Neural Representation of Temporal Prediction in Mouse Primary Visual Cortex

Saghar Mirbagheri

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:

Nicholas A. Steinmetz, Co-Chair

Rajesh P. N. Rao, Co-Chair

Fred Rieke

Program Authorized to Offer Degree:

Graduate Program in Neuroscience

Abstract

Neural Representations of Temporal Prediction in Mouse Primary Visual Cortex

Saghar Mirbagheri

Co-Chairs of the Supervisory Committee:

Nicholas A. Steinmetz, Department of Neurobiology and Biophysics

Rajesh P. N. Rao, Paul G. Allen School of Computer Science & Engineering

Perception requires the nervous system to infer latent causes from inherently ambiguous sensory measurements that do not uniquely specify the states of the world that produced them. Predictive coding frameworks propose that the brain addresses this ambiguity using an internal generative model that captures learned statistical regularities linking sensory input to latent causes and latent causes to one another. Within this model, each latent variable is represented by combining current sensory input with predictions generated by applying these regularities to estimates of other latent variables. Such predictions may arise from earlier estimates at the same representational level or concurrent estimates at higher levels operating over broader temporal scales. This dissertation investigates how the current stimulus, recent history, and broader sequence structure shape representations in mouse primary visual cortex and how their contributions combine.

The first study examined how earlier activity at the same representational level shaped responses to subsequent input. Patterned optogenetic stimulation imposed controlled sequences of population activity in mouse V1 while holding the current input pattern fixed. Responses depended on the preceding patterns and were enhanced for sequences that matched temporal transitions characteristic of natural vision relative to sequences that did not. This sequence dependence extended beyond directly stimulated neurons and included excitation and suppression within the surrounding population, supporting a contribution from

recurrent circuit dynamics rather than an effect confined to neurons driven by the preceding input. The network therefore filtered input sequences according to the temporal statistics of natural vision, consistent with a prior over how represented variables evolve over time.

The second study examined how predictions derived from broader temporal context shaped V1 representations. In hierarchical predictive coding accounts, higher regions may generate these predictions by integrating information over longer temporal spans and convey them to V1. Mice were exposed to structured visual sequences in which the next element depended on more than the immediately preceding stimulus, while a Neuropixels probe spanning all V1 layers recorded population activity. This allowed identical sensory events to be compared across contexts implying different predicted states. Population representations differed according to the prediction implied by the preceding context, even when the current stimulus and immediate history were held constant. Information about contextual prediction was weak and heterogeneous across individual neurons and distributed across cortical layers rather than confined to a specialized population. Analyses of population geometry indicated that the current stimulus, immediate history, temporal order, and contextual prediction contributed separable components to population structure and combined approximately additively.

Together, these studies show that temporal statistics at two timescales reshape representations in V1, the earliest cortical stage of visual processing. At the shorter timescale, predictions from recent activity were expressed through recurrent dynamics within V1, whereas at the longer timescale, predictions were derived from sequence structure extending beyond the interval over which V1 retained information specific to individual stimuli. The approximately additive contributions observed in the second study are consistent with linear-Gaussian formulations of predictive coding, although the findings do not establish a specific anatomical source or circuit implementation for broader contextual prediction.

Table of Contents

List of Figures and Tables....i
Preface....1
The South Gate....1
AI Usage Statement....1
Acknowledgments....2
Chapter 1: Introduction to Predictive Coding Theory....4
Overview....4
Perception as inference....4
Learning hierarchical representations from statistical regularities....6
Invariance, ambiguity, and joint inference....8
Bayesian inference and generative models....9
Predictive coding as a framework for distributed inference....11
Predictive coding as an algorithm for approximate inference....16
The neural code for inferred variables....25
Summary....27
References....28
Chapter 2: Experimental Assessment of Predictive Coding....31
Overview....31
What would count as evidence for predictive inference....31
Experimental designs for identifying predictive influences....33
Operationalizing inferred variables in population firing rates....38
Disentangling concurrent influences on population representations....39
Overview of the studies....43
Chapter 3 :Recurrent cortical networks encode natural sensory statistics via sequence filtering....46
Summary....46
Introduction....47
Results....51
Sequential order affects V1 neurons' responses through non-target neural suppression....53
Patterns corresponding to natural visual inputs produce larger responses....56
A recurrent neural network model that boosts specific sequences produces the context-dependent responses we see in our data....60
Recurrent network mechanism: one input pattern creates responses in other, non-targeted local neurons, which interact with later input patterns....62
Designed input sequences produce predicted effects confirming the network mechanism.65
Discussion....66
Relation to prior work: structure in spontaneous activity, straightening, cortical suppression, and hippocampal sequences....69
The dense E-E cortical network in sensory cortex may encode the structure of natural

sensation....71
Sequence filtering seems likely to occur in other sensory regions beyond visual cortex.73
Conclusion....74
References....75
Acknowledgments....85
Author contributions....85
Methods....86
Animals....86
Optogenetic experiments....86
Neuropixels experiments....86
Viral injection and cranial window implants....87
Optogenetic experiments....87
Neuropixels experiments....87
Retinotopic mapping....88
Two-photon holographic imaging and stimulation....89
Two-photon data analysis....90
Statistics....91
Cell pattern selection....91
Electrophysiology....92
RNN Model....93
Supplemental figures....95
Chapter 4: Prediction Content Shapes Population Representations in Mouse Primary Visual Cortex....100
Abstract....100
Significance statement....101
Introduction....102
Results....104
Visual sequence paradigm for probing prediction-dependent representations....104
Expectation Violations Modulate Responses to Deviant but Not Extra Familiar Elements..124
Expectation-related response modulation is broadly distributed across cortical layers.130
Prediction context alters neural representations beyond nonspecific response enhancement....132
Prediction content systematically organizes the population geometry....139
Predicted context is represented as a common additive transformation....143
The prediction axis is independent of the sensory representation....145
Prediction-content information is weakly distributed across individual units....146
Single-unit examples dissociate the violation response from content selectivity....153
Discussion....155
Mismatch-related enhancement is not uniquely diagnostic of prediction-specific modulation....156
Prediction content contributes a shared linear component across sensory contexts....163

Limitations....167
Conclusion....169
Author Contributions....169
Data Availability:....170
Acknowledgments:....170
References....171
Methods....178
Animals....178
Surgery....178
Experiments....179
Experimental Flow....179
Electrophysiology....181
Session trimming and trial exclusion....182
Analysis....183
Pupil size extraction....183
Lick Extraction....184
Layer Identification....184
Population-level averages....184
Estimation of Trial-by-trial Amplitude of Response....185
Cluster-based permutation test....186
Population decoding along a shared prediction-context axis....187
Representational similarity analysis of within-family geometry....189
Parallelism and cross-family generalization of the prediction axis....193
Alignment with sensory representations....194
Unit-by-unit ROC analysis....195
Response-difference contrasts....195
Decoder-based neuron ordering....196
Supplementary material....198
Chapter 5: General Discussion....219
What these studies establish....219
Where the constraints come from, and what they carry....221
Dynamics and the form of the update....223
Precision and reliability....223
Behavioral goals and active inference....224
What the code carries....226
Summary....227
References....228

List of Figures and Tables

Figure 1: A hierarchical hidden Markov model with two levels of latent variables..... 14
Figure 1: Using two-photon holographic stimulation and imaging to probe the role of excitatory recurrent connections..... 52
Figure 2: Patterns of input produce different responses in the V1 network depending on the sequential context..... 56
Figure 3: Sequence amplification for visual input, observed by simulating responses to a single frame of a movie..... 59
Figure 4: RNN trained to amplify patterns produces context dependent modulation. 61
Figure 5: Sequences designed to be boosted produce strong and sparse responses, validating a recurrent network mechanism..... 64
Figure 6: Mechanism and conceptual model from this work..... 67
Supplementary Figure 1: Two-photon stimulation responses: timing and individual patterns..... 95
Supplementary Figure 2: Sequential context effects are not dependent on distance to previously stimulated cells..... 96
Supplementary Figure 3: Boosting is not due to spike or indicator saturation..... 97
Supplementary Figure 4: RNN training for both amplified and control networks..... 98
Supplementary Figure 5: Constructed boosted and attenuated sequences have similar mean population responses..... 98
Figure 1: Experimental design for probing prediction-dependent representations in mouse visual cortex..... 106
Figure 2: Sequence-locked V1 population activity with matched behavior and arousal across conditions..... 114
Figure 3: Responses at omission times reflect rebound dynamics with additional experience and expectation-dependent components..... 118
Figure 4: Deviant stimuli elicit context-dependent response enhancement specific to the learned sequence..... 126
Figure 5: Response modulation in structured sequences is selective to deviant stimuli..... 129
Figure 6: Prediction context alters neural population representations beyond nonspecific response enhancement..... 134
Figure 7: Prediction content shapes the within-family representational geometry... 140
Figure 8: Prediction content defines a shared representational direction across sequence families..... 144
Figure 9: Stimulus-position (S3 vs S4) information is not resolved by individual units.. 149
Figure 10: The content decoder is distributed across weakly tuned units..... 152
Figure 11: Single-unit examples dissociate the violation response from content selectivity..... 154
Figure S1: Session-by-session anticipatory licking and recording stability..... 199
Figure S2: Within-session behavioral changes and condition-specific licking and pupil

responses..... 200
Figure S3: Omission amplitude by sequence position and exposure, and control-block omission-like and rebound responses..... 200
Figure S4: Within-session stability, positional distribution, and example sessions for the deviant enhancement..... 203
Figure S5: Mismatch-related response enhancement is broadly distributed across cortical layers..... 204
Figure S6:Response amplitudes are comparable across position-2 and position-3 prediction contexts..... 206
Figure S7:One-back stimulus identity is decodable in V1 and decays by the two-back delay..... 206
Figure S8: No laminar concentration of stimulus-position information..... 208
Figure S9: Response-difference contrasts show bidirectional, unit-level structure..210
Figure S10: The deviance axis is concentrated in a tuned subset..... 211
Figure S11: The omission axis is concentrated in a tuned subset..... 212
Figure S12: Decomposing the within-family representational geometry..... 213
Figure S13: Representational shape is preserved across projection spaces..... 214
Figure S14: Regression diagnostics and model specification across spaces.....215
Table S1: Experimental trial conditions and counts per recording session..... 217
Table S2: Number of recorded units..... 218

Preface

The South Gate

A few days before submitting this dissertation, I dreamed of an immense park. I had been there before, although never quite in the same form. It was a place I somehow belonged to, full of paths, structures, and boundaries that revealed themselves only as I moved through it. There was always more of it to discover.

A two-headed snake rested on my shoulder. One head was large; the other was smaller, a simulation of the first.

I made my way toward the south gate. I wanted to reconnect with a friend from an earlier part of my life, someone I had not seen in years. As I approached the gate, I realized that he was nearby. I went beyond the gate, though I had not quite left the park, and tried to reach him so that he could meet me there.

I knew where he was, and I knew where I was. Yet somehow, to communicate my location, I needed the snake. I was trying to use it as a kind of compass, but it would not work.

As I struggled to make sense of what was wrong, the larger head suddenly turned toward the smaller one. There was a sudden rage in it, something raw and instinctive. I felt the force of it without fearing for myself. It attacked the smaller head and swallowed it.

The snake remained on my shoulder.

I went back through the gate. I was told that the snake had malfunctioned because I had taken it beyond the boundary of the park. We tried to fix it.

AI Usage Statement

The author developed the ideas, scientific arguments, conceptual framework, literature synthesis, interpretations, and intellectual direction of Chapters 1, 2, 4, and 5. Generative AI tools, specifically ChatGPT (OpenAI) and Claude (Anthropic), were used to assist with organization, prose drafting and revision, and improvements to clarity and transitions. All AI-assisted material was critically reviewed and revised by the author. The author assumes full responsibility for the scientific content, citations, and final text.

Acknowledgments

This dissertation was shaped not only by the research it contains, but by the many communities and relationships that sustained me through it.

I first learned how to be a scientist from David Heeger during my master's training. More than anything, David taught me science: how to think clearly, ask meaningful questions, and approach evidence with rigor and curiosity. I carried those lessons into my PhD. Just as importantly, he built my confidence as a scientist and continued to show me support and care long after I left his lab. That confidence remained with me during the most difficult periods of this journey and helped me never give up on myself.

I am grateful to Nick Steinmetz, my doctoral advisor, for the opportunity to pursue this work and for the scientific training and intellectual exchanges that shaped it. I thank Raj Rao, my co-advisor for his intellectual generosity and thoughtful guidance. I am also grateful to all the members of the Steinmetz lab, past and present, for their ideas, feedback, technical help, conversations, and companionship throughout these years.

