The retained-response fraction $r$ and adaptation time constant $\tau_a$ control the key temporal dynamics, and S/T jointly form the current population state $z_t$. **(c)** At 125 ms after the history paths converge, the S/T population snapshots still retain differences. Current state and recent context are separated along distinct readout directions, showing that a single current snapshot provides simultaneous access to both types of information.

and traces of preceding events are distributed across parallel populations, and which properties of their dynamics govern this organization.

Here we introduce **RetinaPlex**, a biologically constrained retinal model comprising approximately ten million modeled retinal units and processing native 4K video. Its spatial construction incorporates human photoreceptor statistics and connectivity motifs from human and nonhuman primate retina, including single-cone sampling and local multicone convergence (Cooper et al., 2024; Zhang et al., 2020; Tsukamoto & Omi, 2016). The broader platform supports explicit ON/OFF polarity together with sustained/transient dynamics in four cone-driven channels: ON-sustained, OFF-sustained, ON-transient, and OFF-transient. The present study focuses on the cone-driven sustained and transient bipolar populations, with horizontal-cell-related local gain modulation and physiologically inspired effective dynamics. This spatially explicit construction connects high-resolution visual input to interpretable pathway responses and permits controlled analysis of both full-field activity and compact population representations.

We characterize RetinaPlex using controlled stimuli and natural-video backgrounds, with retinal parameters fixed before the coding experiments. In a fork–rejoin assay, different recent histories converge to exactly the same continuing visual input, while a current-state variable is manipulated independently. Low-capacity readouts receive only the current S/T population representation, testing whether ongoing visual information and preceding events remain jointly accessible after the inputs become identical. We then match the event set, final event, and all subsequent input, testing whether the current representation retains information about the order of earlier events. This additional control distinguishes access to earlier temporal structure from sensitivity to the identity of the most recent event.

To examine how this coding is organized, we independently manipulate retained-response fractions and adaptation timescales while keeping spatial construction fixed. Spatial-budget controls assess

whether comparable readout performance can be maintained with fewer population output sites. Cross-event prediction and targeted representation interventions further examine the structure and accessibility of history traces. A secondary analysis of public primate retinal recordings relates independently measured temporal response phenotypes to current-versus-recent coding bias, while comparisons with frame-like, event-like, and hybrid interfaces place the model representation in the context of artificial visual sensing.

Our main contributions are as follows:

1. **We build a connectomics-informed retinal model comprising approximately ten million units that processes native 4K video.** RetinaPlex integrates human photoreceptor statistics, primate retinal connectivity motifs, and physiologically inspired dynamics into four ON/OFF sustained/transient pathways. This provides a biologically constrained platform for studying how retinal populations encode high-resolution natural visual input.
2. **We show that a single retinal population snapshot supports simultaneous readout of the present and recent past.** With retinal parameters fixed, simple readouts recover current state and recent history on held-out natural-video backgrounds. Earlier event order remains readable even when the final event and all subsequent input are identical, demonstrating history access beyond a trace of the last event.
3. **We show that retained-response differences separate present and history signals, while adaptation redistributes history across pathways over time.** Equalizing S/T retained-response fractions reduces the median state–history angle from 56.8° to 13.7°. Equalizing adaptation times instead largely removes the subsequent shift in relative history contributions, revealing distinct roles for these two temporal properties.

Overall, RetinaPlex connects biologically constrained retinal modeling to the organization of visual information across time. Its current population state supports access to ongoing visual input and traces of earlier events, while controlled interventions distinguish the roles of response persistence and adaptation. For NeuroAI, these findings motivate sensory front ends that make recent temporal context available to simple downstream computations. For neuromorphic sensing and computing, the combination of compact spatial sampling and parallel temporal responses suggests a design direction for representing persistent scene information together with recent change and event history.

## 2 Related Work

**Biologically constrained retinal models.** Virtual Retina and COREM provide biologically motivated frameworks for simulating retinal processing (Wohrer & Kornprobst, 2009; Martínez-Cañada et al., 2016). Learned retinal models connect natural-scene response prediction to circuit interpretation (Maheswaranathan et al., 2023), while openretina supports standardized modeling across datasets and species (D'Agostino et al., 2026). RetinaPlex complements these approaches by combining human photoreceptor statistics, primate connectivity motifs and effective sustained–transient dynamics in approximately ten million modeled units processing native 4K video. This spatially explicit construction provides a common setting for examining visual responses, population coding and controlled parameter interventions.

**Present and past information in population activity.** Retinal adaptation and responses to image transitions establish that visual encoding depends on preceding stimulation (Hosoya et al., 2005; Krishnamoorthy et al., 2017; Krüppel et al., 2023b). Nikolić et al. (2009) demonstrated that current visual cortical activity can simultaneously support linear readout of current and preceding stimuli. Studies of mixed selectivity, sensory–memory representations and retinal information geometry further connect population organization to downstream access and stimulus discriminability (Rigotti et al., 2013; Libby & Buschman, 2021; Ding et al., 2023). Building on these findings, RetinaPlex examines how S/T dynamics organize current and historical evidence after different histories converge to identical continuing input. Additional controls match the final event and subsequent input, testing whether the current population state retains the order of earlier events.

**Temporal diversity and memory.** Bipolar-cell physiology and synaptic studies identify several sources of temporal diversification in retinal pathways (Awatramani & Slaughter, 2000; Baden et al., 2013; Franke et al., 2017; Kuo et al., 2025). Reservoir computing and memory-capacity

theory explain how dynamical states make past inputs available to simple readouts (Maass et al., 2002; Ganguli et al., 2008; Dambre et al., 2012). Neural heterogeneity and structured state-space models extend the design of temporal representations (Perez-Nieves et al., 2021; Gu et al., 2020; 2022). RetinaPlex studies these principles within a fixed retinal architecture. Parameter interventions distinguish how retained-response differences shape early state–history separation and how adaptation-time differences alter subsequent history weighting across pathways. Event-conditioned displacement prediction further tests whether shared temporal gains transfer across stimulus patterns.

# 3 RETINAPLEX: MODEL CONSTRUCTION AND REPRESENTATION INTERFACE

## 3.1 BIOLOGICAL CONSTRAINTS AND SPATIAL CONSTRUCTION

RetinaPlex maps each pixel of a continuous $3840 \times 2160$ video to a photoreceptor site within a fixed spatial population (Figure 1b). The model contains 8,294,400 photoreceptor sites; together with the associated bipolar sampling centres, the modeled population comprises approximately ten million retinal units. Photoreceptor composition, local connectivity, and temporal parameters are specified separately, allowing their contributions to be examined through controlled interventions.

**Photoreceptor composition and physical scale.** Let $e(\mathbf{x})$ denote retinal eccentricity at position $\mathbf{x}$. The cone-density prior $\rho_C(e)$ is informed by human adaptive-optics measurements and anatomical data (Cooper et al., 2024; Masri et al., 2020), whereas the rod-density prior $\rho_R(e)$ follows human rod-pathway measurements (Lee et al., 2019). The local cone proportion and lattice spacing $p$ satisfy

$$\pi_C(e) = \frac{\rho_C(e)}{\rho_C(e) + \rho_R(e)}, \qquad p^{-2} = \mathrm{mean}_{\mathbf{x}\in\Omega_p} \left[\rho_C(e(\mathbf{x})) + \rho_R(e(\mathbf{x}))\right]. \tag{1}$$

Here $\Omega_p$ is the retinal field implied by $p$. Solving this relation gives $p \approx 2.646\,\mu\mathrm{m}$; fixed-seed local quota sampling then assigns rod/cone identities from $\pi_C$. Density anchors, treatment of peripheral priors, bipolar-centre placement, and complete population counts are provided in Supplementary Methods S1 and Table S1.

**Effective cone response and local surround modulation.** For scene-linear input luminance $I(\mathbf{x}, t)$, scaled in the experiments to $0 \le I(\mathbf{x}, t) \le 6$, we define a scaled cone response and a horizontal-cell-related surround signal:

$$\begin{aligned} R_C(\mathbf{x}, t) &= \mathrm{clip}\left(\frac{I(\mathbf{x}, t)}{6}, 0, 1\right), \qquad H_C(\mathbf{x}, t) = \frac{G_\sigma * (M_C R_C)}{G_\sigma * M_C + \varepsilon}(\mathbf{x}, t), \\ \widetilde{R}_C(\mathbf{x}, t) &= R_C(\mathbf{x}, t)\left[1 + \eta\left(R_C(\mathbf{x}, t) - H_C(\mathbf{x}, t)\right)\right]. \end{aligned} \tag{2}$$

Here $M_C$ is the fixed binary cone mask, $G_\sigma$ is a Gaussian spatial kernel, and $*$ denotes convolution. We use $\sigma = 100\,\mu\mathrm{m}$, $\eta = 0.10$, and $\varepsilon = 10^{-6}$. This effective gain modulation is motivated by surround interactions in primate cone photoreceptors (Verweij et al., 2003).

**Local cone-to-bipolar connectivity.** The sustained branch uses single-cone input inspired by human foveal private-line organization (Zhang et al., 2020). The transient branch pools six neighbouring cones with equal weights, following a coarse-grained connectivity choice informed by primate diffuse bipolar anatomy (Tsukamoto & Omi, 2015; 2016). Their native spatial drives are

$$x_{S,i}^{\mathrm{nat}}(t) = \widetilde{R}_{C,c(i)}(t), \qquad x_{T,j}^{\mathrm{nat}}(t) = \frac{1}{6} \sum_{k \in \mathcal{N}_6(j)} \widetilde{R}_{C,k}(t). \tag{3}$$

Here $c(i)$ is the cone assigned to sustained centre $i$, and $\mathcal{N}_6(j)$ contains the six cones connected to transient centre $j$. Connections are stored as fixed sparse indices.

The broader platform accommodates rod bipolar populations, ON/OFF polarity, and downstream extensions. The quantitative analyses presented here focus on the continuous cone-driven S/T bipolar core with horizontal-cell-related modulation.

### 3.2 Sustained and Transient Response Dynamics

Differences in response persistence and temporal filtering across bipolar cells motivate the effective S/T dynamics (Awatramani & Slaughter, 2000; Baden et al., 2013; Kuo et al., 2025). For branch $b \in \{S, T\}$, an input drive $\xi_b(t)$ acts on a fast-response state $u_b(t)$ and an adaptation state $a_b(t)$:

$$\tau_{f,b}\frac{du_b}{dt} = \xi_b - u_b, \qquad \tau_{a,b}\frac{da_b}{dt} = u_b - a_b, \qquad y_b = u_b - (1 - r_b)a_b. \tag{4}$$

The equations are applied componentwise, with parameters shared within each branch. Under constant input $\xi_b(t) = \bar{\xi}_b$,

$$u_b(t) \to \bar{\xi}_b, \qquad a_b(t) \to \bar{\xi}_b, \qquad y_b(t) \to r_b\bar{\xi}_b. \tag{5}$$

Thus, $r_b$ controls the steady response retained relative to the input drive, whereas $\tau_{a,b}$ controls the evolution of adaptation. Both branches use $\tau_{f,b} = 30\,\mathrm{ms}$. The sustained branch uses $r_S = 0.70$ and $\tau_{a,S} = 180\,\mathrm{ms}$; the transient branch uses $r_T = 0.10$ and $\tau_{a,T} = 100\,\mathrm{ms}$. These choices produce a stronger maintained response in S and a more prominent relative transient component in T.

The effective parameters were selected through preliminary response checks and fixed before the coding experiments. Prespecified interventions subsequently vary response retention or adaptation times while keeping the remaining model components fixed.

We integrate Equation 4 exactly under a zero-order-hold (ZOH) assumption at 120 Hz; each nominal 24-Hz input frame advances the model through five integration steps. Natural-video sequences are initialized at the steady state associated with their first-frame drive. The exact discrete-time update and initialization are given in Supplementary Methods S1.

### 3.3 Population Representation and Downstream Readouts

**Current population snapshot.** The main snapshot experiments use a fixed $8 \times 8$ full-field representation for each S/T branch. Each coordinate is count-normalized by the number of branch centres assigned to its spatial region, yielding 64 S-derived and 64 T-derived population coordinates. These coordinates summarize populations rather than individual cells.

At sampling time $t$, the two branch representations are concatenated as

$$\mathbf{z}_t = \begin{bmatrix}\mathbf{y}_S(t)\\ \mathbf{y}_T(t)\end{bmatrix} \in \mathbb{R}^{128}, \qquad \mathbf{y}_S(t), \mathbf{y}_T(t) \in \mathbb{R}^{64}. \tag{6}$$

The downstream readouts receive only $\mathbf{z}_t$. Previous frames and previous population snapshots are not supplied as additional readout inputs.