Beth Buffalo and Adrienne Fairhall each first welcomed me as a rotation student and later served on my dissertation committee. I am grateful to Adrienne for her thoughtful scientific perspective, encouragement, and guidance. I thank Beth for reading my manuscript carefully and for the advice and support she offered during the difficult final years of my PhD, both as a committee member and as department chair.

I owe special thanks to Fred Rieke for the time, care, and seriousness he brought to my work, and for his guidance and support at crucial moments. I am also deeply grateful to Kevin Coffey for joining my committee as GSR when I needed help and for helping make my defense possible.

I thank Ciana Deveau, Zhishang Zhou, Paul LaFosse, Yanting Deng, and Mark Histed for the collaboration that became Chapter 3 of this dissertation, and for the opportunity to contribute to that work.

I am grateful to Anna Gillespie for her mentorship, generosity, and confidence in me, and to Steve Perlmutter for his guidance and support as both a mentor and program director. Andrea Stocco, the first person for whom I taught at UW, trusted my judgment and helped me recognize teaching as an important part of my identity as a scientist. I also thank Andrea Wills and Atom Lesiak for serving as my formal allies, and Astra Bryant for her care, practical help, and steady presence. I am grateful to Dana Miller for generously providing me with a space where I could work when I needed one.

I am grateful to all my colleagues in the Student Representation Committee. Together, we did more than work on policies or respond to problems: we built a community. That community became an important part of my experience at UW and showed me what can be created when students care for one another and work collectively to make an institution better.

I also owe a great deal to my union community, especially Kyle and Vyom. Their advocacy, solidarity, and friendship carried me through moments when I could not have navigated the university alone. Their work reminded me that academic communities are sustained not only through scholarship, but also through people willing to stand beside one another and insist that institutions live up to their values.

I am grateful to all my friends, near and far, who remained present through the many years and transformations that this degree contained. Yeganeh, your friendship has given me continuity, understanding, and a sense of home across distance and time.

Sofia, you began as the postdoctoral mentor supervising my rotation in Beth's lab and somehow became so many other things along the way: a mentor, colleague, ally, and, most importantly, one of my closest friends. You have seen me as both a scientist and a person, including at moments when those two parts of my life could not be separated. Your honesty, loyalty, intellectual companionship, and ability to remain present through difficult years have meant more to me than I can adequately express here.

Finally, I thank my husband, Koosha. As a fellow neuroscientist, you have been my intellectual companion through countless scientific conversations, unfinished ideas, difficult questions, and moments of discovery. But what made this dissertation possible goes far beyond science. Thank you for the care, patience, humor, and love with which you supported me, and for carrying so much when I could not. You lived through every part of this journey beside me, and this accomplishment belongs, in no small part, to you as well.

Chapter 1: Introduction to Predictive Coding Theory

Overview

This thesis's central question is how sensory systems construct and update useful internal representations from ambiguous sensory measurements. Answering this question requires moving across several levels of explanation. In this chapter, I first consider why perception is fundamentally an inference problem, why latent variables provide useful internal descriptions of the world, and how statistical regularities organize them into hierarchical representations. I then formulate the estimation of these variables within Bayesian generative models and discuss predictive coding as a framework and algorithm for approximate distributed inference, alongside alternative computational approaches. Finally, I consider how cortical circuits might realize these computations and which features of neural activity might carry the inferred quantities. The following chapter translates this theoretical framework into empirical criteria and describes the experimental and analytical approaches used to test them in this thesis.

Perception as inference

Sensory systems do not directly measure the variables that matter for behavior. Photoreceptors respond to patterns of photons reaching the retina, cochlear hair cells respond to local mechanical vibrations, and cutaneous mechanoreceptors respond to deformation of the skin. These measurements describe local physical interactions between environmental energy and sensory receptors rather than the distal causes that produced them. For a small number of behaviorally critical variables, evolution has supplied dedicated detectors—receptors tuned to specific chemical signatures, for instance—so that a distal variable is signaled almost directly. Such solutions do not generalize. They require that each variable of interest possess a distinctive proximal signature and that a separate receptor population be devoted to detecting it. Vision affords no comparable shortcut: light arriving at the retina is a projective and lossy

function of the entire visible scene, and essentially nothing of behavioral interest arrives as a dedicated signal. The variables that ultimately guide behavior—where a predator is located, whether a fruit is ripe, or which direction leads home—are therefore latent. They are never directly observed and must instead be inferred from sensory evidence (Kersten et al., 2004; Knill & Pouget, 2004; Pitkow & Angelaki, 2017).

In principle, an organism could dispense with explicit inference and instead learn a direct mapping from sensory measurements to appropriate actions. Such a mapping is not excluded by any computational principle, but it is excluded by finite experience and finite neural resources. The number of distinguishable patterns that can fall on the retina is astronomically larger than the number an animal encounters in a lifetime, and larger still than the number that could be stored by any physically realizable nervous system. Furthermore, a mapping defined over raw sensory patterns would provide no basis for responding to a pattern not previously encountered. Generalization to novel observations therefore requires that sensory measurements be compressed into a far smaller set of explanatory latent variables—object identity, surface layout, or direction of motion, for instance. Compression serves two related purposes: it reduces the dimensionality of the sensory problem while exposing statistical structure that can be reused across previously unseen observations. Such compression necessarily discards information. Because sensory inputs are highly structured, much of what is discarded may be redundant given the rest. However, compression also sacrifices nonredundant information, as illustrated by the reduced representation of fine spatial structure in peripheral vision. What is preserved therefore depends on the value of the resulting variables as intermediate representations, which reflects the statistical regularities they capture, the computations they support at later stages, and their relevance to behavior.

Learning hierarchical representations from statistical regularities

The latent variables are recoverable only because they leave statistical signatures in sensory observations. Common causes manifest as regularities both in the correlation structure of sensory measurements and in how those measurements evolve over time (Barlow, 1961; Fiser et al., 2010; Simoncelli & Olshausen, 2001). Crucially, these regularities are organized across multiple spatial and temporal scales. In the spatial domain, local image statistics define edges and contours, broader regularities define textures and surfaces, and still broader regularities define objects and scenes (Csikor et al., 2025). The temporal domain exhibits an analogous organization: rapidly changing sensory events unfold within slower object and behavioral dynamics, which themselves occur within still slower environmental regularities (Friston et al., 2018; Pezzulo et al., 2022). Regularities at each scale are therefore conditioned by regularities operating over broader spatial and temporal scales. Through experience, nervous systems learn these regularities, allowing latent causes to be inferred from observations that would otherwise be uninterpretable.

A nervous system attempting to exploit these learned regularities must construct internal representations capable of capturing them. The question then becomes which latent variables are worth representing. Not every latent variable capable of explaining sensory observations is equally useful to represent. Useful representations capture latent explanatory factors underlying the observed statistical regularities while supporting the computations performed by later stages (Bengio et al., 2013). Consequently, which variables become represented depends jointly on the statistical structure of sensory observations, the computations that later stages must perform, and ultimately the behavioral demands placed upon the organism. The resulting hierarchy of representations is therefore informed by the statistical structure present in sensory observations but is not simply a copy of the external world's organization.

Variables that explain only a handful of highly specific observations provide little advantage over the observations themselves, whereas variables capturing regularities shared

across many observations are extracted more readily during learning and generalize more effectively to novel situations (Barlow, 1961; Saxe et al., 2019). Such variables also provide a more effective substrate for subsequent inference. Once explanatory variables have been learned, they can be reused and recombined to construct increasingly complex representations rather than being rediscovered from raw sensory measurements. Organizing representations hierarchically therefore promotes both the reuse of previously learned factors and the emergence of progressively more abstract variables, allowing higher-order regularities to be represented without relearning the statistical structure already captured at lower levels (Bengio et al., 2013). At every stage of the hierarchy, a representation consequently serves two complementary roles. It explains the activity at the level below while simultaneously providing the explanatory variables from which the next level constructs its own representations. Representations at one level therefore become both the evidence available to the level above and the contextual information through which lower-level observations are interpreted.

Objects provide perhaps the clearest illustration of what such an explanatory variable looks like. Nothing in the retinal image explicitly identifies where one object ends and another begins, yet object identity provides a remarkably useful explanation of sensory experience. Object identity is not a property read directly from sensory input but a hypothesis about a common cause capable of explaining a large collection of otherwise disparate observations. Three properties make such hypotheses computationally valuable. First, they are compressive: a single explanatory variable accounts for many correlated observations at once. Second, they are stable, in that the cause persists while its sensory consequences change, so an estimate of it remains valid over the timescales on which perception and behavior unfold. Third, they are reusable: a single latent variable can support many different behaviors and be grouped with other variables in different ways depending on the task at hand, rather than being tied to a single sensory pattern or behavioral computation (DiCarlo & Cox, 2007). Object identity is therefore valuable not only because it explains sensory observations, but because it serves as

an efficient intermediate representation from which progressively more abstract representations can be constructed. Importantly, this is a claim about the usefulness of an explanatory model rather than its ontological correctness. Internal variables need not partition the world exactly as it is physically organized; they need only organize sensory observations in a manner that supports successful inference, generalization, and adaptive behavior (Pezzulo et al., 2022).

Invariance, ambiguity, and joint inference

The relationship between environmental causes and sensory measurements is neither one-to-one nor invertible, and this single fact imposes two distinct computational demands. Read in the forward direction, one cause produces many measurements: the same object generates dramatically different retinal images across changes in viewpoint, illumination, distance, and partial occlusion. Recovering the cause across this variation requires invariance, which is naturally achieved through compression—discarding the dimensions along which the behaviorally relevant variable does not vary. This is a many-to-one mapping, and many-to-one mappings are functions. They can in principle be computed by successive convergence, with each stage pooling over a broader extent of the level below, such that the estimate of a higher-level variable comes to depend on the joint pattern of evidence at the level beneath it.

Read in the reverse direction, however, one measurement is consistent with many causes. Here the correct estimate is not determined by the sensory input at all, and no mapping from that input alone can recover it. Disambiguation requires information from elsewhere: other latent variables at the same or higher levels, or estimates carried forward from the recent past. Because those variables are themselves inferred rather than given, the constraints they provide run in both directions. Lower-level evidence informs higher-level estimates through convergence, while higher-level estimates constrain the interpretation of lower-level observations in turn. The same nested statistical structure that favors hierarchical representations therefore also induces statistical dependencies among the corresponding latent

variables, and estimates of rapidly changing variables depend upon slower latent dynamics that are themselves continuously refined as new observations become available. Latent variables distributed across space and time cannot generally be inferred independently. Rather, the central computational problem of perception becomes estimating an interacting hierarchy of latent variables from noisy and incomplete sensory observations while exploiting the statistical regularities learned through experience. Bayesian inference provides a principled framework for describing this estimation problem.

Bayesian inference and generative models

The preceding sections identify the computational problem faced by perception but not how it should be solved. If perception consists of estimating latent variables from noisy, incomplete, and ambiguous sensory observations, then how should those estimates be formed? Bayesian inference provides a normative framework for answering this question by describing how current sensory evidence should be combined with statistical knowledge acquired through previous experience (Knill & Pouget, 2004).

If the brain performs Bayesian inference, it must possess, either explicitly or implicitly, a generative model of the environment. A generative model specifies a probabilistic description of how latent variables give rise to sensory observations and relate to one another. Formally, it defines a joint probability distribution over latent variables and sensory observations,

$$p(x,\, z) \;=\; p(x \mid z)\, p(z),$$

where z denotes latent variables, and x denotes sensory observations. The likelihood p(x | z) describes how latent causes give rise to sensory evidence, whereas the prior p(z) summarizes previously acquired knowledge about the statistical structure of those causes. Perception then consists of estimating the posterior distribution,

$$p(z \mid x) = p(x|z)p(z)/p(x) \quad (p(x) \neq 0)$$

which combines current sensory evidence with prior knowledge to infer the probability of latent variables that have generated the observations. Inference can therefore be understood as an inversion of the generative model: rather than generating observations from latent variables, the observer uses the observations to estimate the probability of latent variables from which they arose.

The hierarchical structure developed in the preceding sections makes this inference problem more complex. Latent variables at different spatial and temporal scales are statistically dependent, such that evidence about one variable can alter the probability assigned to others. The relevant quantity is therefore not generally a collection of independently estimated posteriors, but a joint posterior over an interacting set of latent variables,

$$p(z_1, z_2, \cdots, z_l \mid x).$$

Local sensory ambiguity may be resolved by information about broader contextual variables, while estimates of those broader variables are themselves supported and revised by lower-level sensory evidence. Likewise, estimates of rapidly changing variables may depend on slower latent dynamics, while those slower states are themselves inferred from observations accumulating over time. The optimal estimate of any one latent variable therefore depends, in general, on the evidence available about the others.

This formulation provides a normative target, but it does not require that the nervous system explicitly represent the complete joint posterior or compute every marginal posterior at every stage of processing. A system concerned with a restricted behavioral variable could instead preserve simpler intermediate representations and integrate the relevant evidence only at a later stage. Such a strategy may be entirely adequate when a limited set of mappings reliably produces successful behavior. At the other extreme, maintaining inferred estimates of

latent variables throughout the hierarchy would provide a richer description of the current state of the environment, but would also require substantially more computation.

This tradeoff connects directly to the role of intermediate representations developed above. Variables at different levels of a hierarchy can themselves be behaviorally relevant, rather than serving only as intermediate steps toward a single final estimate. Color may determine behavior in one context, object identity in another, and a more abstract category or behavioral significance in another. Maintaining inferred estimates at multiple levels therefore has a potential computational advantage: the same learned hierarchy can provide reusable estimates at different spatial, temporal, and abstract scales, allowing behavior to flexibly access whichever variables are relevant to the current task. Moreover, because the variables are statistically dependent, incorporating information from the rest of the hierarchy can improve the estimate available at any particular level rather than merely improving a final downstream decision.

There is nevertheless no requirement that a biological system attain this normative ideal. Exact inference over the joint posterior of a complex generative model rapidly becomes computationally intractable, and simpler approximations or task-specific heuristics may provide sufficiently accurate behavior at substantially lower cost. The empirical question is therefore not whether the brain performs exact Bayesian inference, but which aspects of the joint inferential problem are actually reflected in neural computation and representation. Predictive coding provides one influential framework in which inference is distributed across the hierarchy rather than confined to a final downstream readout.

Predictive coding as a framework for distributed inference

In its broadest sense, predictive coding proposes that perception is accomplished through distributed inference within a hierarchical generative model (Friston & Kiebel, 2009; Rao & Ballard, 1999). Rather than treating early sensory representations as fixed feedforward

descriptions whose interpretation occurs only in downstream regions, each level of the hierarchy maintains an estimate of the latent variables represented at that level. These estimates form components of a coupled inferential solution: lower-level estimates are constrained by sensory evidence as well as by information carried by other latent variables, while higher-level estimates are themselves constrained by the evidence represented below. The inferred state at any level can therefore differ from the estimate that would be obtained from its immediate sensory input alone.

This distinction is particularly important when sensory evidence is ambiguous. A feedforward transformation can progressively construct invariant representations by pooling across variations that are irrelevant to a particular latent variable, but when the available evidence is compatible with multiple interpretations, the correct estimate cannot be determined from that evidence alone. Information represented elsewhere in the hierarchy can then constrain the interpretation of the local evidence. Hierarchical latent-variable models make this relationship explicit: the approximate posterior over a lower-level variable can depend jointly on the sensory observation and an inferred higher-level variable, rather than on the sensory observation alone. Recent hierarchical generative models of visual cortex factorize inference in exactly this way, estimating a V1-like latent variable conditional on both the image and an inferred V2-like latent variable (Csikor et al., 2025).

The same principle extends across time. The current value of a latent variable may be constrained not only by variables represented at higher levels, but also by estimates formed from previous observations. Considering a dynamic hierarchical model, and suppressing the required marginalization over uncertain latent variables for notational simplicity, the posterior at time *t* can be written schematically as

$$p(z_t|x_t,\ z_{t-1},\ z_{upper}) \propto p(x_t|z_t)p(z_t|z_{t-1}, z^{upper}).$$

The first term describes the evidence supplied by the current sensory observation. The second describes the expectation for the current latent state given temporal history and hierarchical context. This factorization assumes that once the current latent variable z_t is known the current observation x_t is conditionally independent of previous and higher-level latent variables, z_{t-1} and z^{upper} . The previous and higher-level states are therefore related to the current sensory observation only through their relationship with z_t, the latent cause of that observation. During inference, estimates of z_{t-1} and z_{upper} constrain the estimate of z_t, rather than affecting the sensory observation itself. These relationships can be represented by a graphical model in which latent states are connected across time and across hierarchical levels, while each sensory observation depends on the current latent state.

The structure of the graphical model representative of this process is a form of hierarchical hidden Markov model. The model is hierarchical because it contains interacting latent levels, hidden because these states are inferred rather than directly observed, and Markovian because the current joint state summarizes the past information needed to predict the next state. Temporal and hierarchical information can therefore constrain the same latent estimate while carrying different information about it: previous states provide information derived from temporal regularities, whereas higher-level states can provide contextual information operating over broader spatial, temporal, or abstract scales.

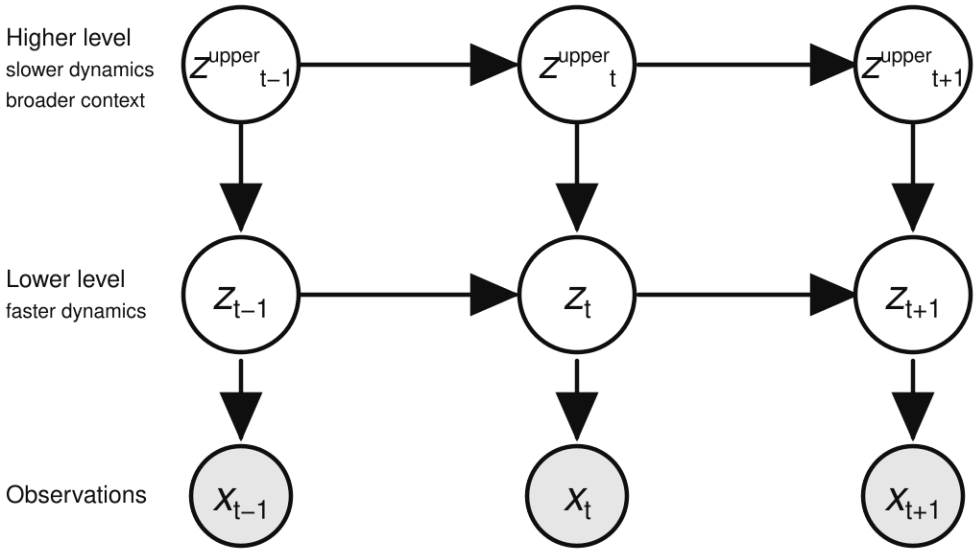


Figure 1: A hierarchical hidden Markov model with two levels of latent variables.
Hidden states evolve across time at both levels, and each lower-level state generates the corresponding sensory observation. The model is Markovian because the joint latent state at one time contains the past information needed to predict the state at the next time. Slower upper-level dynamics nevertheless allow contextual information to be retained over a broader temporal interval.

This is the sense in which the term prediction will be used throughout this thesis. Prediction does not primarily refer to forecasting future sensory observations. Rather, it refers to an estimate of one latent variable conditioned on information supplied by other variables within the generative model. Expected sensory observations follow as a special case, because an estimate of a latent cause implies expected states at lower levels and ultimately expected sensory input. Prediction can therefore operate both across hierarchical levels and across time. In both cases, information unavailable from the current sensory input alone can constrain the inferred state.

Importantly, the variables supplying these constraints are themselves uncertain estimates. A previous latent state is estimated from earlier observations, and a higher-level contextual variable is likewise inferred rather than known. Recursive estimation of the current latent state from observations available up to the current time is known as filtering: the posterior

from the previous time step is marginalized over the temporal transition model to yield a prediction for the current state, which is then updated using the current observation. Uncertainty from previous inference is therefore carried forward through time, while hierarchical inference propagates uncertainty across levels. The expression above suppresses these marginalizations for clarity, but the underlying problem is one of jointly estimating mutually dependent and uncertain latent variables rather than deterministically propagating known states.

Distributed inference offers the computational advantages anticipated by the hierarchical representations described earlier. If each level maintains an estimate of its own latent variables after incorporating the relevant constraints available elsewhere in the model, then intermediate representations remain useful in their own right rather than functioning only as transient steps toward a particular downstream output. The same learned variables can support different behavioral readouts as task demands change, while also serving as evidence for inference at other levels. Maintaining inferred states throughout the hierarchy therefore provides flexibility: information learned because it explains one set of observations can be reused to interpret others and can support behaviors requiring access to different levels of abstraction.

However, distributed inference is not the only architecture capable of producing successful behavior. A system could preserve relatively feedforward intermediate representations and integrate contextual information only at downstream stages relevant to the current behavioral objective. Similarly, a sufficiently powerful feedforward system could learn an amortized approximation to an inferential solution without performing online recurrent refinement, and the distinction is not absolute in any case, since an iterative computation can be unrolled into a sequence of feedforward stages with stage-specific parameters. Such strategies may be computationally cheaper and may be entirely sufficient when the relevant tasks and environmental contingencies are restricted. The question is therefore one of degree rather than an all-or-none opposition between inference and simpler computation: how much of the

information theoretically available from the joint generative model is actually incorporated into latent-variable estimates at each stage of sensory processing?

This distinction generates an experimentally testable prediction. If early sensory representations constitute inferred estimates within a distributed hierarchy, then information relevant to those estimates should alter the representations themselves even when that information originates outside the immediate sensory input. In particular, statistical information acquired from preceding sensory states or from latent variables operating over broader temporal scales should modify the representation of the current sensory state in early sensory cortex. By contrast, architectures in which such information is integrated only by downstream readouts can produce context-dependent perception or behavior without requiring corresponding changes in early sensory representations.

The framework does not, however, specify how these coupled estimates are computed. Exact inference is generally intractable, and different approximations can preserve different aspects of the ideal Bayesian solution while incurring different computational costs. The canonical, or "vanilla," predictive-coding algorithm provides one particular solution, approximating distributed inference through iterative local interactions between latent-state estimates and prediction errors. This algorithm makes stronger and more specific commitments than the framework itself, and those commitments can be tested separately from the broader claim that inference is distributed throughout the sensory hierarchy.

Predictive coding as an algorithm for approximate inference

The framework described above specifies a distributed inferential objective but does not uniquely determine the algorithm by which it is achieved. Multiple algorithms can approximate inference in hierarchical generative models, and not all systems that incorporate prior information into intermediate representations therefore implement predictive coding. Here, I use predictive coding in a narrower sense to refer to the canonical, or "vanilla," algorithm in which

latent-state estimates are iteratively optimized to reduce the discrepancy between states predicted by the generative model and those supported by the available evidence (Friston & Kiebel, 2009; Millidge et al., 2021; Rao & Ballard, 1999).

One simplification is to represent the posterior by a point estimate rather than the complete probability distribution. In the canonical formulation, inference seeks the maximum a posteriori (MAP) estimate: the configuration of latent variables that maximizes posterior probability under the generative model. This discards the covariance structure of the posterior, retaining the most probable configuration rather than the full distribution over configurations. Under linear-Gaussian assumptions, including Gaussian priors and observation noise, the posterior is itself Gaussian, so the retained mode coincides with the posterior mean, and the MAP solution can be obtained by iterative optimization of the negative log posterior (Jiang & Rao, 2022; Millidge et al., 2021; Rao & Ballard, 1999).

Uncertainty remains consequential even when the posterior is reduced to a point estimate. Under Gaussian assumptions, the negative log posterior is a sum of squared discrepancies, each scaled by the inverse variance of the corresponding source of information. Precision therefore enters the objective being optimized rather than being imposed as an additional criterion: a discrepancy associated with a low-variance source contributes more to the objective, and consequently exerts greater influence on the estimate that optimizes it, than an identical discrepancy associated with a high-variance source. A weak or unreliable sensory signal should accordingly produce greater dependence on prior information than an otherwise identical but highly reliable observation.

The linear-Gaussian case makes another property of the algorithm particularly transparent. When the relevant precisions are held fixed, distinct constraints on the same latent variable contribute additively to the objective and therefore to its update. For a latent state constrained simultaneously by current sensory evidence, temporal history, and hierarchical context, the update can be written schematically as the sum of terms reflecting each source.

Hierarchical and temporal predictions can therefore carry different content while contributing to the estimate of the same latent variable through approximately linear combination. This linearity is not a general requirement of Bayesian inference or of distributed hierarchical inference; it follows from the particular assumptions of the canonical approximation and therefore constitutes a more specific algorithmic prediction. Its testability, however, depends on precisions remaining approximately stable across the conditions being compared. Manipulations that alter how predictable a stimulus is generally alter both the content of the predictive prior and its precision, and an apparent departure from additive combination could therefore reflect a change in weighting rather than in the form of the combination itself. Isolating the algebraic form of the combination requires comparing conditions in which prediction content differs while the reliability of the prediction remains comparable.