**Low-capacity readouts.** For task $v \in \{\mathrm{state}, \mathrm{history}\}$, let $\boldsymbol{\mu}_{v,c}^{\mathrm{DEV}}$ denote the mean representation for class $c \in \{0, 1\}$ in the designated fitting/development data. We define a normalized mean-difference direction and scalar score:

$$\mathbf{w}_v = \frac{\boldsymbol{\mu}_{v,1}^{\mathrm{DEV}} - \boldsymbol{\mu}_{v,0}^{\mathrm{DEV}}}{\left\|\boldsymbol{\mu}_{v,1}^{\mathrm{DEV}} - \boldsymbol{\mu}_{v,0}^{\mathrm{DEV}}\right\|_2}, \qquad s_v(t) = \mathbf{w}_v^\top \mathbf{z}_t. \tag{7}$$

Binary predictions use the midpoint of the projected development-set class means:

$$\beta_v = \frac{1}{2}\mathbf{w}_v^\top\left(\boldsymbol{\mu}_{v,1}^{\mathrm{DEV}} + \boldsymbol{\mu}_{v,0}^{\mathrm{DEV}}\right), \qquad \widehat{c}_v(t) = \mathbf{1}[s_v(t) > \beta_v]. \tag{8}$$

**Algorithm 1** RetinaPlex population snapshots and fixed readouts

**Input:** frames $\{I_n\}$; fixed model parameters; DEV readouts; frame interval $\Delta T$ and integration step $h$
Initialize the fixed retinal model and ZOH coefficients; set $K = \Delta T/h$
**for** each frame $I_n$ **do**
    Compute cone response, HC modulation, and S/T drives; initialize steady state on the first frame
    Advance the S/T states through $K$ exact ZOH steps
    If a designated sampling time $t$ is reached, form $\mathbf{z}_t$ and evaluate the frozen readouts
**end for**

Accuracy is computed from thresholded predictions, whereas AUROC is computed from continuous scores. Readout directions and thresholds are determined from fitting/development data and frozen for held evaluation. Experiments involving parameter interventions, readout refitting, or event-dependent prediction specify those operations separately. Display normalization is used only for visualization and does not enter quantitative readouts. Dataset partitions, controlled-event protocols, and evaluation procedures are detailed in Supplementary Methods S2 and Table S2.

# 4 RESULTS

## 4.1 A SINGLE RETINAL SNAPSHOT EXPOSES PRESENT STATE AND RECENT HISTORY

We first asked whether current retinal activity can support access to both present information and preceding events after their sensory inputs become identical. Our fork–rejoin design independently varied current contrast and preceding A/B event order, then presented exactly the same continuing input to matched trials. This separated internally retained history from differences in the current stimulus.

With retinal parameters fixed, simple readouts achieved current/recent accuracies of 1.00/1.00 on eight held procedural carriers and 0.875/0.800 on five held native 4K DAVIS sources (Pont-Tuset et al., 2017) at 125 ms after rejoin (Figures 1c and 2). During continued video input, current-state accuracy remained between 0.85 and 0.95, whereas recent-history accuracy declined to 0.650 at 208.3 ms. Thus, a fixed retinal front end made ongoing sensory evidence and recent context jointly available to simple downstream readouts from one current snapshot. Complete paired-carrier results and source-resolved natural-video performance are shown in Figure S1.

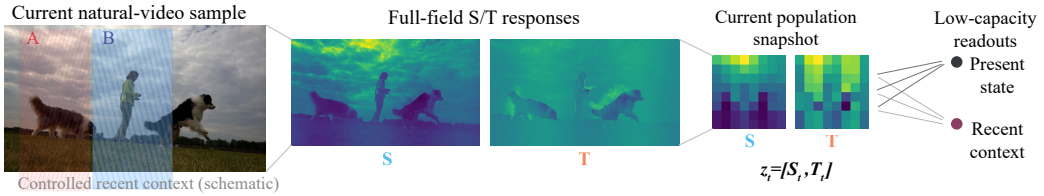
## 4.2 DISTINCT ROLES OF RESPONSE PERSISTENCE AND ADAPTATION

To identify which temporal differences organize this joint representation, we independently manipulated retained-response fractions and adaptation times while keeping spatial construction and input unchanged. Matched-current comparisons between natural histories and repeated-image controls isolated history-related responses across 64 probes from eight VisDrone scenes (Zhu et al., 2018).

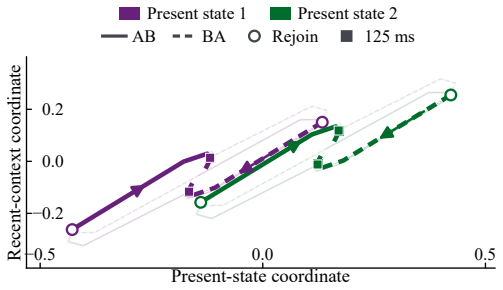
Equalizing retained-response fractions reduced the median angle between current and historical evidence in S/T space from $56.8^\circ$ to $13.7^\circ$. Equalizing adaptation times preserved this separation at $65.7^\circ$ (Figure 3a), but substantially weakened subsequent history reweighting: the relative T contribution declined from 0.68 to 0.41 in the baseline model, compared with 0.76 to 0.71 under equal adaptation times (Figure 3b). Continuous parameter scans supported these distinct effects (Figure S2). Retained-response diversity therefore shapes the separation of current and history-related evidence across the S/T branches, whereas adaptation controls their subsequent relative weighting over time. Complementary readout analyses show that the accessibility reduction associated with equalizing retained-response fractions remains visible on average under alternative linear readouts (Figure S7).

## 4.3 HISTORY ACCESS EXTENDS BEYOND THE FINAL EVENT

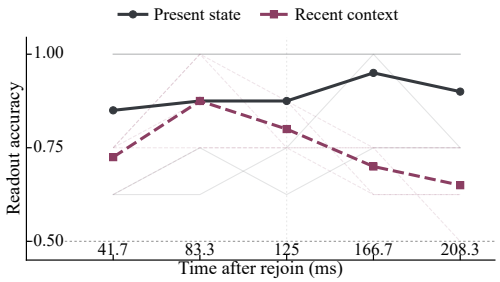
To test whether history decoding captured earlier sequence order, we matched the final event as well as the continuing input. Pairs such as ABC/BAC contained identical events and differed only in their earlier order. Shared temporal gains fitted on four event families were combined with known instantaneous spatial descriptors to predict population displacements for two held event families.




**(a) Pipeline:** From natural video to population readouts.




**(b) Dynamics:** Population trajectories after sensory rejoin.




**(c) Readout:** Concurrent accessibility over time.

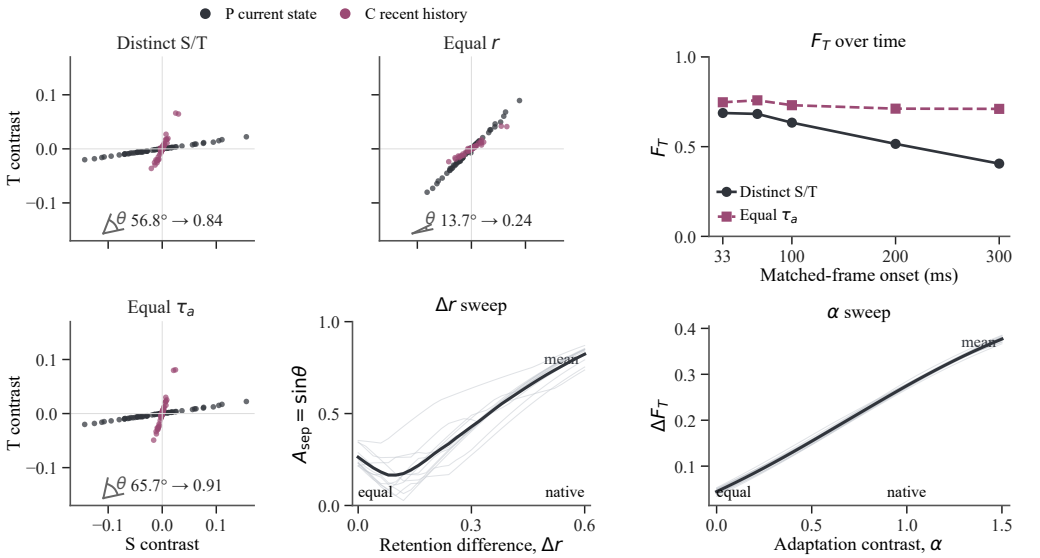
Figure 2: **Population dynamics preserve recent context after sensory rejoin. (a)** Natural video passes through RetinaPlex to produce full-field S/T responses, which are summarized into the current S/T population snapshot; lightweight readouts use only this snapshot to read out current state and recent context. Translucent A/B regions indicate experimentally manipulated recent history. **(b)** After sensory inputs rejoin, different histories still correspond to different population states and gradually approach one another along different trajectories under subsequent identical input. **(c)** Under continuing natural-video input, current state remains highly readable, while recent context remains readable early after rejoin and then gradually declines.

At 125 ms after common-suffix onset, earlier-order decoding achieved a macro AUROC of 0.683 across six held family–terminal conditions; all 60 prespecified pair evaluations yielded positive margins (Figure 4a; Figure S3). We separately tested representation sensitivity in the original binary assay. After readout refitting, history-targeted attenuation reduced history held balanced accuracy from 0.800 to 0.575, and current-targeted attenuation reduced current held balanced accuracy from 0.875 to 0.700. Both losses exceeded matched random perturbations (Figure 4b; Figure S4). Original non-target readout scores were protected by construction. Together, these tests show that retinal dynamics retain information beyond the final event and organize task-relevant population components whose accessibility remains sensitive to targeted perturbation.

### 4.4 Temporal Phenotypes Show Concordant Coding Biases in Model and Primate Retina

To assess the biological relevance of the model's coding organization, we reanalyzed public marmoset RGC recordings (Krüppel et al., 2023a;b). Sustained-like and transient-like groups were defined using independent full-field step responses, then evaluated by decoding current stimuli under fixed preceding conditions and preceding stimuli under fixed current conditions.

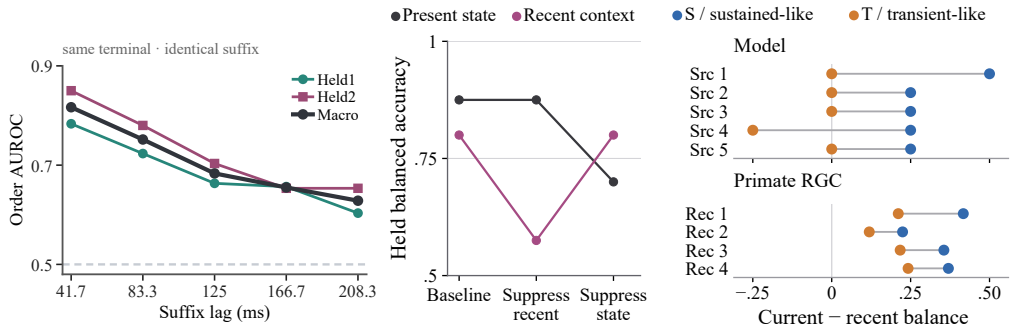
Across four recordings, sustained-like populations showed a consistently greater relative bias toward current information. The between-group difference in current-minus-history balanced accuracy ranged from 0.106 to 0.206 (Figure 4c; Figure S6). A corresponding secondary analysis of model S64/T64 representations produced a difference of 0.35, positive across all five videos and greater than every one of 1,000 spatially matched mixed-group controls. Both model branches also supported current-state decoding. This concordance links temporal response phenotype to relative coding preference across model populations and primate retinal output, providing biological support for the organization revealed by the model interventions.




(a) Retained-response diversity separates state and history.

(b) Adaptation time reweights recent history.

Figure 3: **Mechanisms of S/T population coding.** **(a)** Combinations of current-state contrast $P$ and recent-history contrast $C$ across the S/T branches. Equalizing the retained-response fraction $r$ reduces their separation, whereas increasing $\Delta r$ generally increases the separation $A_{\text{sep}} = \sin\theta$. **(b)** The adaptation time constant $\tau_a$ primarily regulates the relative weighting of history residues between the S/T branches. Equalizing $\tau_a$ weakens the temporal variation of $F_T$ (the relative T share of the history residual), whereas larger timescale differences produce stronger reweighting.




(a) **Generalization:** terminal-matched order remains accessible in held events.

(b) **Intervention:** targeted suppression is selective.

(c) **Correspondence:** coding balance agrees across systems.

Figure 4: **Generalization, selective intervention, and cross-level correspondence of population coding.** **(a)** After matching the terminal event and subsequent input, earlier event order remains decodable and shows a consistent temporal trajectory across two held event families. **(b)** Targeted suppression of recent-context or current-state directions selectively reduces the corresponding readout while leaving the other largely preserved. **(c)** Model and primate RGC data both show a coding-balance relation in which sustained-like populations are relatively biased toward current information and transient-like populations toward recent information; the two sample sets are independent.

### 4.5 Compact Retinal Sampling Supports Joint Current–History Readout

To connect these coding properties to computational interfaces, we compared frame, event, hybrid, and temporal-filter representations using the same 128-dimensional output interface and readout protocol (Gallego et al., 2022; Scheerlinck et al., 2019; Rebecq et al., 2019). RetinaPlex achieved current/recent AUROCs of 0.9975/0.9300, compared with 0.9950/0.5950 for the specified Two-

Table 1: **Access to current and recent information under matched output dimensionality.** The primary time point is 125 ms after the start of the common suffix. All scores are AUROC, and $J = \min(\mathrm{AUROC}_{\mathrm{current}}, \mathrm{AUROC}_{\mathrm{recent}})$. Retinal Two-IIR and RetinaPlex share the same retinal spatial front end.

| Representation | Current | Recent | $J$ |
|---|---|---|---|
| Current frame | 1.0000 | 0.5000 | 0.5000 |
| Adjacent-frame pair | 1.0000 | 0.5000 | 0.5000 |
| Event surface | 0.5000 | 1.0000 | 0.5000 |
| Explicit frame–event hybrid | 1.0000 | 1.0000 | 1.0000 |
| Raw Two-IIR | 0.9925 | 0.6450 | 0.6450 |
| Retinal Two-IIR (same spatial front end) | 0.9950 | 0.5950 | 0.5950 |
| RetinaPlex | 0.9975 | 0.9300 | 0.9300 |

IIR control sharing its spatial front end. The explicit frame–event hybrid achieved 1.0000/1.0000 (Table 1; Figure S5).

We also examined the spatial sampling budget before aggregation. The native layout used 943,616 S/T output sites and achieved joint accessibility—defined as the lower task AUROC—of 0.9300. A dense All-Cone layout used about 17.6 times as many sites and achieved 0.9275 (Table S3), with both representations reduced to 128 coordinates. The native construction therefore supported similar observed readability with substantially fewer spatial sampling sites. For NeuroAI, these results provide a concrete sensory-front-end design in which compact sampling and parallel temporal dynamics expose complementary current and recent information to low-capacity downstream computation.