Inference in this formulation is also iterative and local. An initial estimate need not already incorporate all information available elsewhere in the generative model; instead, estimates are repeatedly revised as constraints propagate through the hierarchy until the system approaches a mutually consistent solution. The locality is a property of the update itself. Solving for the joint posterior as a global optimization problem would require simultaneous access to variables distributed across the hierarchy. In the local iterative formulation, by contrast, revising the estimate at one level requires only the estimates of the variables to which it is directly coupled in the generative model and the discrepancy between them, so no individual update requires direct access to variables elsewhere in the hierarchy. Globally coherent inference emerges from repeated interactions of this restricted kind. Rao and Ballard's formulation already incorporated temporal dynamics through its relationship to Kalman filtering, such that estimates of the current state could be constrained both by hierarchical context and by information propagated from previous states (Rao & Ballard, 1999). Hierarchical and temporal influences are therefore not separate forms of prediction within the algorithm; they provide different constraints entering the iterative estimation of a common latent state.

These properties distinguish the canonical algorithm from the broader framework of distributed inference developed above. Evidence that temporal or hierarchical context alters an early sensory representation supports the claim that inference is distributed across the hierarchy, but does not by itself establish predictive coding as the underlying algorithm. Stronger evidence requires testing properties of the transformation by which those sources of information are integrated. The approximately additive combination predicted by the linear-Gaussian approximation provides one such property: it concerns not merely whether prior information reaches a sensory representation, but the algebraic relationship by which multiple constraints contribute to the inferred state.

Other algorithms can pursue a similar distributed inferential objective while making different assumptions about how constraints are combined. In predictive coding/biased competition with divisive input modulation (PC/BC-DIM), reconstructions modulate lower-level activity multiplicatively rather than being subtracted from it, such that predictions scale rather than offset the evidence on which they act (Spratling, 2008, 2017). A separate proposal derives cortical computation from an inference objective and arrives at normalization as its canonical operation, so that the influence of any input depends divisively on the activity of the surrounding population rather than entering as an independent additive contribution (Heeger, 2017). More general message-passing accounts further demonstrate that probabilistic inference need not produce linear combinations even in relatively simple graphical models: updates to one variable can depend jointly and nonlinearly on the estimated values and uncertainties of the variables constraining it (Pitkow & Angelaki, 2017).

A related distinction concerns how top-down information enters the computation at all. Contextual input may contribute directly to the estimate of a lower-level variable, shifting it independently of its current value, or it may modulate the dynamics by which that estimate evolves, changing the mapping between the current state and its subsequent trajectory. In dynamic predictive coding, a higher-level state generates modulatory weights that select among

learned transition matrices, so that the same lower-level state can evolve according to different effective dynamics depending on the inferred context (Jiang & Rao, 2024). These alternatives can all produce context-dependent representations consistent with distributed inference, but they predict different relationships among the sources of information being combined, and observing that context influences a lower-level representation does not by itself distinguish them.

A further class of alternatives concerns not how constraints are combined but the form in which the posterior and its uncertainty are represented. In sampling-based schemes, a posterior distribution can be represented through the distribution of neural activity over time rather than by a single point estimate (Hoyer & Hyvärinen, 2003; Orbán et al., 2016), whereas probabilistic population codes encode probabilistic information in patterns of activity across neuronal populations (Ma et al., 2006; Pouget et al., 2013). These distinctions need not determine whether higher-level or temporal information influences a lower-level representation, but they change what information about the posterior is available in neural activity and how it can be used by subsequent computations. Thus, observing an influence of prior information on sensory representations is insufficient to identify a particular algorithm. Distinguishing among alternative algorithms requires asking more specifically how multiple sources of information are combined, how uncertainty influences that combination, and how the resulting estimates evolve over time.

The algorithmic description nevertheless remains abstract with respect to its physical realization. Local iterative optimization does not by itself specify which neuronal populations represent the relevant quantities, how those quantities are communicated between cortical areas or layers, or whether the discrepancies required by the optimization are explicitly represented by dedicated neurons. These questions concern the neural implementation of predictive coding and impose a further set of empirical commitments beyond those of either the distributed-inference framework or the algorithm itself.

Neural implementation of predictive coding

The algorithmic description above specifies the quantities that must interact during inference but does not uniquely determine how those quantities are instantiated by neural circuits. In particular, the presence of a discrepancy term in an iterative optimization does not require that this discrepancy be explicitly represented by a dedicated neuronal population. A neural circuit could, in principle, implement the same update through mixed representations, recurrent population dynamics, dendritic computations, or interactions among activity subspaces without assigning each computational quantity to a distinct cell class. Canonical neural implementations of predictive coding make a stronger proposal: latent-state estimates and discrepancies between predicted and observed activity are represented by partially distinct neural populations, whose reciprocal interactions implement the iterative updates described above (Friston & Kiebel, 2009; Rao & Ballard, 1999).

In the simplest version of this proposal, populations representing latent-state estimates generate predictions of the activity expected at the level below. The difference between predicted and observed activity is represented as a prediction error, which provides the information required to revise the corresponding latent-state estimates. Prediction errors are signed because observations may either exceed or fall below their predicted values. Because neuronal firing rates cannot be negative, canonical implementations often assign each sign to separate populations. This division into positive and negative prediction error populations was present in the original formulation and has been elaborated in subsequent accounts (Keller & Mrsic-Flogel, 2018; Rao & Ballard, 1999). Reciprocal interactions among state and error populations allow updated estimates to generate new predictions, producing the recurrent dynamics required for iterative inference. Learning operates over a longer timescale: discrepancies that update latent states during inference can also modify the synaptic

parameters of the generative model, allowing predictions themselves to change as environmental regularities are learned.

This organization has been mapped onto the recurrent and hierarchical anatomy of the cortex. Feedforward and feedback projections connect cortical areas reciprocally but exhibit different laminar origins and termination patterns, motivating proposals in which ascending and descending pathways carry different computational quantities. In the canonical microcircuit account of predictive coding, superficial pyramidal populations and the feedforward projections arising from them have been associated with prediction-error signals, while deep pyramidal populations and their descending projections have been associated with latent-state estimates and the predictions generated from them (Bastos et al., 2012; Friston & Kiebel, 2009). Within each area, recurrent excitatory and inhibitory circuitry could then perform the local comparisons and state updates required by the algorithm. These anatomical correspondences are hypotheses about implementation rather than necessary consequences of predictive coding, however, and the mapping between particular computational quantities, cortical layers, and neuronal populations remains substantially less established than the existence of reciprocal feedforward and feedback connectivity itself.

The diversity of cortical inhibitory interneurons provides additional candidate mechanisms for implementing these computations. Parvalbumin-positive (PV), somatostatin-positive (SST), and vasoactive intestinal peptide-positive (VIP) interneurons differ in their connectivity and in the compartments of pyramidal neurons that they regulate, allowing recurrent circuits to control the gain and integration of feedforward, recurrent, and feedback inputs in different ways. Such circuitry could contribute to comparisons between expected and observed activity, regulate the effective influence of contextual input, or adjust the weighting applied to different sources of evidence.

This last possibility deserves emphasis, because precision cannot generally be identified with the magnitude of a prediction-error signal itself. A large discrepancy carrying low precision

may influence an estimate less than a smaller discrepancy carrying high precision, so the size of a mismatch response is not by itself a measure of the influence that mismatch exerts on inference. Candidate implementations of precision therefore involve mechanisms that alter the gain or effective synaptic weighting of discrepancy-related information, including inhibitory control and neuromodulatory influences. The resulting neural signal may consequently reflect both the magnitude of a discrepancy and the precision with which it is weighted. Measurements of response magnitude alone therefore underdetermine both the discrepancy and the weight assigned to it.

Temporal inference introduces a related but partly distinct implementation problem. Information about previous latent states need not arrive through feedback from a hierarchically higher cortical area. Recurrent connectivity within a cortical area can allow previous population states to influence subsequent activity, providing a substrate for temporal predictions based on recently observed regularities; local recurrent interactions in mouse V1 have been shown to alter responses to an input pattern depending on the patterns that preceded it, on timescales of tens to hundreds of milliseconds (Deveau et al., 2026). Contextual variables operating over longer timescales may additionally recruit interactions with areas whose dynamics integrate information over broader temporal windows and that can influence earlier sensory areas through feedback. Local recurrence and hierarchical feedback therefore provide anatomically distinct routes by which information unavailable in the current feedforward input can modify a sensory representation. Both can implement components of the same distributed inferential computation even though the source and characteristic timescale of the contextual information may differ.

The physical implementation of contextual modulation also need not correspond to the simple addition of two activity patterns. The distinction introduced at the algorithmic level between changing an estimate and changing the dynamics of that estimate has several possible neural realizations. Feedback may alter synaptic gain, recurrent coupling, dendritic integration, or the effective state-space dynamics of a population, thereby changing how subsequent

sensory input is transformed rather than simply adding a fixed response to it. Compartmentalized dendritic processing provides one possible substrate: feedforward and feedback inputs arriving onto different dendritic compartments could interact nonlinearly before influencing somatic output. At the population level, contextual information could similarly alter the subspace or dynamical regime in which recurrent activity evolves without requiring a dedicated population whose firing rate directly represents the contextual contribution. Such implementations could realize the context-dependent transition functions discussed above while producing neural signatures quite different from a literal additive prediction signal.

These possibilities illustrate why identifying prediction-error-like responses is insufficient to establish a particular implementation of predictive coding. Sensory mismatch is a computationally useful signal in many settings, and responses to unexpected events can arise through multiple circuit mechanisms. Stimulus-specific adaptation, recurrent inhibition, disinhibitory circuits, neuromodulatory signals, and interactions involving thalamic and corticothalamic pathways can all produce responses that depend on whether an observation matches recent or contextual expectations. Conversely, a circuit performing predictive inference need not produce a population whose activity can straightforwardly be labeled a prediction error. The existence of mismatch responses and the implementation of predictive coding are therefore related but separable questions.

The resulting hierarchy of claims is important experimentally. A change in an early sensory representation produced by temporal or hierarchical context provides evidence for distributed inference; the algebraic relationship by which multiple constraints combine can distinguish among candidate algorithms; and the populations, pathways, dynamics, and synaptic mechanisms required to produce those transformations concern their neural implementation. Evidence at one level therefore constrains, but does not by itself establish, claims at the others.

The neural code for inferred variables

The preceding sections describe what should be estimated, how those estimates might be computed, and how the computation might be realized by cortical circuits. A further question determines how any of these levels can be tested experimentally: which feature of neural activity carries the quantities in question. A latent variable is a construct of the computational model, and its neural representation is whatever pattern of activity encodes it. Identifying that pattern is not incidental to testing predictive coding, because every measurement presupposes an answer. An analysis of trial-averaged firing rates can only detect influences of prediction that appear in trial-averaged firing rates, and a null result under one assumed code constrains other codes only weakly.

The most common assumption is that inferred quantities are carried by firing rates, either of individual neurons or of populations. Rate codes have the advantage of being straightforwardly measurable and of connecting naturally to the algorithmic descriptions above, in which latent states and discrepancies are real-valued quantities that can be identified with mean activity levels. They are also the basis for most reported evidence of predictive processing, including mismatch responses, omission responses, and expectation-dependent modulation of stimulus-evoked activity.

Rate is not the only candidate. An influential alternative holds that contextual and prior information may be carried not by discharge rate alone but by temporal relationships among spikes—through synchrony, oscillatory phase, and correlation structure (Singer, 2021). In such accounts, firing rates primarily signal the presence and strength of sensory features, while temporal coordination provides an additional dimension through which contextual information and grouping can be represented. Coherently firing groups of neurons can thereby signal which cells are currently participating in a common interpretation. A closely related proposal treats the functional unit of cortical representation as the neuronal assembly rather than the individual neuron, with inference corresponding to the activation and competition of assemblies rather

than solely to graded changes in single-cell activity (Yuste et al., 2024). These proposals matter here for a specific reason: information carried in correlation structure or assembly membership can vary while mean rates remain unchanged, so measurements confined to rates could fail to detect a prediction-dependent signal that is nonetheless present.

A further alternative is sequence-based coding, in which information is carried by the relative temporal order in which neurons within a population become active rather than by their firing rates alone. Neuronal populations can exhibit brief bursts in which participating neurons fire in stereotyped sequences, providing a combinatorially rich code in which different information can be represented by different firing orders using relatively few spikes. Recent recordings from human anterior temporal cortex, for example, showed that the relative order of neuronal firing within population bursts carried information about visual categories and individual exemplars that was separable from and complementary to information carried by firing rates (Xie et al., 2024). Such a code is particularly relevant to predictive processing because contextual information could alter which sequence of population activity unfolds, even when its effect on spike counts within a fixed analysis window is small. A predicted state might therefore be expressed not as a displacement of a static population-rate representation, but through the ordering of neuronal activation by which that state is reached. Such a code is naturally suited to representing dynamical structure, and is therefore relevant to the algorithmic distinction drawn above between context that shifts a current estimate and context that modifies the dynamics by which that estimate evolves, although observing a sequence code alone would not distinguish between these possibilities. It is also directly relevant to the present work, since the paradigm used here establishes temporal regularities across sequences of stimuli; exploratory analyses of sequence-level structure in this dataset were pursued during the course of this thesis but are not reported here.

These proposals concern the neural dimension in which information is encoded. A partly separate question is what information about the posterior that code carries. If activity encodes

only a point estimate of the inferred variable, trial-to-trial variability need not itself represent posterior uncertainty. If activity samples the posterior, by contrast, variability is part of the representational scheme, and its structure across trials should reflect properties of the posterior (Hoyer & Hyvärinen, 2003; Orbán et al., 2016). If populations encode distributional parameters directly, information about uncertainty may instead be recoverable from the pattern of population activity on a single trial (Ma et al., 2006; Pouget et al., 2013). These schemes therefore assign different interpretations to the same neural measurements. For example, changes in response variability may reflect changes in posterior uncertainty under a sampling account, whereas variability need not carry this interpretation under a point-estimate account.

Summary

The theoretical framework developed in this chapter distinguishes the computational problem of perceptual inference, the algorithms that may approximate it, their possible neural implementations, and the neural activity through which inferred variables may be represented. This framework does not by itself specify which empirical observations would distinguish predictive inference from other forms of contextual modulation. The following chapter translates these theoretical claims into experimental criteria and describes the design and analytical choices used to test them in this thesis.

References

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). MIT Press.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695–711.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(8), 1798–1828.

Csikor, F., Meszéna, B., Ócsai, K., & Orbán, G. (2025). Top-down perceptual inference shaping the activity of early visual cortex. *Nature Communications*, *16*(1), 9998.

Deveau, C. E., Zhou, Z., LaFosse, P. K., Deng, Y., Mirbagheri, S., Steinmetz, N., & Histed, M. H. (2026). Recurrent cortical networks encode natural sensory statistics via sequence filtering. *Neuron*, *114*(8), 1489-1503.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*(8), 333–341.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences, 14*(3), 119–130.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1211–1221.

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews, 90*, 486–501.

Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences, 114*(8), 1773–1782.

Hoyer, P. O., & Hyvärinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior. In *Advances in Neural Information Processing Systems* (Vol. 15, pp. 293–300). MIT Press.

Jiang, L. P., & Rao, R. P. N. (2022). Predictive coding theories of cortical function. *Oxford Research Encyclopedia of Neuroscience*. Oxford University Press.

Jiang, L. P., & Rao, R. P. N. (2024). Dynamic predictive coding: A model of hierarchical sequence learning and prediction in the neocortex. *PLOS Computational Biology, 20*(2), e1011801.

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron, 100*(2), 424–435.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271–304.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*(11), 1432–1438.

Millidge, B., Seth, A., & Buckley, C. L. (2021). *Predictive coding: A theoretical and experimental review* (arXiv:2107.12979). arXiv.

Orbán, G., Berkes, P., Fiser, J., & Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron, 92*(2), 530–543.

Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. *Philosophical Transactions of the Royal Society B: Biological Sciences, 377*(1844), 20200531.

Pitkow, X., & Angelaki, D. E. (2017). Inference in the brain: Statistics flowing in redundant population codes. *Neuron, 94*(5), 943–953.

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience, 16*(9), 1170–1178.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2019). A mathematical theory of semantic development in deep neural networks. *Proceedings of the National Academy of Sciences, 116*(23), 11537–11546.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193–1216.

Singer, W. (2021). Recurrent dynamics in the cerebral cortex: Integration of sensory evidence with stored knowledge. *Proceedings of the National Academy of Sciences, 118*(33), e2101043118.

Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience, 2*, 4.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition, 112*, 92–97.

Xie, W., Wittig Jr, J. H., Chapeton, J. I., El-Kalliny, M., Jackson, S. N., Inati, S. K., & Zaghloul, K. A. (2024). Neuronal sequences in population bursts encode information in human cortex. *Nature*, *635*(8040), 935-942.

Yuste, R., Cossart, R., & Yaksi, E. (2024). Neuronal ensembles: Building blocks of neural circuits. *Neuron*, *112*(6), 875-892.

Chapter 2: Experimental Assessment of Predictive Coding

Overview

The preceding chapter developed a theoretical account of how sensory representations may be shaped by current input, earlier states at the same representational level, and contextual states at higher levels. This chapter translates that account into empirical criteria and describes the experimental and analytical framework used in this thesis. I first consider what would count as evidence for predictive inference and why contextual modulation alone is insufficient, especially when prediction is confounded with recent sensory history, temporal order, adaptation, or local stimulus statistics. I then discuss experimental designs that compare responses to the same sensory input while varying the information available to predict it, focusing on predictions derived from recent population states and from broader sequence context. Next, I specify how inferred variables are operationalized in population firing rates and consider the limitations of this measurement choice. Finally, I describe how analyses of population geometry can separate the contributions of concurrent influences on neural representations and determine how those contributions combine, before providing an overview of the two experimental studies.

What would count as evidence for predictive inference

The theoretical distinctions developed above lead to a series of progressively stronger experimental questions. A first requirement is that the representation of a sensory variable at an intermediate stage of the hierarchy should not be determined solely by the current feedforward input. If latent variables are estimated jointly, information carried by other statistically coupled variables—including previous states and variables operating over broader spatial or temporal scales—should influence the representation of the current state itself.

Context dependence alone, however, is a weak signature of inference. The stronger prediction is that these influences should reflect the statistical relationships among variables that have been learned from experience. A response may depend on recent activity for many reasons, and a contextual signal may reach the sensory cortex without constituting an estimate of the latent causes generating its input. Showing that the form of contextual modulation depends systematically on the statistical regularities relating sensory events therefore provides an additional link between the structure learned from the environment and the structure of the resulting neural representation. Even this does not establish that neural activity represents the full Bayesian posterior, its MAP estimate, or any particular approximation to it. Rather, each additional property of the representation progressively constrains the class of inferential accounts capable of explaining it.

Stronger distinctions require asking what information is represented and how the contributing constraints are combined. As described in the preceding section, different algorithms for approximate inference make different predictions about which features of population activity should carry the influence of prediction and how those influences should relate to one another. The present work addresses only a subset of these possibilities. In particular, it asks whether different sources of information produce distinguishable transformations of population firing-rate representations and whether some of those transformations combine approximately linearly under conditions in which their reliability is comparable. Other algorithmically informative properties—including changes in covariance, trial-to-trial variability, precision, and fine temporal dynamics—are left to future investigation.

A further question concerns how such information is organized across the population. Prediction-related signals might be concentrated in a specialized subpopulation, aligned with a neuron's sensory selectivity, or restricted to particular cortical layers; alternatively, they might be distributed weakly and heterogeneously across neurons that also carry sensory information. These possibilities bear on whether sensory and contextual variables are multiplexed within

shared populations or assigned to more segregated channels, and they also have consequences for how the representation should be measured: a contextual variable can be reliably represented in the collective pattern of activity even when individual neurons carry only weak signatures of it.

The two studies presented in this thesis address these questions from complementary directions. One examines contextual influences arising from the statistics of natural sensory experience and the local recurrent circuitry through which they act; the other examines contextual influences arising from regularities acquired within the experiment and the representational structure they produce. Together they ask whether intermediate sensory representations incorporate contextual information, whether that information reflects statistical regularities of the environment and of recent experience, and how different sources of information are organized within the resulting population representation. The following sections consider the experimental designs and analytical approaches through which such questions can be addressed.

Experimental designs for identifying predictive influences

Testing predictive inference requires more than demonstrating that neural responses depend on context. The theoretical framework developed above makes progressively stronger claims about the source and content of that contextual dependence, and experimental designs can provide corresponding degrees of leverage. At minimum, the current sensory evidence must be distinguished from information supplied by preceding or broader context. Stronger tests ask whether that contextual information reflects statistical regularities in sensory experience, whether it contains information about the particular sensory state implied by those regularities, and whether the mechanisms carrying that information can be identified causally. No single manipulation establishes all of these claims, but several general design principles can progressively constrain the possible explanations.

Primary visual cortex provides a particularly informative setting in which to test these claims. V1 lies near the bottom of the cortical visual hierarchy, yet its activity represents relatively low-level latent visual variables such as local orientation, edges, and other image features rather than simply reproducing retinal input. Under distributed hierarchical inference, estimates of these lower-level latent variables should themselves be conditioned by other statistically coupled variables in the generative model. The representation of an edge, for example, should depend not only on the local sensory evidence supporting that edge but also on information about broader causes and preceding states. If contextual information were instead incorporated only at later stages while V1 supplied a relatively context-independent representation for downstream interpretation, successful inference and behavior could still result, but the stronger claim that inference is distributed throughout the hierarchy would not hold at this early level. V1 therefore provides a stringent test of whether contextual inference shapes the representation of latent variables near the bottom of the sensory hierarchy. This distinction motivates the focus on V1 in both studies presented here.

Holding sensory input constant while varying context provides a direct experimental test of this possibility. If the same sensory input produces different activity depending on what preceded it, then the current response cannot be explained by the current feedforward input alone. Such dependence is consistent with the framework developed above, in which temporal and hierarchical information jointly constrain the current latent-state estimate. It is not, however, uniquely predictive. Recent sensory experience can alter subsequent responses through adaptation, through changes in recurrent network state, and through learned temporal relationships, and these are not alternatives of the same kind: adaptation is a competing explanation for a context effect, whereas local recurrence is one mechanism through which information about previous states could influence the current estimate. The relevant experimental question is therefore what information about the preceding context is preserved and how that information transforms the representation of the current input.

A second principle is to establish a relationship between contextual modulation and the statistical regularities from which predictions can be derived. Statistical structure can become incorporated into the nervous system over different timescales. Circuit organization may reflect regularities encountered over development and extended natural experience, while new relationships can also be acquired through controlled exposure within an experiment. The first study approaches the problem from the former direction by asking whether recurrent interactions in V1 reflect statistical relationships characteristic of natural visual inputs. The second approaches it from the latter by imposing a temporal regularity through repeated exposure and asking how the resulting learned structure changes sensory representations. These approaches provide complementary evidence: one asks whether cortical interactions reflect the statistical structure of the environment in which the system developed and learned, whereas the other tests whether experimentally controlled statistical experience can generate corresponding contextual influences on sensory representations.

The timescale and structure of the relevant statistical dependency provide additional leverage on where contextual information can originate. A temporal relationship recoverable directly from the immediately preceding sensory state could, in principle, be computed using information already available locally. Conversely, if determining the current prediction requires information extending beyond what the recorded population reliably retains, a context-dependent representation suggests that information from a broader temporal context has become available to that population. This consideration motivated the temporal structure of the second study: the immediately preceding image alone was insufficient to determine the next element, whereas the preceding two-image context disambiguated the prediction. Control measurements therefore asked whether V1 itself reliably retained information about the required two-back stimulus over the relevant interval; this information was not reliably decodable from population firing rates. This does not identify the source from which the broader contextual information ultimately arrives, but it constrains a simple account in which the prediction is

reconstructed solely from an explicitly retained representation of the required recent input within V1.

A still stronger test concerns prediction content rather than prediction strength. Expected and unexpected events naturally differ along dimensions other than the sensory state being predicted. Unexpected events are often rarer and more novel, can recruit orienting and arousal, and can provide learning signals indicating that an internal model should be updated. Predictability can also alter precision, changing the relative weighting assigned to sensory evidence and prior information without changing the content of either. Local inhibitory circuits, neuromodulatory systems, and thalamocortical interactions involving higher-order thalamic nuclei can contribute to these processes. Importantly, these mechanisms need not be alternatives to predictive processing in a broad sense: inference, model updating, attentional allocation, and precision regulation can occur simultaneously and may share circuitry or signals.

An enhanced or suppressed response to an unexpected event therefore establishes sensitivity to statistical context but does not uniquely reveal either what was predicted or which predictive process produced the response. This is nonetheless the manipulation on which much of the experimental literature rests, and the resulting findings are correspondingly heterogeneous, with reported effects varying in sign, magnitude, and laminar or cell-type specificity across paradigms and species. That literature is reviewed in the chapters to which it is directly relevant rather than here, since the present concern is with what such measurements can establish rather than with what they have reported.