## 5 Discussion and Conclusion

RetinaPlex integrates approximately ten million biologically constrained retinal units with native 4K input, linking large-scale retinal modeling to testable principles of population coding. The broader platform supports ON/OFF polarity together with sustained–transient pathways, while the present analyses focus on the cone-driven S/T population representation. From a single current snapshot, these dynamics make present state and recent history jointly accessible to simple downstream readouts. Parameter interventions distinguish response persistence, which primarily shapes state–history separation, from adaptation, which regulates how history is distributed across pathways over time. Terminal-matched sequences further show that accessibility extends beyond the identity of the most recent event, while targeted representation interventions demonstrate that current and recent information occupy selectively perturbable population coordinates. Together, these results show how biological temporal diversity can organize complementary information for downstream computation.

The model also provides a bridge between this representation-level mechanism and broader biological and engineering contexts. An independent reanalysis of primate retinal recordings revealed a concordant association between temporal response phenotype and relative current-versus-recent coding bias, linking the organization identified in the model to retinal output. At the same time, the native retinal layout achieved similar joint accessibility to a dense all-cone construction while using a substantially smaller pre-aggregation S/T output-site budget. These results position RetinaPlex as both a biologically grounded model of temporal population coding and a controlled platform for examining how spatial sampling and temporal dynamics shape accessible visual representations.

For NeuroAI and neuromorphic sensing, the central opportunity is to combine persistent state with sensitivity to change at the sensory front end. Event cameras report local brightness changes (Gallego et al., 2022), whereas RetinaPlex maintains a stateful population representation in which both current and recent information are accessible. In our representation-level comparison, RetinaPlex placed both variables in the high-accessibility regime, approaching the joint readability of an explicit frame–event hybrid while retaining a single biologically constrained stateful representation. This motivates sensory architectures that couple event-like responses with sustained internal states, potentially preserving scene information during periods of limited motion while maintaining recent temporal context for lightweight downstream computation. Physical implementations will be important for establishing the resulting bandwidth, energy, latency, and communication trade-offs.

### AI USE STATEMENT

AI tools were used as editorial aids for manuscript refinement, including language revision and improvement of written presentation, and for assistance with literature search and discovery. The authors independently reviewed the resulting text and relevant references and retain full responsibility for the scientific content, interpretation, conclusions, and accuracy of the manuscript.

### REPRODUCIBILITY STATEMENT

The main text and supplementary material document the model construction, temporal dynamics, population representation, experimental protocols, and evaluation procedures used in this study. Supplementary Methods provide additional implementation details, including physical-scale calibration, spatial sampling and connectivity, exact temporal integration, initialization, dataset partitions, experimental units, and readout definitions. Supplementary figures and tables report source-level results, intervention controls, temporal-filter comparisons, biological analyses, and model-construction parameters needed to reproduce the reported experiments. All quantitative results are based on fixed experimental protocols and explicitly defined development and evaluation partitions.

## REFERENCES

Gautam B. Awatramani and Malcolm M. Slaughter. Origin of transient and sustained responses in ganglion cells of the retina. *Journal of Neuroscience*, 20(18):7087–7095, 2000. doi: 10.1523/JNEUROSCI.20-18-07087.2000.

Tom Baden, Philipp Berens, Matthias Bethge, and Thomas Euler. Spikes in mammalian bipolar cells support temporal layering of the inner retina. *Current Biology*, 23(1):48–52, 2013. doi: 10.1016/j.cub.2012.11.006.

Robert F. Cooper, Snega Kalaparambath, Geoffrey K. Aguirre, and Jessica I. W. Morgan. Morphology of the normative human cone photoreceptor mosaic and a publicly available adaptive optics montage repository. *Scientific Reports*, 14:23166, 2024. doi: 10.1038/s41598-024-74274-y.

Lisa J. Croner and Ehud Kaplan. Receptive fields of P and M ganglion cells across the primate retina. *Vision Research*, 35(1):7–24, 1995. doi: 10.1016/0042-6989(94)E0066-T.

Dennis M. Dacey and Michael R. Petersen. Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences*, 89 (20):9666–9670, 1992. doi: 10.1073/pnas.89.20.9666.

Federico D'Agostino, Thomas Zenkel, Baptiste Lorenzi, Michaela Vystrčilová, Dominic Gonschorek, Samuel Suhai, Samuele Virgili, Alexander S. Ecker, Olivier Marre, Larissa Höfling, Thomas Euler, and Matthias Bethge. The openretina project: Collaborative retina modelling across datasets and species. *eLife*, 2026. doi: 10.7554/eLife.111328.1. URL https://elifesciences.org/reviewed-preprints/111328. Reviewed Preprint, version 1.

Joni Dambre, David Verstraeten, Benjamin Schrauwen, and Serge Massar. Information processing capacity of dynamical systems. *Scientific Reports*, 2:514, 2012. doi: 10.1038/srep00514.

Xuehao Ding, Dongsoo Lee, Joshua Melander, George Sivulka, Surya Ganguli, and Stephen Baccus. Information geometry of the retinal representation manifold. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://papers.neurips.cc/paper_files/paper/2023/hash/8a267516a7a697965c6ae4f48b908605-Abstract-Conference.html.

Felice A. Dunn and Fred Rieke. Single-photon absorptions evoke synaptic depression in the retina to extend the operational range of rod vision. *Neuron*, 57(6):894–904, 2008. doi: 10.1016/j.neuron.2008.01.031.

Katrin Franke, Philipp Berens, Timm Schubert, Matthias Bethge, Thomas Euler, and Tom Baden. Inhibition decorrelates visual feature representations in the inner retina. *Nature*, 542(7642):439–444, 2017. doi: 10.1038/nature21394.

Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2022. doi: 10.1109/TPAMI.2020.3008413.

Surya Ganguli, Dongsung Huh, and Haim Sompolinsky. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48):18970–18975, 2008. doi: 10.1073/pnas.0804451105.

Albert Gu, Tri Dao, Stefano Ermon, Atri Rudra, and Christopher Ré. HiPPO: Recurrent memory with optimal polynomial projections. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1474–1487, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/102f0bb6efb3a6128a3c750dd16729be-Abstract.html.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022.

Toshihiko Hosoya, Stephen A. Baccus, and Markus Meister. Dynamic predictive coding by the retina. *Nature*, 436:71–77, 2005. doi: 10.1038/nature03689.

Vidhyasankar Krishnamoorthy, Michael Weick, and Tim Gollisch. Sensitivity to image recurrence across eye-movement-like image transitions through local serial inhibition in the retina. *eLife*, 6: e22431, 2017. doi: 10.7554/eLife.22431.

Steffen Krüppel, Dimokratis Karamanlis, Yunus C. Erol, Sören J. Zapp, and Tim Gollisch. Dataset – marmoset retinal ganglion cell responses to saccadic grating shifts and supporting stimuli. G-Node Open Data, 2023a. URL https://doi.org/10.12751/g-node.thlt1j.

Steffen Krüppel, Mohammad H. Khani, Dimokratis Karamanlis, Yunus C. Erol, Sören J. Zapp, Matthias Mietsch, Dario A. Protti, Fernando Rozenblit, and Tim Gollisch. Diversity of ganglion cell responses to saccade-like image shifts in the primate retina. *Journal of Neuroscience*, 43(29): 5319–5339, 2023b. doi: 10.1523/JNEUROSCI.1561-22.2023.

Sidney P. Kuo, Wan-Qing Yu, Prerna Srivastava, Haruhisa Okawa, Luca Della Santina, David M. Berson, Gautam B. Awatramani, Rachel O. L. Wong, and Fred Rieke. Cone bipolar cell synapses generate transient versus sustained signals in parallel ON pathways of the mouse retina. *eLife*, 13:RP98817, 2025. doi: 10.7554/eLife.98817.3.

Sammy C. S. Lee, Paul R. Martin, and Ulrike Grünert. Topography of neurons in the rod pathway of human retina. *Investigative Ophthalmology & Visual Science*, 60(8):2848–2859, 2019. doi: 10.1167/iovs.19-27217.

Alexandra Libby and Timothy J. Buschman. Rotational dynamics reduce interference between sensory and memory representations. *Nature Neuroscience*, 24(5):715–726, 2021. doi: 10.1038/s41593-021-00821-9.

Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002. doi: 10.1162/089976602760407955.

Niru Maheswaranathan, Lane T. McIntosh, Hidenori Tanaka, Satchel Grant, David B. Kastner, Joshua B. Melander, Aran Nayebi, Luke E. Brezovec, Julia H. Wang, Surya Ganguli, and Stephen A. Baccus. Interpreting the retinal neural code for natural scenes: From computations to neurons. *Neuron*, 111(17):2742–2755.e4, 2023. doi: 10.1016/j.neuron.2023.06.007.

Pablo Martínez-Cañada, Christian Morillas, Begoña Pino, Eduardo Ros, and Francisco Pelayo. A computational framework for realistic retina modeling. *International Journal of Neural Systems*, 26(7):1650030, 2016. doi: 10.1142/S0129065716500301.

Rania A. Masri, Ulrike Grünert, and Paul R. Martin. Analysis of parvocellular and magnocellular visual pathways in human retina. *Journal of Neuroscience*, 40(42):8132–8148, 2020. doi: 10.1523/JNEUROSCI.1671-20.2020.

Danko Nikolić, Stefan Häusler, Wolf Singer, and Wolfgang Maass. Distributed fading memory for stimulus properties in the primary visual cortex. *PLOS Biology*, 7(12):e1000260, 2009. doi: 10.1371/journal.pbio.1000260.

Nicolas Perez-Nieves, Vincent C. H. Leung, Pier Luigi Dragotti, and Dan F. M. Goodman. Neural heterogeneity promotes robust learning. *Nature Communications*, 12:5791, 2021. doi: 10.1038/s41467-021-26022-3.

Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbeláez, Alexander Sorkine-Hornung, and Luc Van Gool. The 2017 DAVIS challenge on video object segmentation. *arXiv:1704.00675*, 2017. URL https://arxiv.org/abs/1704.00675.

Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. Events-to-video: Bringing modern computer vision to event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3852–3861, 2019. doi: 10.1109/CVPR.2019.00398.

Mattia Rigotti, Omri Barak, Melissa R. Warden, Xiao-Jing Wang, Nathaniel D. Daw, Earl K. Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497 (7451):585–590, 2013. doi: 10.1038/nature12160.

Cedric Scheerlinck, Nick Barnes, and Robert Mahony. Continuous-time intensity estimation using event cameras. In *Computer Vision – ACCV 2018*, pp. 308–324. Springer, 2019. doi: 10.1007/978-3-030-20873-8_20.

P. H. Schiller, J. H. Sandell, and J. H. Maunsell. Functions of the ON and OFF channels of the visual system. *Nature*, 322(6082):824–825, 1986. doi: 10.1038/322824a0.

Enrica Strettoi, Elio Raviola, and Ramon F. Dacheux. Synaptic connections of the narrow-field, bistratified rod amacrine cell (AII) in the rabbit retina. *Journal of Comparative Neurology*, 325 (2):152–168, 1992. doi: 10.1002/cne.903250203.

Yoshihiko Tsukamoto and Naoko Omi. OFF bipolar cells in macaque retina: Type-specific connectivity in the outer and inner synaptic layers. *Frontiers in Neuroanatomy*, 9:122, 2015. doi: 10.3389/fnana.2015.00122.

Yoshihiko Tsukamoto and Naoko Omi. ON bipolar cells in macaque retina: Type-specific synaptic connectivity with special reference to OFF counterparts. *Frontiers in Neuroanatomy*, 10:104, 2016. doi: 10.3389/fnana.2016.00104.

Margaret Lin Veruki and Espen Hartveit. Electrical synapses mediate signal transmission in the rod pathway of the mammalian retina. *Journal of Neuroscience*, 22(24):10558–10566, 2002. doi: 10.1523/JNEUROSCI.22-24-10558.2002.

Jan Verweij, Eric P. Hornstein, and Julie L. Schnapf. Surround antagonism in macaque cone photoreceptors. *Journal of Neuroscience*, 23(32):10249–10257, 2003. doi: 10.1523/JNEUROSCI.23-32-10249.2003.

Adrien Wohrer and Pierre Kornprobst. Virtual retina: A biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience*, 26(2):219–249, 2009. doi: 10.1007/s10827-008-0108-4.

Chi Zhang, Yeon Jin Kim, Ana R. Silverstein, Akina Hoshino, Thomas A. Reh, Dennis M. Dacey, and Rachel O. L. Wong. Circuit reorganization shapes the developing human foveal midget connectome toward single-cone resolution. *Neuron*, 108(5):905–918.e3, 2020. doi: 10.1016/j.neuron.2020.09.014.

Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Ling, and Qinghua Hu. Vision meets drones: A challenge. *arXiv:1804.07437*, 2018. URL https://arxiv.org/abs/1804.07437.

# SUPPLEMENTARY METHODS

This section provides the construction and experimental details referenced from the main text. The main text defines the biological construction, continuous S/T dynamics, and task-facing population representation. Here we provide the corresponding construction parameters, exact numerical integration, initialization, and assay-level implementation details.

## S1. RETINAPLEX MODEL CONSTRUCTION AND IMPLEMENTATION

**Population composition and physical scale.** RetinaPlex uses a fixed $3840 \times 2160$ photoreceptor lattice in which every site is assigned to either a rod or a cone. For the fixed $3840 \times 2160$ one-pixel/one-receptor lattice, we solved the physical site spacing self-consistently so that the lattice site density matched the mean human rod-plus-cone density evaluated over the retinal field implied by that spacing, yielding $p = 2.646462374\,\mu\mathrm{m}$. The frozen model contains 7,575,680 rod sites and 718,720 cone sites, giving 8,294,400 photoreceptors. The associated bipolar populations contain 725,913 rod-bipolar centers, 718,720 sustained (S) centers, and 224,896 transient (T) centers. The photoreceptor and bipolar populations therefore comprise 9,963,929 explicitly instantiated retinal units.