A more informative design compares contexts that imply different predicted sensory states while allowing the current sensory input itself to remain the same. Two contexts may generate predictions of comparable strength while differing in their content. If the representation of an identical sensory event varies systematically according to the state implied by the broader context, the contextual influence contains information about what is predicted, rather than merely whether the event is expected or surprising. This is the logic of presenting the same

image under different predictive contexts in the second study. Because the compared conditions are designed to differ primarily in what is predicted rather than in how reliably it is predicted, they also provide a better approximation to holding predictive precision constant, which is important when asking how distinct contextual constraints combine. Such content-specific effects are particularly informative at the population level because prediction may change the pattern of activity across neurons without producing a corresponding uniform enhancement or suppression of the population response.

Omission paradigms provide complementary leverage by removing the sensory event whose prediction is under investigation. Activity occurring when an expected event fails to occur cannot be directly driven by that event and can therefore reveal information generated from the preceding context. An omission response alone is nevertheless ambiguous. Temporal expectation, recurrent dynamics, release from adaptation or inhibition, and other consequences of the preceding sequence can all generate activity in the absence of the omitted stimulus. Many alternative accounts can therefore explain the magnitude of an omission response without predicting that the response should contain information about the identity of the omitted stimulus. The stronger question is whether activity generated during an omission contains information about the content of the missing event and whether that information follows the statistical relationships from which the expectation was derived. Omissions are useful not because any response to an omitted event constitutes a prediction error, but because removing the current sensory evidence creates an opportunity to ask what information the contextual state itself supplies.

Causal manipulation provides a different form of experimental leverage by moving from identifying contextual information toward identifying the mechanisms through which it is generated or propagated. Observing that the same sensory input is represented differently after different histories establishes contextual dependence; perturbing a candidate circuit asks whether that circuit is necessary for producing the effect. This distinction is central to the first

study, in which patterned optogenetic stimulation was used to deliver a known input pattern to a defined group of neurons under different preceding contexts, providing leverage on the contribution of local recurrent interactions rather than merely establishing a correlation between preceding stimulation and subsequent activity. Causal and representational evidence therefore address complementary questions: one concerns the mechanism through which contextual dependence arises, while the other concerns the information carried by the resulting neural state.

No single manipulation completely resolves these ambiguities. Prediction is derived from sensory regularities, and many variables that appear as experimental confounds—history, repetition, probability, and temporal order—are simultaneously the information from which predictions are learned. Eliminating them entirely would therefore eliminate the structure from which the phenomenon of interest can arise. The goal is instead to construct complementary manipulations in which candidate explanations make different predictions and to determine which components of neural activity follow each variable. Experimental controls and causal manipulations constrain the possible sources of an effect, while representational analyses can ask what information remains and how multiple influences are organized within population activity. How these variables are separated analytically, and how their contributions can be compared within a common representational space, is considered in the following section.

Operationalizing inferred variables in population firing rates

We test in the work presented in this thesis for inferred variables in a rate-based population code. Neuronal responses are summarized as spike counts within defined time windows, and prediction-dependent effects are sought in the geometry of the resulting population responses rather than in spike timing, correlation structure, or trial-to-trial variability. Importantly, a rate-based code does not require prediction to produce a uniform increase or decrease in firing rate. Changes in the pattern of activity across neurons can alter population

geometry even when mean activity across the population changes little or not at all. Population activity also provides a natural substrate for representing the continuous quantities assumed by the inferential algorithms discussed above, and, as noted, patterns of activity across neurons can in principle parameterize probability distributions over latent variables rather than single values. The analyses used here make the weaker assumption that inferred variables are reflected in population firing-rate patterns, without requiring those patterns to encode the complete posterior distribution. This choice is partly practical and partly principled: rate-based population geometry is well matched to the algorithmic questions posed above, since the algebraic relationship by which multiple constraints combine can be expressed as a relationship among population response vectors.

This choice nevertheless carries a specific interpretive cost. Effects of prediction carried by synchrony, assembly membership, neuronal firing sequences, other fine temporal structure, or the shape of response distributions would be attenuated or invisible in these measurements. Negative results reported here therefore constrain the rate-based population code specifically and should not be read as excluding prediction-dependent signals in other features of neural activity. The choice of code determines what can be measured; the following sections consider what can be inferred from those measurements.

Disentangling concurrent influences on population representations

The experimental strategies described above provide leverage for separating prediction from correlated features of sensory experience, but they do not imply that the resulting neural responses can be partitioned into independent processes in a simple way. At any moment, activity in the sensory cortex reflects the interaction of current sensory evidence with the state produced by preceding activity and with recurrent and long-range influences structured by learned or developmentally acquired regularities. These influences can be simultaneously present in the same neurons and need not act independently. One source of information may

alter the effect of another, as in the context-dependent dynamics discussed above, or several influences may contribute jointly to the resulting population state. The analytical problem is therefore not to identify a response component that is uniquely and exclusively "prediction," but to determine whether experimentally defined sources of information produce distinguishable and theoretically meaningful transformations of the neural representation.

Recent sensory history illustrates this problem particularly clearly. A preceding stimulus can alter the response to a subsequent stimulus through reduced excitability of neurons activated by the preceding input, but history dependence is not restricted to such stimulus-specific suppression. Recurrent interactions can propagate the consequences of previous activity to neurons that were not themselves strongly driven by the preceding stimulus, thereby changing how the network responds to subsequent input. In the first study, these off-target effects are central rather than incidental: the transformation induced by recent sensory history acts as a form of filtering whose structure is related to statistical relationships present in natural visual experience. The relevant distinction is therefore not between prediction on one hand and history dependence on the other. Temporal history is itself a source of information from which predictions can be constructed, and recurrent circuitry provides one mechanism through which that information can alter the current estimate.

The same issue becomes more explicit in the structured sequences examined in the second study. Sensory identity, recent sensory history, temporal position or order, and learned predictive context are correlated within the experienced sequence and can all influence the response to a given element. The question is consequently whether the transformation associated with predictive context can be distinguished from transformations attributable to these other variables, and how the different transformations relate when they act on the same population. Importantly, separability at the level of analysis does not imply biological independence. Components that can be experimentally estimated separately may combine

additively, interact nonlinearly, or change one another's effective influence through context-dependent recurrent dynamics.

This problem motivates analyzing the geometry of population representations rather than relying primarily on changes in response magnitude. A neuronal population can simultaneously encode multiple variables through distributed patterns of activity, and the relevant information need not be concentrated in neurons selectively responsive to a single variable. Different experimental conditions can instead be treated as points in a high-dimensional representational space, with the relationships among those points characterizing which distinctions are preserved by the population code. Representational geometry therefore provides an intermediate description between individual neuronal responses and the computational variables of the theory: it characterizes both what distinctions are present in population activity and the format in which those distinctions are organized.

Representational similarity analysis (RSA) provides a particularly useful framework for this purpose. Rather than requiring a correspondence between individual neurons and particular theoretical variables, RSA characterizes a representation through the pairwise relationships among population activity patterns. This abstraction is useful here because the theoretical predictions concern transformations of inferred variables, whereas there is no reason to expect individual recorded neurons to correspond one-to-one with those variables. RSA specifically makes the representational dissimilarities among experimental conditions explicit, allowing hypotheses about which distinctions are emphasized or de-emphasized by the neural population to be compared directly. The same abstraction is also its principal limitation: because dissimilarity structure discards the directions along which conditions differ, RSA can establish that two conditions are differently represented without specifying the geometric relationship between the transformations producing that difference. Analyses that retain the directions themselves are therefore complementary to it rather than redundant with it.

This perspective also clarifies the relationship between representational and single-neuron analyses. Establishing that predictive context changes population geometry does not specify how that information is distributed across the neurons producing the geometry. The same population-level effect could arise from a small specialized group of strongly modulated neurons, from modulation aligned with each neuron's sensory preference, or from weak and heterogeneous contributions distributed across many neurons; contextual information could likewise be concentrated in particular cortical layers or distributed throughout the cortical column. The existence of a population-level representational effect does not by itself distinguish among these possibilities, and a transformation that is difficult to identify from the response of any single neuron may nonetheless be apparent in the collective geometry of population activity. The analyses in this thesis therefore move between these levels: population analyses ask what information is present and how experimentally defined components relate within the representational space, whereas analyses of individual neurons, sensory alignment, and cortical layers ask how that information is distributed across the biological population.

Several analytical frameworks could, in principle, be used to disentangle such concurrent influences. Linear dimensionality-reduction approaches such as demixed principal component analysis can seek axes associated with experimentally defined variables, while nonlinear representation-learning approaches such as CEBRA can identify low-dimensional structure related to behavioral or experimental variables. Latent dynamical models address a somewhat different question by attempting to separate observed population activity into underlying dynamical structure and external inputs. During this work, LFADS was used exploratorily to ask whether stimulus-driven input could be distinguished from dynamics generated within the recurrent network, and initial analyses indicated that the distinguishing information was carried by the inferred inputs rather than by the recurrent dynamics. These analyses were not pursued to a conclusion. Establishing what such a model's inferred inputs and dynamics correspond to would require substantial validation in its own right, and the experimental design already

provided a more direct source of leverage on the same question through the timescale of the imposed statistical dependency.

The primary analyses ultimately remained closer to the measured representational geometry because this was sufficient for the questions posed here and required fewer additional assumptions about the latent structure generating the data. RSA does not require specifying a particular low-dimensional coordinate system or assigning individual dimensions to computational variables; instead, the relationships among conditions can be compared directly with experimentally or theoretically defined models. This is also why RSA, encoding models, and related representational approaches are better regarded as complementary tools rather than mutually exclusive descriptions: they expose different aspects of the same underlying population representation.

Within this framework, disentanglement does not mean that prediction, recent history, temporal order, and sensory evidence must occupy orthogonal neural dimensions. Rather, the goal is to determine whether their contributions can be distinguished, whether they exhibit systematic geometric relationships, and whether the representation observed when multiple constraints are present is consistent with particular rules for their combination. The approximately additive relationship examined in the second study is one such test. The analytical strategy adopted here therefore does not attempt to recover every latent component of the underlying computation. It asks a more constrained question: given experimentally controlled sources of information, what transformations do they produce in the measured population representation, and what does the relationship among those transformations imply about the computation that generated them?

Overview of the studies

The studies presented in this thesis examine how predictions derived from statistical regularities at two temporal scales shape representations of current sensory input in mouse

primary visual cortex. The first concerns information carried forward by earlier population states at the same representational level. The second concerns information supplied by broader sequence context, which hierarchical accounts associate with states operating over longer temporal scales. Together, the studies ask how these two predictive sources alter the representation of current input and how their contributions combine with sensory evidence.

The first study, described in the following chapter (Chapter 3), examines how earlier activity at the same representational level shapes responses to subsequent input. Patterned optogenetic stimulation was used to impose controlled sequences of population activity in V1 while holding the current input pattern fixed. The study asks whether responses to the same input depend on the preceding patterns, whether sequences matching temporal transitions characteristic of natural vision are selectively enhanced, and whether this sequence dependence extends beyond directly stimulated neurons through excitation and suppression within the surrounding population. By examining how recurrent dynamics carry information from one population state into the next, the study relates same-level temporal prediction to filtering in predictive coding and Kalman filter models and asks whether recurrent interactions embody a prior over how represented variables evolve through time. This work was published as Deveau et al. (2026), to which I contributed the electrophysiological recordings characterizing the duration of V1 responses to natural images.

The second study, described in Chapter 4, examines how predictions derived from broader temporal context shape V1 representations. In hierarchical predictive coding accounts, such predictions may be supplied by higher levels that integrate information over longer temporal spans. Mice were exposed to structured visual sequences in which the identity of the next element depended on more than the immediately preceding stimulus, while I recorded population activity across all layers of V1 using Neuropixels probes. This design allowed identical sensory events to be compared across contexts implying different predicted states, including comparisons in which the current stimulus and immediate history were held constant.

The study asks whether population representations differ according to prediction content, whether this contribution can be distinguished from those of the current stimulus, immediate history, and temporal order, how these contributions combine within representational space, and how contextual information is distributed across neurons and cortical layers.

Together, the studies connect recurrent filtering over shorter timescales with prediction derived from broader sequence context. They provide complementary circuit and representational perspectives on how temporal statistics become incorporated into V1 activity and how current sensory evidence is combined with information carried by recent population states and broader contextual structure.

Chapter 3[1] :Recurrent cortical networks encode natural sensory statistics via sequence filtering

Ciana E. Deveau[1,2,3,6], Zhishang Zhou*[1,6], Paul K. LaFosse[1,2,4], Yanting Deng[1], Saghar Mirbagheri[5], Nicholas Steinmetz[5], Mark H. Histed[1,7*]

[1] Intramural Program, National Institute of Mental Health, National Institutes of Health; Bethesda, MD USA.

[2] NIH Graduate Partnership Program, Bethesda, MD USA,

[3] Department of Neuroscience, Brown University, Providence RI USA,

[4] Neuroscience and Cognitive Science Program, University of Maryland, College Park MD USA [5] Department of Biological Structure, University of Washington, Seattle, WA USA

[6] These authors contributed equally

[7] Lead contact * Correspondence: mark.histed@nih.gov

Summary

Recurrent neural networks can generate dynamics, but in the sensory cortex it has been unclear if any dynamic processing is supported by the dense recurrent excitatory-excitatory network. Here we show a role for recurrent connections in mouse visual cortex: they support powerful dynamical computations, but by filtering sequences of input instead of generating sequences. Using two-photon optogenetics, we measure neural responses to natural images and play them back, finding responses are boosted when inputs are played back during the correct movie dynamic context— when the preceding sequence corresponds to natural vision. This sequence selectivity depends on a network mechanism: earlier input patterns produce responses in other local neurons, which interact with later input patterns. We confirm this mechanism by designing sequences of inputs that are boosted or attenuated by the network.

[1] This chapter is adapted from the following published work: Deveau, C. E., Zhou, Z., LaFosse, P. K., Deng, Y., Mirbagheri, S., Steinmetz, N., & Histed, M. H. (2026). Recurrent cortical networks encode natural sensory statistics via sequence filtering. Neuron, 114(8), 1489-1503. My contribution to this work was performing the electrophysiological recordings used to characterize the duration of responses to natural images in mouse V1.

These data suggest recurrent cortical connections perform predictive processing, encoding the statistics of the natural world in input-output transformations.

Introduction

A defining feature of the cerebral cortex is its extensive local recurrent connectivity. Excitatory cortical cells make up approximately 80% of the neurons in all cortical areas, and each excitatory cell receives hundreds or thousands of inputs.[1] A majority of those input synapses come from other local neurons, within a distance of a few hundred microns.[2–4]

However, in many cortical areas, the function of this extensive local recurrent network is not well understood. For example, the sensory functions of primary visual cortex (V1) have often been explained by feedforward or feedback inputs alone, with little need for recurrent computations. Foundational work on V1 from Hubel and Wiesel[5,6] showed that feedforward summation of visual thalamic inputs can create the edge detection properties (i.e. orientation and direction tuning) that are a defining characteristic of early visual cortex. Later work using inactivation and intracellular recording,[7–9] confirmed this, finding that even though recurrent inputs to V1 neurons were large, recurrent input did not substantially refine visual tuning, for example not sharpening orientation responses.[10]

However, other work did find some contribution to visual function made by recurrent connections. Connectivity studies that combined *in vitro* and *in vivo* measurements showed neurons with similar orientation tuning are preferentially recurrently connected.[11–13] This observation suggested that the recurrent network could perform preferential amplification of feedforward inputs driving orientation tuning[3,14]. That prediction was borne out using two-photon holographic stimulation, which confirmed that cells with similar orientation sensitivity produced amplified responses when stimulated simultaneously.[15,16]

In sum, many years of study of V1 responses has found a primary role of the area is to extract edges from the visual field,[17,18] and also found that the recurrent network of V1 facilitates

that function. Population-level recording studies confirm this idea, showing that much variation in V1 neurons' responses is well-described by receptive field estimates derived from simple edges.[19,20]

But this framework is not the whole story. When V1 responses are recorded during natural vision, e.g. in response to natural movies, there is significant variation in neurons' responses that are not well-explained by mere edge detection. That is, V1 response estimates derived from the tuning of the classical receptive field "center" do not come close to predicting all the variance of V1 neurons during natural vision. As Olshausen et al. put it,[21] "What is the other 85% of V1 doing [during natural vision]?". Up to 85% of the variability of V1 responses has not been explained by orientation or direction-based receptive field models, both when that question was posed and at the present time.

One possibility is that much of the responses of V1 neurons are explained by non-visual factors. Over the past 15 years, a large number of studies have focused on non-visual responses in V1 — responses influenced by the self-motion of the organism (e.g. running,[22–24] top-down inputs,[25,26] prediction,[27] or cross-modal responses to non-visual sensory input, such as tactile sensation or auditory stimuli).[28–30,but see 31]

On the other hand, it may be that the remaining unexplained V1 function is due to visual properties that have not yet been uncovered. There are hints that unknown visual responses drive this missing variance. The first hint of this sort is that V1 responses to natural movies are stronger and sparser — statistically quite different — than predicted from models that use V1 receptive field (edge) properties alone,[32–34] implying some lawful rules of visual processing that have not yet been identified. Also suggesting there may be some V1 role in complex natural vision, neurons co-activated in natural vision are also more likely to be connected.[35] And using large-scale recording methods, the responses of V1 neural populations have been found to be high-dimensional, as if they describe a very large space of visual responses extending beyond classical oriented-edge receptive fields.[36] In fact, studies of V1 receptive fields generally

observe a 'non-classical surround' — a large region outside the classical receptive field that affects V1 neurons' responses, with hints of complexity[37] that have not been so far fully deciphered (reviewed in [38]).

One thing that could explain the missing variance in visual cortical responses is information about time. During natural vision, the visual input the organism receives changes as the visual world changes, both due to motion of objects and due to the organism itself moving. That means that the natural dynamic stimuli which produce strong and sparse V1 responses contain complex motion of various sorts.

But a major challenge to this idea, that the missing variance in V1 responses is explained by temporal information, is that many past studies have found only simple modulation of V1 neurons' activity with time. Responses of V1 neurons to a sequence or movie of full-field stimuli are predicted by linear combinations of responses to single component stimuli,[19,39,40] and V1 neurons often do not have complex spatiotemporal receptive fields (STRFs) as have been seen in audition.[41–43] That is, receptive field models that explain V1 responses as well as currently possible are often — but not always[34,44,45] — spatiotemporally separable, so that the temporal kernel and spatial kernel of a V1 receptive field can be simply multiplied together to give its spatiotemporal response.[44,34,46]

We thus are left with a question about how the V1 recurrent network transforms spatiotemporally dynamic visual cortical responses. Much of what we can explain of V1 function can be explained by feedforward input and edge processing, but this does not explain V1 population responses to natural vision. What explains this missing variance in V1, and is that unknown non-classical function created by processing in the recurrent network?

Here we provide an answer to that question, showing that V1 responses are indeed modulated by temporal context, and that the recurrent network encodes features of natural vision, linking responses of neurons across time. We use two-photon holographic stimulation to simulate patterns of input derived from natural movies. We find the gain, or magnitude, of the

response to a given feedforward input is affected by other neurons' activity in the recurrent network, showing that V1 responses to a visual stimulus vary in a predictable way based on temporal context. When a given patterned input is provided in the correct temporal context, preceded by the visual input that occurs in natural vision, the response to that patterned input is larger than when the same input pattern is delivered in a non-natural context. If one frame of a natural movie is preceded by the 'correct' prior input, as expected from natural vision, the response to that frame is boosted. Our data may explain why past studies have found a large non-classical surround effect, but have not been previously able to find lawful rules of how simple stimuli predict non-classical surround responses: the non-classical surround is too complex to fully explain by exploring the space of visual stimuli using recording methods. Here, by activating parts of known, identified natural stimuli using patterned stimulation, we show stimuli matching natural statistics specifically boost V1 responses, as populations of neurons in V1 responding to stimulus subparts influence each other through recurrent mechanisms.

Our results thus support the idea that the V1 recurrent network is performing complex computations over time not by generating sequential responses, but by filtering sequences of input. Natural vision produces a changing set of patterned input to the cortex, and either via experience or development, the cortical recurrent network has learned, and boosts, input sequences that correspond to natural vision. We confirm this in an artificial recurrent network, showing that a network selecting natural sequences of input is consistent with our data. We train a network to boost certain sequences and show that the network reproduces our experimental results. Then, returning to experiment, we use two-photon stimulation to identify a recurrent mechanism for this sequence filtering, construct novel sequences of input patterns following this mechanism, and show that sequences designed to be boosted do indeed produce larger responses. This confirms a recurrent origin for the sequence filtering that can underlie natural visual processing in V1.

Taken together, our results show that V1 neural responses are indeed sensitive to temporal context, but in a way that is not easily described by single cell response properties or receptive fields, but instead by populations of neurons affecting other populations via recurrent influences. Because we find that neurons' gain is affected by the activity of other local neurons, the information for this boosting, or sequence amplification, process can be encoded in the large number of excitatory-excitatory synaptic weights in the V1 local network. These results thus provide one possible solution for what "the other 85%" of V1 is doing: beyond just edge extraction, the V1 recurrent network is sensitive to the complex temporal structure of the natural world.

Results

Two-photon holographic stimulation (Fig. 1E)[47,48] allows us to deliver direct inputs to populations of cells in the cortex and directly study the effects of local neural activation on other local neurons. We use a viral expression strategy[49] that co-expresses an opsin and a calcium indicator in excitatory neurons in layer 2/3 of mouse primary visual cortex, stably over days to weeks, with minimal optical crosstalk between imaging laser and opsin (Fig. 1A-C; as we previously characterized[49]). This approach allows us to directly access the dense local recurrent circuitry between excitatory cells (Fig. 1D). With two lasers we simultaneously stimulate groups of neurons with single-cell resolution and measure responses to the optogenetic stimulation and/or visual stimuli (Fig. 1E, details in Methods). The optogenetic inputs are precise spatially (Fig. 1F) and temporally, with one input pattern changing to the next within tens of milliseconds (Fig. 1I,J; Supp. Fig. 1B,C).

Two-photon optogenetics can mirror features of natural visual input by driving population activity in the cortical network. When animals are shown natural visual movies, many neurons in the visual cortex respond (Fig. 1G). Using two-photon stimulation we can simulate these activity patterns by giving input to many neurons at once, activating the network through excitatory cell

stimulation (Fig. 1H; in this work we target 15-30 cells,[50]). This all-optical approach allows us to measure sensory responses, select neurons that respond, and activate the excitatory patterns.

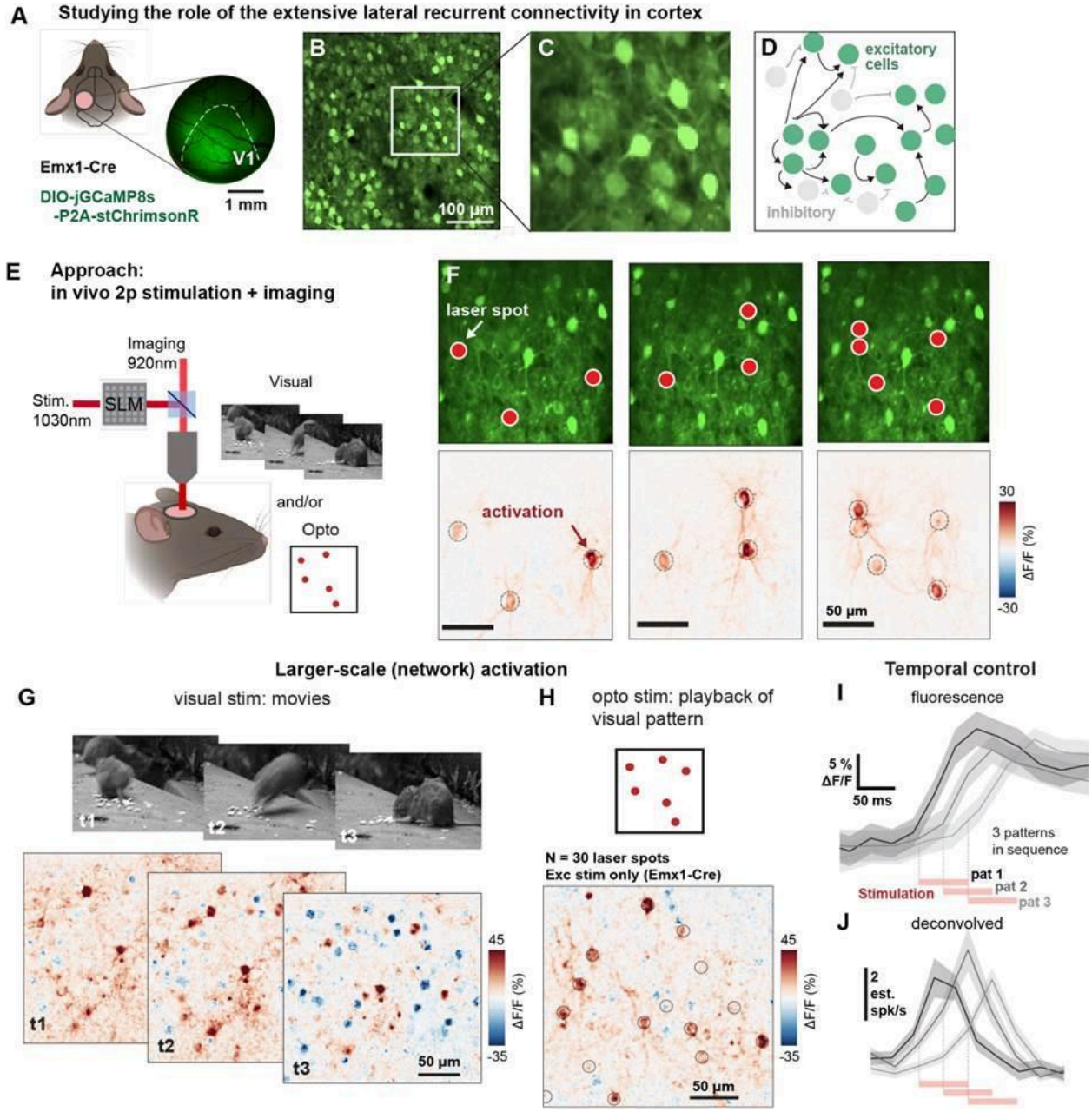


Figure 1: Using two-photon holographic stimulation and imaging to probe the role of excitatory recurrent connections.