The cone-density prior $\rho_C(e)$ is constructed from human photoreceptor measurements (Cooper et al., 2024; Masri et al., 2020). The central and near-central region uses population and principal-meridian statistics from Cooper et al. (2024). Densities are interpolated in log space using a shape-preserving piecewise cubic interpolant. The 3.0-mm endpoint includes the frozen monotonic-envelope adjustment used during model construction, and the 3.5–6.0-mm peripheral continuation uses continuity-scaled human temporal retinal measurements from Masri et al. (2020).

The rod-density prior follows the human retinal fit used in the frozen model (Lee et al., 2019):

$$\rho_R(e) = \max\left[585500\,\exp(-0.1388e) - 598900\,\exp(-0.2998e),\,0\right], \tag{9}$$

where $e$ is expressed in mm and density in cells/mm$^2$. The central rod-free region is determined by the zero crossing of this function.

The cone fraction and lattice-spacing calculation follow Eq. (1) of the main text. The frozen lattice spacing is approximately $p = 2.646\,\mu\mathrm{m}$, corresponding to a modeled rectangular retinal field of approximately $10.16 \times 5.72$ mm.

Rod/cone identity is generated with fixed-seed local quota sampling (seed 40). Within each spatial block, the expected cone count is obtained from the local cone probabilities and the required locations are sampled without replacement according to those probabilities. Locations assigned as cones are removed from the rod pool. This procedure produces the frozen receptor composition used in all subsequent experiments.

**Bipolar centers and local connectivity.** The S pathway uses single-cone sampling motivated by human foveal private-line organization (Zhang et al., 2020). Each modeled cone is associated with one S sampling center and provides its input.

The T pathway uses local multicone convergence informed by primate diffuse bipolar connectivity (Tsukamoto & Omi, 2015; 2016). The frozen model uses six distinct nearby cones with equal weights for each T center. The value six was selected during the earlier construction stage from the tested local convergence range and was fixed before the population-coding experiments.

T and rod-bipolar centers are generated with coverage constraints. Candidate centers are placed according to the relevant receptor population and target convergence rule, nearest-neighbor connections are constructed, and additional centers are added where required to complete receptor coverage. Connections are stored as fixed sparse indices.

The rod-bipolar convergence rule is based on human rod-density and dendritic-field measurements (Lee et al., 2019). The reference convergences are 13, 35, and 49 rod inputs at eccentricities of 1, 4, and 12 mm, respectively, with interpolation between reference points. The realized frozen model contains 1–38 rod inputs per rod-bipolar center over the simulated field.

The horizontal-cell-related modulation follows the operator defined in the main text and is motivated by surround interactions measured in primate cone photoreceptors (Verweij et al., 2003). The frozen parameters are $\sigma = 100\,\mu\mathrm{m}$, $\eta = 0.10$, and $\varepsilon = 10^{-6}$.

**Temporal parameter selection and initialization.** The effective S/T dynamics are motivated by physiological evidence for sustained and transient retinal signaling (Awatramani & Slaughter, 2000; Baden et al., 2013; Kuo et al., 2025). Both branches use $\tau_f = 30$ ms. The frozen S parameters are $r_S = 0.70$ and $\tau_{a,S} = 180$ ms, whereas the T parameters are $r_T = 0.10$ and $\tau_{a,T} = 100$ ms.

These settings were selected from prespecified candidate configurations using preliminary step-response, paired-pulse, and frequency-response checks and were frozen before the state/history coding experiments. The coding-task labels were not used during this parameter-selection stage.

Each nominal 24-Hz video frame advances the temporal states through five integration steps. Natural-video sequences are initialized at the steady state corresponding to their first-frame drive and then proceed through the specified warm-up period. Isolated response-characterization sequences use their corresponding protocol-specific initialization.

**Exact temporal integration and initialization.** The continuous S/T dynamics defined in the main text are integrated exactly under a zero-order-hold assumption. For an integration interval $h = 1/120\,\mathrm{s}$, define

$$q_{f,b} = \exp\left(-\frac{h}{\tau_{f,b}}\right), \qquad q_{a,b} = \exp\left(-\frac{h}{\tau_{a,b}}\right), \qquad \kappa_b = \frac{\tau_{f,b}}{\tau_{a,b} - \tau_{f,b}}\left(q_{a,b} - q_{f,b}\right). \tag{10}$$

For a branch input $\xi_{b,k}$ held constant over integration step $k$, the exact update is

$$\begin{aligned} u_{b,k+1} &= q_{f,b}u_{b,k} + (1 - q_{f,b})\xi_{b,k}, \\ a_{b,k+1} &= q_{a,b}a_{b,k} + \kappa_b u_{b,k} + (1 - q_{a,b} - \kappa_b)\xi_{b,k}, \\ y_{b,k+1} &= u_{b,k+1} - (1 - r_b)a_{b,k+1}. \end{aligned} \tag{11}$$

Both state updates use the states from the previous integration step; in particular, the adaptation update uses $u_{b,k}$ rather than the newly updated $u_{b,k+1}$. Each nominal 24-Hz input frame advances the model through five integration steps.

Natural-video sequences are initialized at the steady state associated with their first-frame branch drive,

$$u_{b,0} = a_{b,0} = \xi_{b,0}, \qquad y_{b,0} = r_b\xi_{b,0}. \tag{12}$$

Internal temporal states then persist continuously across frame boundaries.

**Task-facing population representation.** The quantitative snapshot interface contains separate $8 \times 8$ spatial summaries for the S and T populations. Each spatial coordinate is normalized by the number of branch centers assigned to that region. With zero-based image coordinates, the row-major bin index is

$$k = 8\left\lfloor \frac{8\,\mathrm{row}}{2160} \right\rfloor + \left\lfloor \frac{8\,\mathrm{col}}{3840} \right\rfloor. \tag{13}$$

The resulting 64 S coordinates and 64 T coordinates form the 128-dimensional population representation defined in the main text.

Only the current population output is supplied to the downstream readouts. Previous video frames, previous population snapshots, and the internal fast/adaptation states are not provided as additional readout variables. Display interpolation and normalization used for full-field visualization are applied separately from the quantitative readout representation.

### ON/OFF POLARITY AND DOWNSTREAM PLATFORM EXTENSIONS

In addition to the cone-driven S/T interface used in the main population-coding experiments, RetinaPlex implements explicit ON/OFF pathways, an effective AII-like rod–cone crossover, and a fixed GC-like population readout. These modules provide an extended platform output while retaining the same photoreceptor and bipolar spatial construction.

**ON/OFF pathway construction.** Parallel ON and OFF signaling motivates separate polarity-specific pathways (Schiller et al., 1986). The model expresses polarity in the luminance-oriented coordinates of the effective cone signal $\widetilde{R}_C$ defined in the main text. Let $W_b$, for $b \in \{S, T\}$, denote the fixed, row-normalized cone-to-bipolar sampling operator: single-cone sampling for S and equal-weight six-cone sampling for T. The two drives are

$$\xi_b^{\mathrm{ON}} = W_b \widetilde{R}_C, \qquad \xi_b^{\mathrm{OFF}} = W_b(\mathbf{1} - \widetilde{R}_C). \tag{14}$$

The complement is applied after horizontal-cell-related modulation, without an additional clamp. ON and OFF use the same centers, connectivity, and branch-specific temporal parameters, with separate fast and adaptation states for each polarity. Both follow the main-text two-state dynamics and exact zero-order-hold update. At the start of a sequence, each polarity is initialized at the steady state corresponding to its own first-frame drive.

Because $W_b\mathbf{1} = \mathbf{1}$, the drives satisfy $\xi_b^{\mathrm{ON}} + \xi_b^{\mathrm{OFF}} = \mathbf{1}$. With the matched initialization and shared temporal coefficients, the outputs consequently satisfy $y_b^{\mathrm{ON}} + y_b^{\mathrm{OFF}} = r_b\mathbf{1}$, up to floating-point precision. The two continuous response streams are retained separately for downstream polarity-specific operations. They share the existing spatial sites and do not change the reported spatial-unit count.

**Rod-bipolar input and AII-like transmission.** The rod branch uses $R_R = \mathrm{clip}(I/2, 0, 1)$, the corresponding same-type surround modulation, and the fixed rod-to-bipolar connectivity described above. Its single ON rod-bipolar population follows the same two-state response family with $\tau_{f,R} = 40$ ms, $\tau_{a,R} = 250$ ms, and $r_R = 0.55$.

Anatomical rod–AII–cone-pathway connections and AII-to-ON-cone electrical coupling motivate the effective crossover (Strettoi et al., 1992; Veruki & Hartveit, 2002). Activity-dependent gain control at the rod-bipolar–AII synapse motivates a local transmission-efficacy state (Dunn & Rieke, 2008). For rod-bipolar center $j$, define

$$\rho_j(t) = \frac{u_{R,j}(t)}{u_{R,j}^{\mathrm{sat}}}, \qquad \tau_D \frac{\mathrm{d}d_j}{\mathrm{d}t} = 1 - \big(1 + \rho_j(t)\big)d_j, \qquad \tau_D = 200\ \mathrm{ms}, \tag{15}$$

where $u_{R,j}$ is the rod-bipolar fast state and $u_{R,j}^{\mathrm{sat}}$ is its independently computed steady-state reference under uniform input $I = 2$. The fixed 200-ms value specifies the effective model timescale. The transmitted signal is $y_{R,j}^{\mathrm{eff}} = d_j y_{R,j}$.

After each parent update, the newly obtained $\rho_{j,k}$ is held constant for the efficacy update over interval $h$:

$$d_{j,k} = d_{j,k}^{\infty} + \big(d_{j,k-1} - d_{j,k}^{\infty}\big)\exp\left[-\frac{(1+\rho_{j,k})h}{\tau_D}\right], \qquad d_{j,k}^{\infty} = \frac{1}{1+\rho_{j,k}}. \tag{16}$$

Initialization uses $d_{j,0} = 1/(1+\rho_{j,0})$ at the first-frame parent steady state.

For each receiving S/T center $i$, the support $\mathcal{N}_{b,i}$ contains the nearest at most 16 rod-bipolar centers within $30\,\mu$m. Distance ties are resolved by canonical rod-bipolar index; support is not filled from beyond this radius. The local rod contribution uses the fixed divisor 16, including at sites with fewer neighbors:

$$A_{b,i}^{R} = \frac{1}{16}\sum_{j \in \mathcal{N}_{b,i}} y_{R,j}^{\mathrm{eff}}, \qquad \kappa_{b,i} = \frac{|\mathcal{N}_{b,i}|}{16}, \qquad \lambda_{b,i} = \frac{\kappa_{b,i}}{1+\kappa_{b,i}}. \tag{17}$$

The retained four-channel crossover implementation combines cone-derived and rod-derived drives by support-gated mixing. With $B_S = 0.70$, $B_T = 0.10$, and $\beta_b = B_b/0.55$, define

$$v_{b,i}^{\mathrm{ON}} = \beta_b A_{b,i}^{R}, \qquad v_{b,i}^{\mathrm{OFF}} = B_b - v_{b,i}^{\mathrm{ON}}. \tag{18}$$

For polarity $p \in \{\mathrm{ON}, \mathrm{OFF}\}$, the composite drive is

$$Z^p_{b,i} = (1 - \lambda_{b,i}) y^p_{b,i} + \lambda_{b,i} v^p_{b,i}. \tag{19}$$

Sites without local rod-bipolar support have $\lambda_{b,i} = 0$ and preserve the cone-driven output. This effective transmission stage retains signed outputs and introduces no further rectification or temporal state beyond the inherited efficacy state.

**Eccentricity-dependent GC spatial support.** The fixed full-field GC-like readout uses local and broad spatial operators. Eccentricity-dependent ganglion-cell receptive fields and dendritic scales motivate this spatial organization (Croner & Kaplan, 1995; Dacey & Petersen, 1992); the kernel values below are fixed computational settings of the full-field v3 implementation.

Native T-center values are assigned to a $20\,\mu\mathrm{m}$ computational grid of $286 \times 510$ bins by within-bin averaging. Let $\mathcal{B}_T X$ be this rasterization of a native T-population vector $X$, and let $M_T$ be the binary occupancy mask obtained from the fixed T-center positions. Output values are sampled back at the native receiving positions.

Three kernel pairs use local scale parameters $(75, 115.909092, 150)\,\mu\mathrm{m}$ and broad scale parameters $(400, 618.181825, 800)\,\mu\mathrm{m}$. Each two-dimensional kernel is a separable Gaussian with $\sigma = r/2$, sampled at $20\,\mu\mathrm{m}$ spacing and truncated at $\lfloor r/(20\,\mu\mathrm{m}) \rfloor$ bins on each side of each axis. These pairs are associated with eccentricity control points $0$, $2.5$, and $4.5$ mm. For receiving position $i$ with eccentricity $e_i$, define

$$\begin{aligned} a_i &= \mathrm{clip}\Big(\frac{e_i}{2.5\,\mathrm{mm}}, 0, 1\Big), \quad b_i = \mathrm{clip}\Big(\frac{e_i - 2.5\,\mathrm{mm}}{2.0\,\mathrm{mm}}, 0, 1\Big), \\ (w_{i,1}, w_{i,2}, w_{i,3}) &= (1 - a_i,\ a_i - b_i,\ b_i). \end{aligned} \tag{20}$$

For $\ell \in \{\mathrm{local}, \mathrm{broad}\}$, pooling is

$$(\mathcal{P}_\ell X)_i = \frac{\sum_{m=1}^{3} w_{i,m} \left[K_{\ell,m} * (M_T \mathcal{B}_T X)\right]_{\mathrm{bin}(i)}}{\sum_{m=1}^{3} w_{i,m} \left[K_{\ell,m} * M_T\right]_{\mathrm{bin}(i)}}. \tag{21}$$

The kernel-weighted numerators and denominators are combined before the occupancy normalization. The denominator is positive at occupied native receiving sites. Spatial weights depend only on the fixed receiving coordinates. This GC support is distinct from the 16-neighbor rod-bipolar support used in the crossover stage.

**Fixed parallel GC-like outputs.** The readout retains four functional streams, termed Robust (R), Adaptive (A), Context (C), and Sustained (S), separately for ON and OFF polarity. For each polarity, the T-derived local and broad drives are

$$L^p_k = \mathcal{P}_{\mathrm{local}} Z^p_{T,k}, \qquad B^p_k = \mathcal{P}_{\mathrm{broad}} Z^p_{T,k}, \qquad R^p_k = [L^p_k - B^p_k]_+, \tag{22}$$

where $[x]_+ = \max(x, 0)$ componentwise. A separate reference state at each native T position follows

$$Q^p_k = q_{\mathrm{ref}} Q^p_{k-1} + (1 - q_{\mathrm{ref}}) L^p_k, \qquad q_{\mathrm{ref}} = \exp(-h/\tau_{\mathrm{ref}}), \qquad \tau_{\mathrm{ref}} = 80\ \mathrm{ms}. \tag{23}$$

Using the updated reference, the unit-scale temporal and contextual outputs are

$$\begin{aligned} D^p_k &= [L^p_k - Q^p_k]_+, \\ C^p_{\mathrm{unit},k} &= \mathcal{P}_{\mathrm{broad}} D^p_k, \qquad A^p_{\mathrm{unit},k} = [D^p_k - C^p_{\mathrm{unit},k}]_+. \end{aligned} \tag{24}$$

Here $D^p_k$ is defined at native T positions and rasterized through $\mathcal{B}_T$ before broad pooling. The contextual threshold is fixed at zero. The sustained stream directly preserves the native S output, $S^p_{\mathrm{out},k} = Z^p_{S,k}$, with its signed values and spatial detail. The eight output streams are therefore R/A/C at T positions and direct S at S positions, each with explicit ON/OFF polarity.

The reference is initialized as $Q^p_0 = L^p_0$ after first-frame steady initialization of the parent model. The full-field implementation updates C4 and GC states at $h = 1/120$ s; its 30-Hz video interface uses four updates per frame. Spatial kernels, the 80-ms reference timescale, and the zero contextual threshold are fixed, with no trainable GC coefficients.

The archived readout configuration calibrates A/C amplitudes separately for each polarity. Its positive scale $s_p$ is the RMS of the corresponding C4 T-route raster over the fitting sequences. This gives $A^p = A^p_{\text{unit}}/s_p$ and $C^p = C^p_{\text{unit}}/s_p$, while R and direct S retain their original scale. RMS accumulation uses frame-end raster values, including unoccupied zeros and excluding the initialization sample. Subsequent task-feature standardization is a separate readout operation. The crossover and GC extensions reuse the existing spatial substrate; their dynamic states and functional output streams are not counted as additional anatomical units.

## S2. Experimental protocols and evaluation

**Fork–rejoin experiments and DAVIS partition.** The controlled fork–rejoin experiments use 16 fitting backgrounds and eight evaluation carriers. Each background is crossed with two current-state conditions and two recent-history orders.

Natural-video experiments use DAVIS (Pont-Tuset et al., 2017). Fifteen complete source videos are used for development and readout fitting: *boxing-fisheye*, *classic-car*, *dog-gooses*, *dogs-scale*, *drone*, *kid-football*, *lindy-hop*, *night-race*, *schoolgirls*, *scooter-board*, *skate-park*, *stunt*, *tuk-tuk*, *varanus-cage*, and *walking*. Five complete sources are used for evaluation: *dogs-jump*, *gold-fish*, *judo*, *lab-coat*, and *loading*. Each source contributes two deterministic clips.

For image width $W$, the two controlled event patterns are

$$A(x) = \sin\left(\frac{4\pi x}{W}\right), \qquad B(x) = \cos\left(\frac{4\pi x}{W}\right), \tag{25}$$

and the current-state modulation uses

$$D(x) = \frac{A(x) - B(x)}{\sqrt{2}}. \tag{26}$$

For normalized scene luminance $Y_t$ and $q \in \{-1, +1\}$, the carrier is

$$C_t(q) = 0.20 + 0.60Y_t + 0.04qD. \tag{27}$$

During each controlled-event epoch, the corresponding A or B pattern is added with amplitude 0.08.

Each sequence contains 26 frames:

- frames 0–11: natural-video warm-up;
- frames 12–14: first controlled event;
- frames 15–17: natural-video interval;
- frames 18–20: second controlled event;
- frames 21–25: common continuing natural-video suffix.

The two deterministic clip starting positions are obtained from the rounded one-third and two-thirds points of the valid starting-index range.

AB and BA conditions share the same underlying video and current-state condition. Their sensory input is identical throughout the common suffix. The primary sampling point is the third processed suffix frame, nominally 125 ms after rejoin. The complete sampled suffix corresponds to 41.7, 83.3, 125.0, 166.7, and 208.3 ms.

At each lag, the development set contains 120 population snapshots and the evaluation set contains 40 snapshots, including 20 matched AB/BA history pairs. State and history readout directions and thresholds are obtained from the development representation using the mean-difference procedure defined in Section 3.3 and are frozen before evaluation.

The AB/BA experiment measures recent-context accessibility after exact sensory rejoin. A separate terminal-matched experiment, described below, holds the final controlled event fixed while exchanging the order of the two earlier events.

**Matched-current mechanism analysis.** The mechanism experiments in Figure 3 use VisDrone scenes (Zhu et al., 2018). The primary analysis contains eight scenes with eight fixed probes per scene, giving 64 probes. For each probe, a natural seven-frame history ending at target frame $F_t$ is compared with a stabilized sequence in which $F_t$ is repeatedly presented. The two conditions therefore share the same current image at the measurement time.

For each probe, let

$$P = (P_S, P_T)$$

denote the target-versus-context contrast associated with the current image, and let

$$C = (C_S, C_T)$$

denote the branch contrast between the natural-history and stabilized conditions. Their acute separation is

$$\theta = \arccos\left(\frac{|P^\top C|}{\|P\|_2\|C\|_2}\right), \qquad A_{\text{sep}} = \sin\theta. \tag{28}$$

History-related response magnitude in branch $b$ is

$$E_b = \text{RMS}\left(y_b^{\text{natural}} - y_b^{\text{stable}}\right), \qquad F_T = \frac{E_T}{E_S + E_T}. \tag{29}$$

The *Equal retention* condition sets $r_S = r_T = 0.40$. The *Equal adaptation* condition sets

$$\tau_{a,S} = \tau_{a,T} = \sqrt{180 \times 100}\ \text{ms},$$

and the fully shared condition combines both manipulations.

Continuous retained-response scans vary $\Delta r = r_S - r_T$ while holding the mean retained fraction at 0.40. Continuous adaptation scans preserve the geometric-mean adaptation time $g$ and use

$$\tau_{a,S} = g(1.8)^{\alpha/2}, \qquad \tau_{a,T} = g(1.8)^{-\alpha/2}. \tag{30}$$

Four additional reserve scenes are included in the frozen directionality checks. Sweep summaries use the mean of per-scene medians. Time zero for this experiment is the onset of the matched current frame.

**Terminal-matched earlier-order assay.** To test earlier-event order while holding the terminal event fixed, each spatial event family $k$ contains three equal-status phase events:

$$A_k = \sin(\phi_k), \qquad B_k = \sin(\phi_k + 2\pi/3), \qquad C_k = \sin(\phi_k + 4\pi/3), \tag{31}$$

with event amplitude 0.08.

| Family | $\phi_k(x, y)$ | Coordinates | $(f_x, f_y)$ | Role |
|---|---|---|---|---|
| HX1 | $2\pi x$ | $x = j/W,\ y = i/H$ | $(1, 0)$ | TRAIN |
| HX2 | $4\pi x$ | $x = j/W,\ y = i/H$ | $(2, 0)$ | TRAIN |
| VY1 | $2\pi y$ | $x = j/W,\ y = i/H$ | $(0, 1)$ | TRAIN |
| DP1 | $2\pi(x + y)$ | $x = j/W,\ y = i/H$ | $(1, 1)$ | TRAIN |
| HX3 | $6\pi x$ | $x = j/W,\ y = i/H$ | $(3, 0)$ | EVALUATION |
| DM2 | $4\pi(x - y)$ | $x = j/W,\ y = i/H$ | $(2, -2)$ | EVALUATION |

All six permutations are generated and grouped into the three terminal-matched contrasts

$$ABC/BAC, \qquad ACB/CAB, \qquad BCA/CBA.$$

Within each pair, the three event identities, event durations, terminal event, and continuing visual suffix are identical; only the order of the first two events changes.

Each sequence contains 32 frames:

- frames 0–11: natural-video warm-up;
- frames 12–14: event 1;
- frames 15–17: natural-video interval;
- frames 18–20: event 2;

- frames 21–23: natural-video interval;
- frames 24–26: common terminal event;
- frames 27–31: common continuing natural-video suffix.

The current-state modulation is fixed at $q = 0$. The primary checkpoint is the third processed suffix frame, 125 ms after common-suffix onset and 250 ms after onset of the shared terminal event.

The four event families `HX1`, `HX2`, `VY1`, and `DP1` are used to estimate shared temporal gains. The two held event families are `HX3` and `DM2`. The development/evaluation video partition is the same as above.

For event family $k$, terminal-matched contrast $p$, and lag $L$, define the observed order-dependent population displacement as

$$d_{k,p}(L) = \mathbb{E}\left[z^{(+)}_{k,p}(L) - z^{(-)}_{k,p}(L)\right]. \tag{32}$$

An instantaneous spatial descriptor $g_{k,p}$ is computed from the exchanged event pair through the fixed retinal spatial front end. The predicted population displacement is

$$\widehat{d}_{k,p}(L) = \begin{bmatrix}\Delta\gamma_S(L)\, g^S_{k,p} \\ \Delta\gamma_T(L)\, g^T_{k,p}\end{bmatrix}, \tag{33}$$

where the two branchwise temporal gains are fitted on the four fitting event families:

$$\Delta\gamma_b(L) = \frac{\sum_i \langle g^b_i, d^b_i\rangle}{\sum_i \|g^b_i\|_2^2}. \tag{34}$$

The fitted gains are frozen before evaluation of the two held event families. For each held condition, the normalized predicted displacement defines the order-readout direction and AUROC is computed from the resulting sequence scores. The primary summary is the equal-weight macro average over the two held event families and three terminal conditions.

**Selective representation intervention.** The selective intervention analysis uses the 128-dimensional population snapshot at the primary 125-ms lag. Let $w_{\mathrm{state}}$ and $w_{\mathrm{hist}}$ denote the development-set readout directions defined in Section 3.3. A history-selective direction orthogonal to the state direction is

$$u_{\mathrm{hist}} = \mathrm{normalize}\left[w_{\mathrm{hist}} - (w_{\mathrm{hist}}^\top w_{\mathrm{state}}) w_{\mathrm{state}}\right], \tag{35}$$

with the state-selective direction defined analogously.

For target direction $u$, the rank-one attenuation is

$$z' = z - \lambda u u^\top (z - \mu_{\mathrm{DEV}}), \tag{36}$$

where

$$\lambda \in \{0,\ 0.25,\ 0.50,\ 0.75,\ 1.00\}.$$

At each intervention strength, new state and history readouts are fitted on the intervened development representation and frozen before held evaluation.

Each target intervention is compared with 64 matched random rank-one directions. Random directions are constrained to be orthogonal to the protected readout direction and are scaled to match the development-set perturbation energy of the selective intervention.

**Public primate RGC analysis and model correspondence.** The biological comparison uses the public marmoset retinal ganglion-cell recordings and associated dataset of Krüppel et al. (2023a;b).

For each saccadic transition, the preceding fixation defines the recent stimulus and the following fixation defines the current stimulus. Population responses are summarized as spike counts in a 50-ms window, with the primary coding window spanning 100–150 ms after current-fixation onset.

Two conditional four-class decoding problems are evaluated. Current-stimulus decoding predicts the current stimulus under fixed preceding conditions, whereas recent-stimulus decoding predicts

the preceding stimulus under fixed current conditions. Decoding uses conditional nearest-class centroids, training-fold standardization, and five-fold chronological cross-validation.

Temporal phenotype is measured independently using the full-field ON/OFF-step responses. For each cell, $E_i$ and $L_i$ denote the positive baseline-corrected mean firing rates in the early $[0, 200)$-ms and late $[200, 400)$-ms windows, respectively, summed over gray-to-white and gray-to-black steps. Baseline firing is measured over $[-200, 0)$ ms. The persistence score is

$$P_{\mathrm{persist},i} = \frac{L_i}{E_i + L_i}. \tag{37}$$

Cells with $E_i + L_i = 0$ have undefined persistence.

For each group, coding balance is

$$B = \mathrm{BA}_{\mathrm{current}} - \mathrm{BA}_{\mathrm{recent}}, \tag{38}$$

where BA denotes balanced accuracy.

The corresponding model analysis uses the S64 and T64 components of the frozen RetinaPlex representation and the same current-versus-recent conditioning logic. The branch coding-balance contrast is

$$I = (C_S - H_S) - (C_T - H_T), \tag{39}$$

where $C_b$ and $H_b$ denote current and recent decoding performance for branch $b$. The model comparison additionally includes 1,000 spatially and dimensionally matched mixed-group controls. Model and RGC coding-balance effects are summarized separately and compared at the population-pattern level.

**Representation interfaces and spatial-budget comparison.** The representation comparison in Table 1 uses the same fork–rejoin task, the same development/evaluation split, 128 task-facing output coordinates, and development-fitted mean-difference readouts.

The seven interfaces are defined as follows:

**Current frame.** Current $8 \times 8$ scene information represented within the 128-coordinate interface.

**Adjacent-frame pair.** Concatenated $8 \times 8$ representations of the current and immediately preceding frames.

**Event surface.** Positive and negative inter-frame luminance changes accumulated with a fixed 100-ms exponential decay. This follows the event-based representation family used for the comparison (Gallego et al., 2022; Scheerlinck et al., 2019; Rebecq et al., 2019).

**Explicit frame–event hybrid.** Current scene intensity combined with a signed event representation.

**Raw Two-IIR.** A generic temporal interface using a 200-ms exponential moving average together with a 50-ms-minus-200-ms temporal residual.

**Retinal Two-IIR.** The same retinal S/T spatial front end used by RetinaPlex, with 100-ms and 50-ms exponential temporal states applied to the S and T spatial inputs, respectively.

**RetinaPlex.** The frozen heterogeneous S/T population representation used throughout the main snapshot experiments.

Current-state and recent-history performance are evaluated separately by AUROC, and joint accessibility is summarized by the smaller of the two AUROCs, $J$.

For the spatial-budget comparison, the native RetinaPlex layout contains

$$718720 + 224896 = 943616$$

pre-aggregation S/T output sites. The dense All-Cone construction assigns dense S and T outputs across all 8,294,400 receptor locations, giving

$$2 \times 8294400 = 16588800$$

S/T output sites. The dense-to-native output-site ratio is therefore approximately $17.58\times$. Both constructions are reduced to the same 64-S plus 64-T task-facing representation and are evaluated with the same readout protocol.

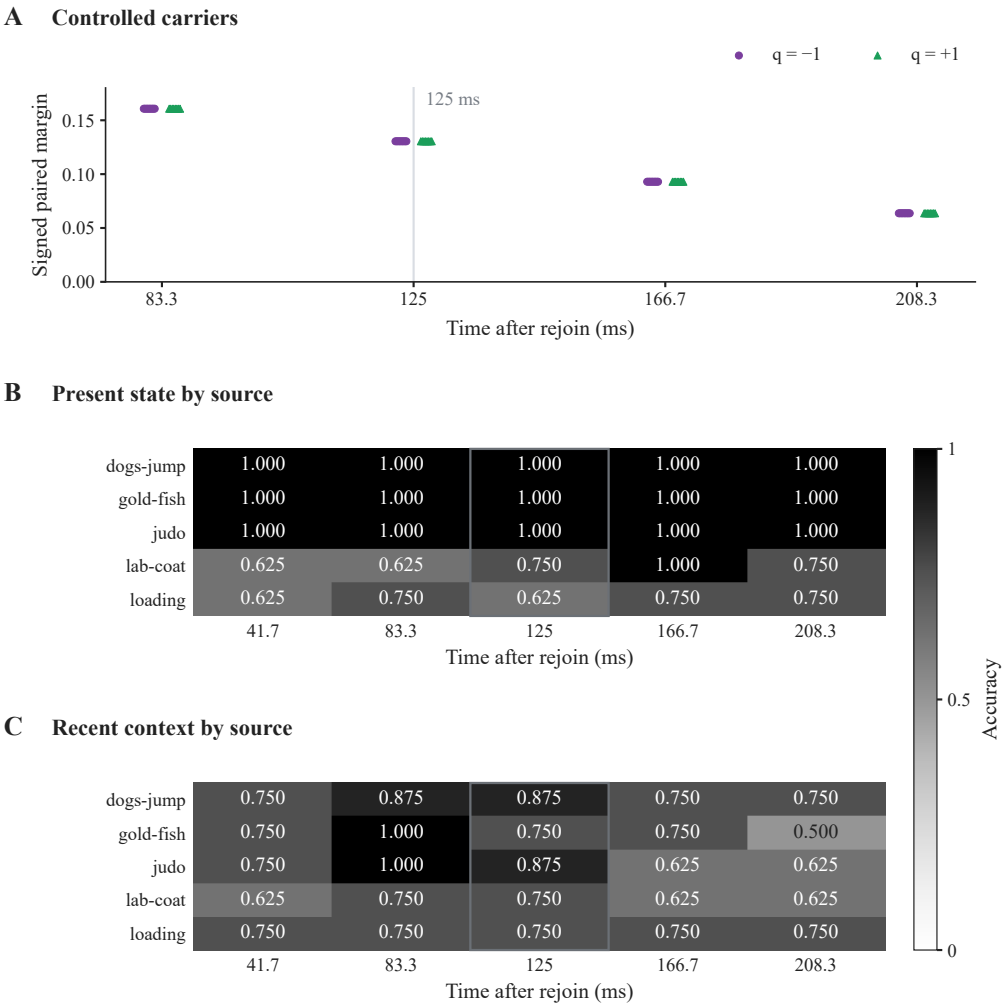



Figure S1: **Complete paired-history and source-level evidence after sensory rejoin. (A)** Signed recent-context margins for all eight controlled carriers and both current-state conditions at four recorded lags (83.3, 125, 166.7, and 208.3 ms; 16 matched pairs per lag). Positive margins show separation of the paired histories throughout this interval. Small horizontal offsets separate overlapping points; the vertical line marks 125 ms. **(B,C)** Present-state and recent-context accuracy for each DAVIS evaluation video at all five suffix lags. Each cell contains eight snapshots (two clips, two current states, and two histories). Outlined columns mark the primary 125-ms checkpoint. The shared color scale reveals the source-level distribution underlying the main-text aggregate results. All readouts use the development-fitted directions and thresholds.

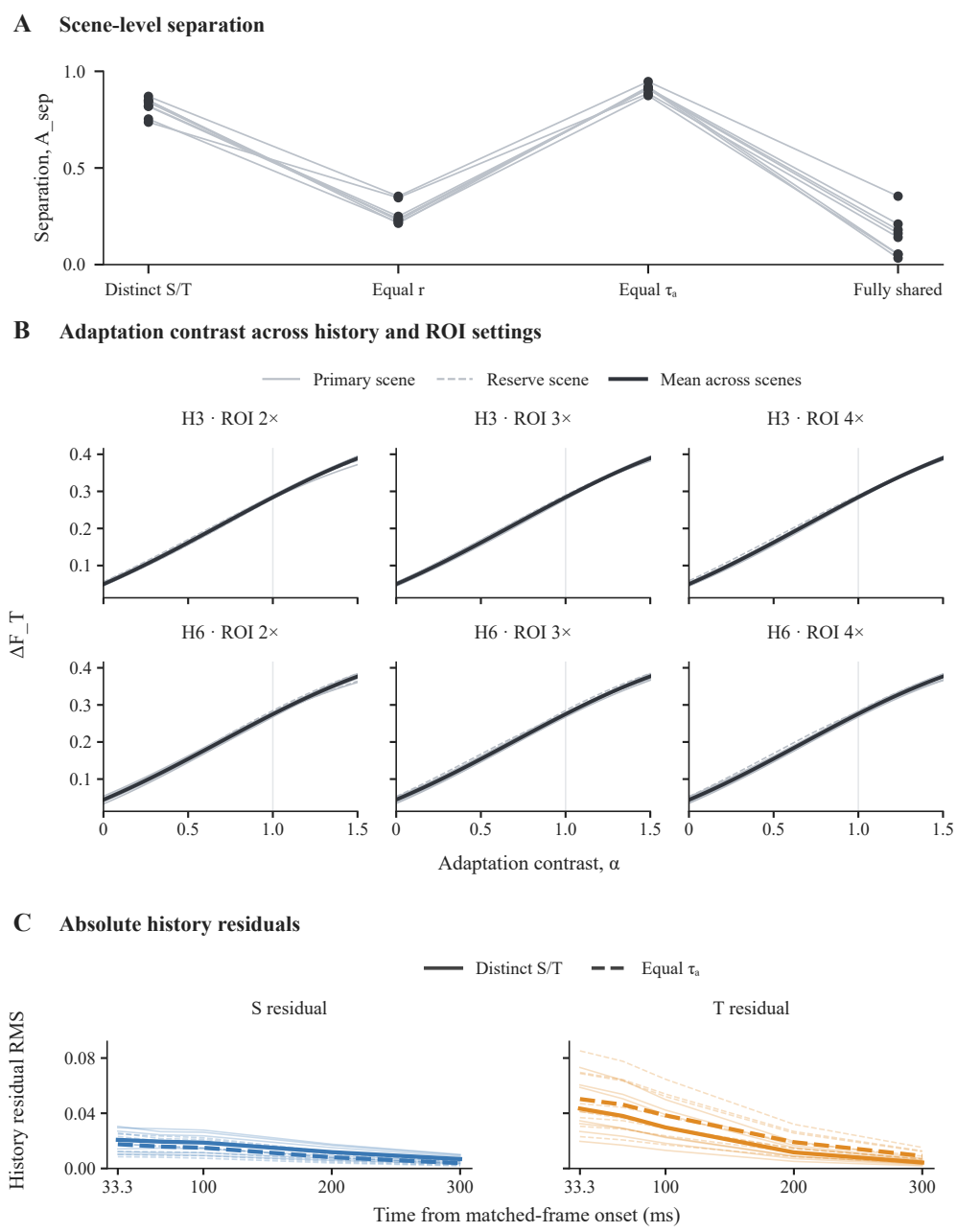



Figure S2: **Temporal mechanisms across scenes and measurement settings. (A)** Separation $A_{\text{sep}}$ at 66.7 ms for four temporal conditions. Each point is the median across eight probes in one scene; lines connect the same scene. Equal retained response reduces separation in all eight scenes, whereas equal adaptation preserves strong separation. **(B)** Adaptation-contrast scans across two history spans (H3: 100 ms; H6: 200 ms) and three context-region scales (ROI $2\times$, $3\times$, $4\times$). Each setting shows 31 parameter values and twelve scenes: eight primary (solid thin lines) and four reserve (dashed thin lines). Thick lines are scene-equal arithmetic means. For each scene, $\Delta F_T$ is the difference between probe-median $F_T$ at 66.7 and 300 ms. Vertical lines mark $\alpha = 1$. **(C)** Absolute S/T history-residual RMS at all five recorded checkpoints. Thin lines are within-scene probe medians; thick lines average the eight scene medians. Shared vertical scales show the branch magnitudes under distinct and equal-adaptation dynamics. All times are measured from matched-frame onset.

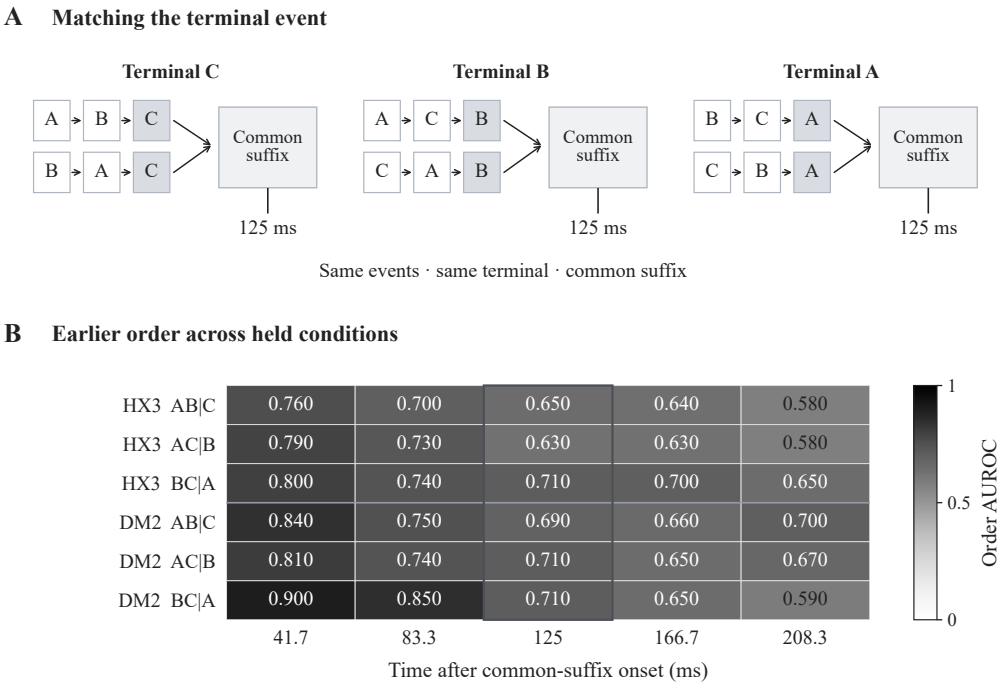



Figure S3: **Earlier-order accessibility with a common terminal event.** **(A)** Three sequence pairs exchange the first two events while preserving the terminal event and continuing suffix. The diagrams show sequence order; 125 ms marks the primary sampling time after suffix onset. **(B)** Order AUROC for every held event-family/terminal combination at five lags. For example, $AB|C$ denotes $ABC$ versus $BAC$. The outlined column identifies 125 ms. Earlier order remains accessible across both held families and all three terminal conditions. **(C)** The same assay for Raw Two-IIR, Retinal Two-IIR, and RetinaPlex at 125 ms. Small points are the six conditions; diamonds are their equal-weight macro averages. Readout directions combine event-specific spatial descriptors with two branchwise temporal gains fitted on the four development families. The dashed line marks chance (0.5). Sampling units are specified in Table S2.

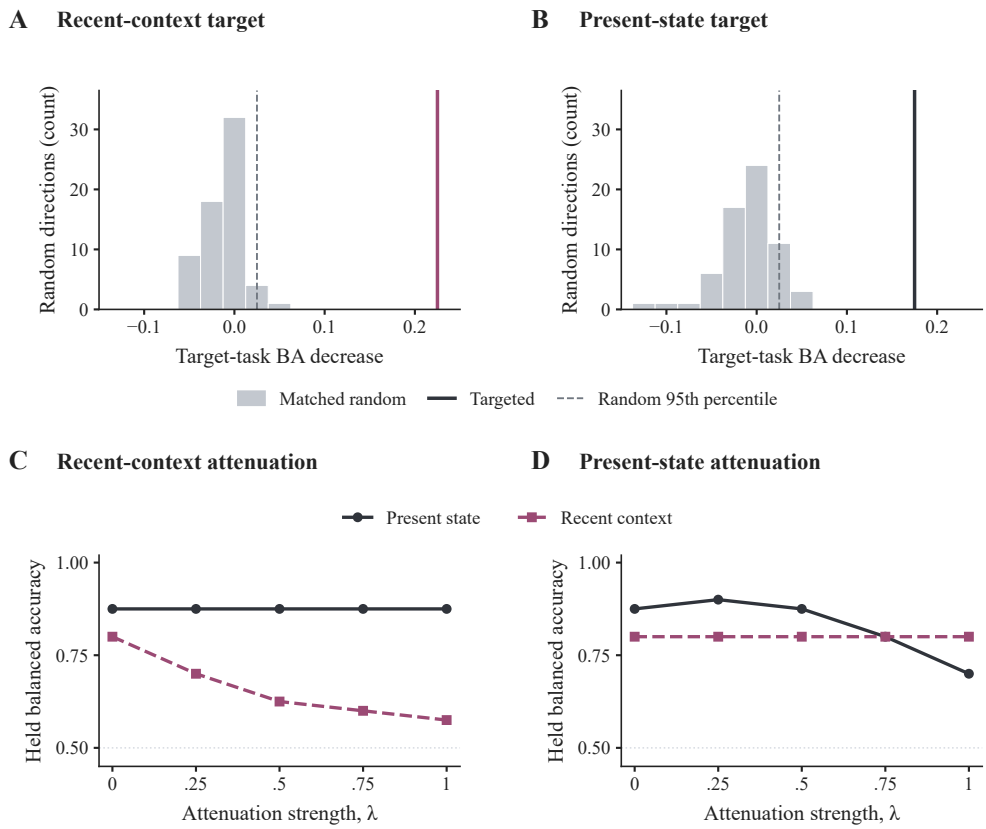



Figure S4: **Selective attenuation across intervention strengths and matched random controls. (A,B)** Target-task balanced-accuracy decreases for 64 random rank-one directions per target. Controls match the protected-axis constraint and development-set perturbation energy of maximal targeted attenuation. Solid lines mark targeted decreases (0.225 for recent context and 0.175 for present state); dashed lines mark the random 95th percentile (0.025 for each target). Negative decreases correspond to improved accuracy. **(C,D)** Present-state and recent-context balanced accuracy across all five attenuation strengths. Readouts were refitted on each intervened development representation before evaluation. The curves show how selective loss develops across intervention strength while the other task is preserved. All panels use 40 balanced evaluation snapshots at 125 ms. Random-control distributions refer to the maximal-attenuation endpoint.

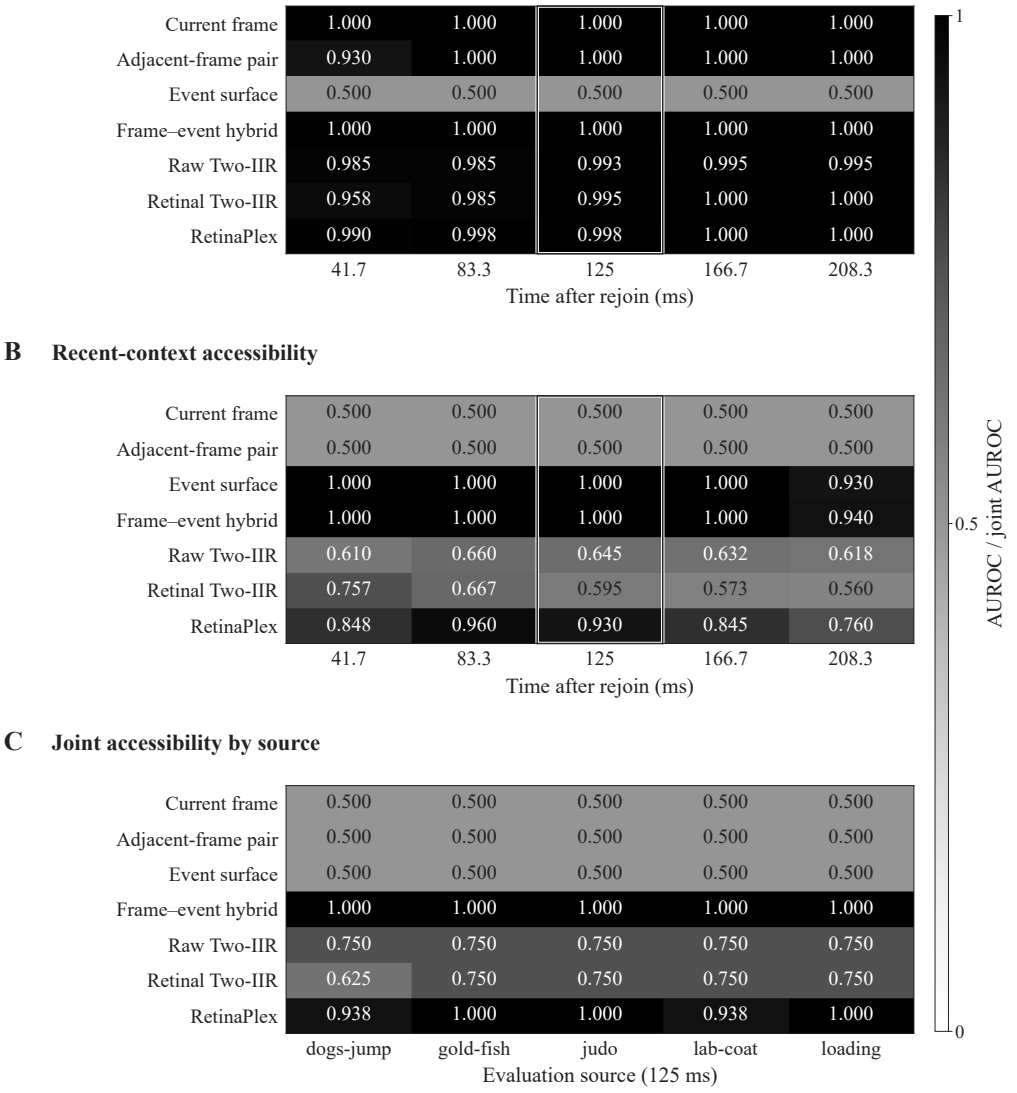



Figure S5: **Representation accessibility across time and video sources. (A,B)** Pooled present-state and recent-context AUROC for all seven interfaces at five post-rejoin lags. Outlined columns identify the 125-ms checkpoint reported in main-text Table 1. **(C)** Joint accessibility within each evaluation video at 125 ms, $J_s = \min(\mathrm{AUROC}_{\mathrm{current},s}, \mathrm{AUROC}_{\mathrm{recent},s})$. The source-level results show high joint accessibility for RetinaPlex across all five videos and place it alongside the explicit frame–event hybrid and temporal-filter controls. Panels A/B use pooled scores, whereas C evaluates each source separately. All interfaces allocate 128 output coordinates and use the same development-fitted mean-difference readout protocol. Values are displayed to three decimals with a shared 0–1 scale; binary chance AUROC is 0.5.

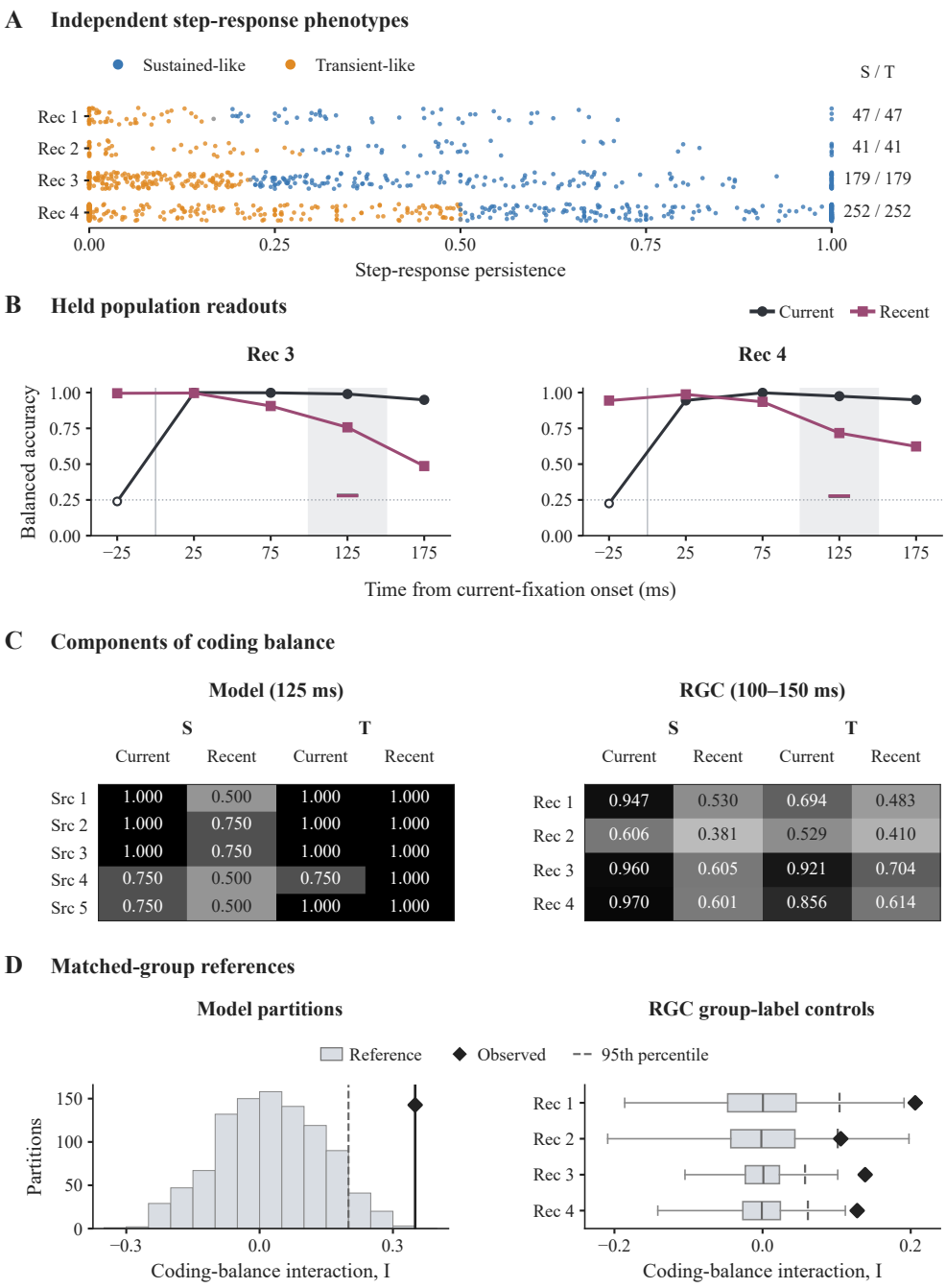



Figure S6: **Independent temporal phenotypes and the components of coding balance.** **(A)** Step-response persistence with independently assigned sustained-like and transient-like groups; counts and recording identities are given in Table S2. **(B)** Current/recent decoding in the two originally held recordings. Shading marks 100–150 ms; open pre-onset current markers are future-current controls. Ticks show the reference 95th percentiles. **(C)** Current/recent balanced accuracies underlying the model and RGC coding-balance comparison, by source or recording (chance: 0.5 and 0.25, respectively). **(D)** Coding-balance interactions against 1,000 model mixed partitions and 1,000 RGC group-label controls per recording. Diamonds mark observed values; dashed markers indicate reference 95th percentiles. Boxes show quartiles and medians; whiskers span the full range. The observed interactions have consistent positive signs across recordings.

Table S1: **Construction parameters and density anchors for the retinal substrate.** The population counts, sampling rules, and cone-density anchors specify the spatial model underlying Supplementary Methods S1. The last block records the fixed parameters used by the cone-driven S/T measurements.

**A. Geometry, population counts, and sampling**

| | |
|---|---|
| Photoreceptor lattice | $3840 \times 2160$ |
| Rod / cone sites | $7{,}575{,}680$ / $718{,}720$ |
| Rod-bipolar / sustained / transient centers | $725{,}913$ / $718{,}720$ / $224{,}896$ |
| Total photoreceptor and bipolar spatial units | $9{,}963{,}929$ |
| Lattice spacing | $2.646462374\,\mu\mathrm{m}$ |
| Field width $\times$ height | $10.162416 \times 5.716359$ mm |
| Local quota block / random seed | $54 \times 54$ sites / 40 |
| Cone inputs per S / T center | 1 / 6 |
| Realized rod inputs per rod-bipolar center | 1–38 |

**B. Cone-density anchors**

| $e$ (mm) | $\rho_C(e)$ (cells/mm$^2$) | $e$ (mm) | $\rho_C(e)$ (cells/mm$^2$) |
|---|---|---|---|
| 0.00 | 152,906.0 | 2.00 | 11,154.7 |
| 0.10 | 83,664.0 | 2.50 | 9,787.7 |
| 0.20 | 61,738.9 | 3.00$^\dagger$ | 9,787.7 |
| 0.30 | 46,305.5 | 3.50$^\ddagger$ | 8,783.952641 |
| 0.40 | 36,066.8 | 4.00$^\ddagger$ | 7,883.141510 |
| 0.50 | 30,269.5 | 4.50$^\ddagger$ | 7,074.710282 |
| 0.75 | 22,348.1 | 5.00$^\ddagger$ | 6,349.185222 |
| 1.00 | 17,723.9 | 5.50$^\ddagger$ | 5,698.064143 |
| 1.25 | 15,093.4 | 6.00$^\ddagger$ | 5,113.716776 |
| 1.50 | 13,483.3 | – | – |

**C. Fixed cone-driven S/T parameters**

| Branch | $\tau_f$ (ms) | $\tau_a$ (ms) | Retained fraction $r$ |
|---|---|---|---|
| Sustained | 30 | 180 | 0.70 |
| Transient | 30 | 100 | 0.10 |

HC-related modulation: $\sigma = 100\,\mu\mathrm{m}$, $\eta = 0.10$, $\varepsilon = 10^{-6}$.
Time update: 120 Hz; nominal DAVIS input: 24 Hz (five substeps per frame).
Snapshot interface: $64$ S coordinates $+$ $64$ T coordinates $= 128$.

Cone anchors through 2.5 mm use human measurements (Cooper et al., 2024). † The 3.0-mm value is the construction's monotonic-envelope adjustment; ‡ values from 3.5–6.0 mm form the continuity-scaled peripheral continuation informed by Masri et al. (2020). The profile is interpolated in log density using a shape-preserving piecewise cubic interpolant. The rod-density prior and crossover/GC extensions are specified in Supplementary Methods S1. The spatial-unit count sums the five explicitly instantiated populations; polarity states and functional readout streams share these spatial sites.

Table S2: **Experimental units, evaluation settings, and source identities.** The assay map links each supplementary display to its source-level sampling and primary readout. Recording and video labels used in Figure S6 are resolved below.

| Assay | Data and sampling | Readout / primary measurement | Display |
|---|---|---|---|
| Controlled rejoin | 16 fitting backgrounds; 8 evaluation carriers; two current states and two histories | Development-fitted mean-difference scores; paired history margins | Fig. S1A |
| Natural-video rejoin | 15 development and 5 evaluation videos; two clips per source; 40 evaluation snapshots per lag | Mean-difference state/history readouts; accuracy at 125 ms and across suffix lags | Fig. S1B,C |
| Temporal mechanisms | 8 primary scenes, 8 probes each; scans add 4 reserve scenes and six history/ROI settings | Scene-level separation at 66.7 ms; residual RMS and early-to-late $F_T$ change | Fig. S2 |
| Terminal-matched order | 4 fitting families; held HX3/DM2; 3 terminal conditions; 5 evaluation videos | Predicted-axis AUROC at 125 ms; 10 nominal clip pairs per condition | Fig. S3 |
| Selective attenuation | 40 evaluation snapshots; 5 attenuation strengths; 64 random directions per target | Intervened-DEV readout refit; held balanced accuracy at 125 ms | Fig. S4 |
| Representation interfaces | 7 interfaces; the same 15/5 video split and 128-coordinate output allocation | Pooled task AUROCs by lag; within-source $J_s$ at 125 ms | Fig. S5 |
| Primate RGC output | Original held analysis: Rec 3/4; phenotype analysis: Rec 1–4 | Conditional four-class centroid decoding; 50-ms windows; primary 100–150 ms | Fig. S6 |
| Model coding balance | S64/T64; five evaluation videos; 1,000 spatially matched mixed partitions | Conditional binary centroid decoding; branch balance at 125 ms | Fig. S6C,D |

**RGC recording identities and phenotype groups**

| Label | Recording | S-like | T-like | Middle[a] | Undefined[b] |
|---|---|---|---|---|---|
| Rec 1 | `180710_YE` | 47 | 47 | 1 | 0 |
| Rec 2 | `180712_YE` | 41 | 41 | 0 | 1 |
| Rec 3 | `180724_SZ` | 179 | 179 | 1 | 47 |
| Rec 4 | `180808_DK` | 252 | 252 | 1 | 83 |

**Video labels.** Src 1–5 denote *dogs-jump*, *gold-fish*, *judo*, *lab-coat*, and *loading*, respectively. The DAVIS-based assays reuse this evaluation partition. The terminal-matched protocol has ten nominal clip slots corresponding to nine distinct source/start clips, and 60 nominal matched pairs across its six held conditions. Time zero is suffix onset for DAVIS, matched-frame onset for the mechanism assay, and current-fixation onset for RGC decoding.

**RGC grouping.** Persistence is measured from independent full-field step responses in the public recordings (Krüppel et al., 2023a;b). Rec 1/2 are the phenotype-development recordings; Rec 3/4 provide the secondary phenotype comparison and the original held population-readout analysis. [a]One middle-ranked cell is left unassigned when the number of valid cells is odd; the three such cells appear in gray in Figure S6A. [b]Undefined persistence is excluded from that panel. In total, 1,041 finite scores are displayed and 1,038 cells form the balanced S-like/T-like groups.

Table S3: **Native retinal sampling retains comparable joint accessibility with a compact output-site budget.** All three constructions use the same S/T temporal parameters, 64-S plus 64-T readout interface, development-fitted mean-difference readouts, and 125-ms natural-video rejoin checkpoint. $B_{\mathrm{ST}} = N_S + N_T$ counts the S/T output sites before aggregation.

**A. Spatial output-site allocation**

| Construction | $N_S$ | $N_T$ | $B_{\mathrm{ST}}$ |
|---|---|---|---|
| Native | 718,720 | 224,896 | 943,616 |
| All-Cone, native centers | 718,720 | 224,896 | 943,616 |
| All-Cone, dense | 8,294,400 | 8,294,400 | 16,588,800 |

**B. Accessibility at the shared 128-dimensional interface**

| Construction | Relative budget | Current AUROC | Recent AUROC | $J$ |
|---|---|---|---|---|
| Native | $1.00\times$ | 0.9975 | 0.9300 | 0.9300 |
| All-Cone, native centers | $1.00\times$ | 0.9975 | 0.9300 | 0.9300 |
| All-Cone, dense | $17.58\times$ | 1.0000 | 0.9275 | 0.9275 |

The native-center All-Cone condition replaces the receptor substrate while retaining the original S/T center coordinates. The dense condition assigns one S and one T output to every $3840 \times 2160$ receptor location; T inputs retain nearest-six pooling. Relative budget is normalized to the native 943,616 sites. Here $J$ is the minimum of the two pooled task AUROCs, using the same five evaluation videos as the other natural-video rejoin analyses. The similar $J$ values locate the native construction on a much smaller spatial sampling budget at the common task-facing interface.

## S3. LINEAR-READOUT ROBUSTNESS OF FIXED POPULATION REPRESENTATIONS

**Fixed representations and evaluation data.** We examined whether the accessibility of current state and recent history persists across linear readout rules. The analysis uses the saved 128-dimensional snapshots from the two-event natural-video fork–rejoin assay at 125 ms after common-suffix onset. All 15 development videos provide 120 fitting snapshots; the same five evaluation videos provide 40 snapshots, with eight balanced snapshots per video. The retinal parameters and representations remain fixed; only the downstream readouts are estimated. The four within-model conditions are Distinct S/T, Equal $r$, Equal $\tau_a$, and Fully shared dynamics. Retinal Two-IIR is evaluated under the same readout protocol. This is a secondary analysis of the existing source partition.

**Readout rules and fitting.** Current state and recent history are decoded separately from the entire snapshot, without conditioning on the other task label. Raw mean difference (MD) uses the development class-mean difference as its readout direction. Standardized MD first centers and scales each coordinate using the fitting mean and population standard deviation, then computes the same direction. Standard deviations below $10^{-12}$ are replaced by one; mean-difference vectors with norm at most $10^{-12}$ are assigned a zero direction. The third rule is a standardized linear logistic readout. It minimizes the sum of binary cross-entropies plus $\|\mathbf{w}\|_2^2/(2C)$, with an unpenalized intercept. For each representation and task, $C$ is selected from $\{0.01, 0.1, 1, 10, 100\}$ by three-fold source-grouped cross-validation within the development set. Each validation fold contains five complete videos. Selection maximizes the mean validation-source AUROC; scores tied within $10^{-12}$ select the smaller $C$. Standardization is recomputed within every training fold. The final normalizer and readout are then fitted using all 15 development videos and frozen before evaluation.

**Source-resolved joint accessibility.** Let $A_{v,s}$ denote the AUROC for task $v$ within evaluation video $s$. We summarize joint accessibility as

$$J_s = \min(A_{\mathrm{current},s}, A_{\mathrm{recent},s}), \qquad \overline{J} = \frac{1}{5}\sum_{s=1}^{5} J_s. \tag{40}$$

Figure S7 shows the within-model comparisons and the source-specific logistic differences between Distinct S/T and each control. Table S4 reports all five representations under all three readout rules. It also reports pooled joint accessibility, $J_{\mathrm{pool}} = \min(A_{\mathrm{current,pool}}, A_{\mathrm{recent,pool}})$, computed from all 40 evaluation snapshots as in main-text Table 1. Source-macro and pooled AUROC measure within-video and pooled discrimination, respectively, and are reported separately.

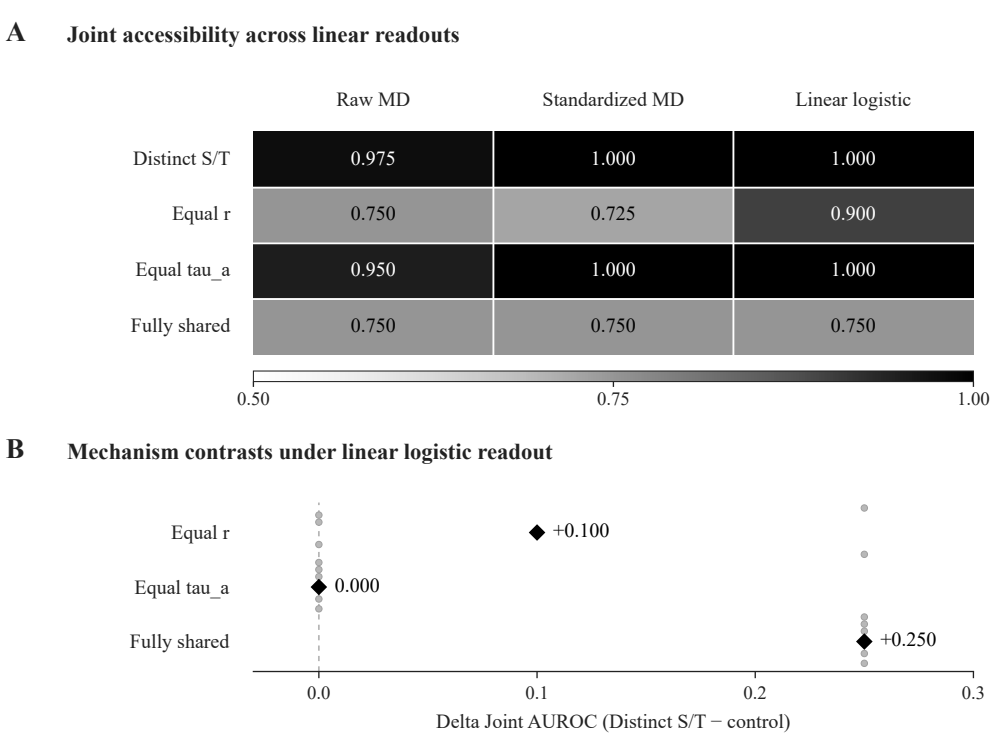



Figure S7: **Readout robustness of S/T temporal-mechanism controls. (A)** Source-macro joint AUROC for four fixed representations and three linear readout rules at 125 ms after common-suffix onset. All readouts use the full 15-video development set. **(B)** Logistic-readout differences, Distinct S/T minus control. Points denote the five evaluation videos; diamonds denote their means. Vertical offsets separate coincident values. Equal $r$ has two differences of 0.250 and three ties (mean 0.100); Equal $\tau_a$ has five ties; Fully shared has five differences of 0.250. The average retention contrast therefore persists after learnable linear reweighting. Table S4 includes Retinal Two-IIR and pooled metrics.

Table S4: **Complete full-development-set comparison across linear readouts.** All five representations use the same 128-coordinate interface, 15-video fitting set, five-video evaluation set, and 125-ms checkpoint. Source-macro $\overline{J}$ averages the within-video task minima; pooled $J_{\mathrm{pool}}$ takes the smaller task AUROC after pooling all evaluation scores. Values are reported to four decimals.

| | Source-macro $\overline{J}$ | | | Pooled $J_{\mathrm{pool}}$ | | |
|---|---|---|---|---|---|---|
| Representation | Raw MD | Standardized MD | Linear logistic | Raw MD | Standardized MD | Linear logistic |
| Distinct S/T | 0.9750 | 1.0000 | 1.0000 | 0.9300 | 0.9775 | 1.0000 |
| Equal $r$ | 0.7500 | 0.7250 | 0.9000 | 0.7100 | 0.7150 | 0.7450 |
| Equal $\tau_a$ | 0.9500 | 1.0000 | 1.0000 | 0.9075 | 0.9800 | 1.0000 |
| Fully shared | 0.7500 | 0.7500 | 0.7500 | 0.7150 | 0.7250 | 0.7125 |
| Retinal Two-IIR | 0.7250 | 0.6750 | 1.0000 | 0.5950 | 0.6250 | 0.9050 |

Source-macro current-state AUROC is 1.0000 for every row/readout combination; thus, in this table, $\overline{J}$ is determined by recent-history accessibility. The raw-MD pooled entries for Distinct S/T and Retinal Two-IIR reproduce the corresponding main-text Table 1 values. These AUROC summaries are distinct from thresholded classification accuracy.