(A) 3mm window surgical implant over V1. **(B)** FOV image of excitatory cells expressing both GCaMP8s and opsin stChrimsonR (via AAV-DIO-jGCaMP8s-P2A-stChrimsonR in Emx1-Cre line). **(C)** Enlarged region of (b). **(D)** Schematic of the ubiquitous recurrent connections between excitatory cortical neurons. **(E)** Schematic of two-photon imaging and stimulation experiments, allowing pairing of visual and patterned optogenetic stimuli. SLM: Spatial Light Modulator; splits the stimulation laser beam to target selective patterns of neurons. **(F)** Example of spatial precision of 2p stimulation; left to right: three different patterns, with 3, 3, and 5 laser spots, each spot targeted to a neuron and

stimulated simultaneously. The patterns are non-overlapping in time. **(G)** Network activation from natural movie input at three timepoints. **(H)** Network activation, with non-target changes in non-stimulated neurons, also results from optogenetic patterns (black circles: laser spots, 30 total, a subset of image shown here for visual clarity). **(I-J)** Example of temporal precision of sequential 2p stimulation: 3 patterns, 15 cells each, here 60 ms stim per pattern with 30 ms between pattern onset; for other experiments overlap is 0-30 ms and pattern duration is 30-121 ms (Methods). **(I)** Average fluorescence traces across cells and trials, reflecting calcium responses of stimulated cells; **(J)** Corresponding deconvolved traces using OASIS (Methods). These data show that even with some overlap, imaging can resolve distinct temporal peaks with stimulation onsets separated by 30 ms.

Sequential order affects V1 neurons' responses through non-target neural suppression

If recurrent effects influence V1 responses to temporally-varying input, then changing the order of patterns of input might be expected to produce differences in neurons' responses. To determine if this is true, we chose patterns of neurons at random (three patterns; A, B, and C, each a fixed duration, 30 or 60 ms in different experiments, spanning a range of the temporal frequencies often seen in vision: e.g. frame times used in videos,[51] flicker fusion frequencies,[52,53] and photoreceptor dynamics.[54] We presented these patterns in two different orders (ABC or CBA, Fig. 2A).

We find the sequential order of these population input patterns does indeed affect neurons' responses (Fig. 2B-E). The first pattern produces a greater response than the same pattern when it is delivered later in the sequence (Fig. 2B-E). This sequential context dependence occurs in the cells stimulated as part of the A, B, or C patterns. On average, the cells stimulated first in the sequence show stronger responses (Fig. 2C-E; the median of the A and C distributions differ in Fig. 2E). Large effects are seen in individual neurons, which are strongly modulated by sequential position (e.g. there is significant variance in the distributions in Fig. 2E: 37% of A neurons and 45% of C neurons are significantly different from zero, K-S test $p < 0.05$).

To understand whether prior patterns of input to other cells could change the response of neurons to a particular input, we analyzed responses in the B cells (Fig. 2F). The B input pattern is preceded by one pattern in each sequential order: the A pattern in the ABC sequence, and the C pattern in the reverse sequence. Thus, if there is a specific effect of sequential context on individual neurons, the firing of some B cells to the B input should be modulated also by the

pattern that precedes it. Indeed, we found that many of these cells were significantly modulated by sequential context (21% of B cells $p<0.05$, Mann-Whitney U test, Fig. 2G-L). Even though the B input was always the same, some B cell responses were larger in the ABC sequence (Fig. 2G-I) and some were larger in the CBA sequence, as the pattern that preceded the B pattern was changed (Fig. 2J-L; variance differences via A-D test, Fig. 2N-O). This variation in response was not dependent on the distance between each B cell and the previously stimulated A or C cell (Supp. Fig. 2). Further, stimulation of each pattern individually yielded a similar distribution of excitation and suppression in the non-stimulated neurons (Fig. 2P,Q). Thus these randomly selected patterns produce similar non-target activity that suppresses incoming inputs on average. In sum, stimulation with sequences of random population input patterns shows that V1 neurons' responses to fixed inputs are highly dependent on sequential context. Responses to an input pattern are affected by prior patterns of input to different neurons. This suggests that the visual cortex recurrent network may perform a time-based computation for sensory stimuli, and to understand this, we next investigated how sequential modulation relates to natural vision.

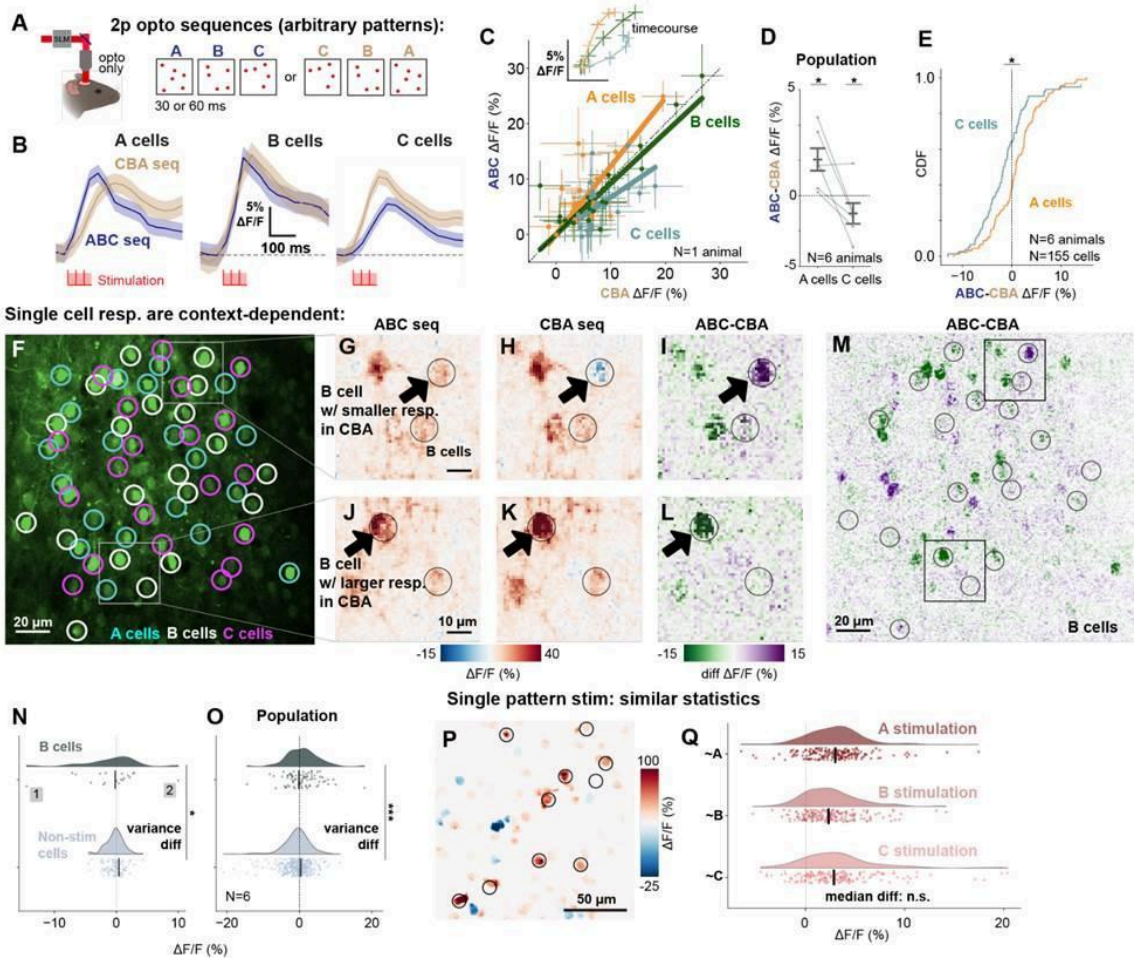

A
2p opto sequences (arbitrary patterns):
opto only
A B C
or
C B A
30 or 60 ms
B
A cells
B cells
C cells
CBA seq
ABC seq
5% ΔF/F
100 ms
Stimulation
C
timecourse
5% ΔF/F
A cells
B cells
C cells
N=1 animal
ABC ΔF/F (%)
CBA ΔF/F (%)
D
Population
ABC-CBA ΔF/F (%)
N=6 animals
A cells C cells
E
CDF
C cells
A cells
N=6 animals
N=155 cells
ABC-CBA ΔF/F (%)
Single cell resp. are context-dependent:
F
20 μm
A cells B cells C cells
ABC seq
CBA seq
ABC-CBA
G
B cell w/ smaller resp. in CBA
B cells
H
I
J
B cell w/ larger resp. in CBA
10 μm
K
L
-15
40
ΔF/F (%)
-15
15
diff ΔF/F (%)
M
ABC-CBA
20 μm
B cells
Single pattern stim: similar statistics
N
B cells
1
2
Non-stim cells
variance diff
O
Population
variance diff
N=6
ΔF/F (%)
P
100
ΔF/F (%)
-25
50 μm
Q
A stimulation
B stimulation
C stimulation
~A
~B
~C
median diff: n.s.
ΔF/F (%)

Figure 2: Patterns of input produce different responses in the V1 network depending on the sequential context.

(A) Experiment schematic: stimulation of three distinct patterns (e.g. N=30 cells per pattern, cells chosen at random for each pattern; each pattern 30 ms duration). Two sequences: forward (ABC) and reverse order (CBA). **(B)** Responses depend on sequential order. Mean responses averaged across cells of A (left), B (middle), and C (right) cells in the ABC (blue) or CBA (beige) trials (N=1 expt; example). **(C)** Points: average response (across trials) of individual neurons, same data as in B. Change in slope for A, B, C indicates differences in response based on sequential order. Inset: Timecourse of average response across cells for each pattern (begins 5 frames prior to stimulation; each point one imaging frame). **(D)** Population data (N=6 expts). gray lines: ABC - CBA responses of A (left) and C (right) cells; averaged over A, C cells per experiment. Error bars: SEM. (t-test, A cells (N=79): $p<0.0001$, C cells (N=76): $p=0.001$). **(E)** ABC and CBA responses differ for A cells (N=79) and C cells (N=76) (K-S test, $p<0.001$). **(F-M)** B cell responses reveal that effects of sequential context vary from cell to cell. **(F)** Anatomical image (green: bicistronic GCaMP8s and stChrimsonR expression). **(G-L)** Neurons' (B cells, black circles) responses to stimulation change based on which other neurons (A, C) are stimulated before them. B cells. **(G-I)**: example B cell with stronger activation in the ABC sequence compared to CBA. **(J-L)**: example B cell with stronger CBA activation. **(M)** Larger view of cells shown in g-i. Image matches anatomical FOV in f. Black squares: regions shown in G-I. **(N)** Some stimulated cells' firing is increased and some decreased by previous stimulation pattern. x-axis: B cell response difference between ABC and CBA pattern; if responses were not modulated by sequence, variability would be similar to non-stimulated neurons (bottom: gray, cells not stimulated in any A, B, or C pattern); instead variability is larger (A-D test for variance diff, N=19 B cells, N=95 non-stim cells, expt shown in F-M, two-tailed $p<0.01$). **(O)** Same as N, across N=6 expts (A-D test, N=81 B cells, N=339 non-stim cells, $p<0.001$). **(P-Q)** Sequence modulation is not due to response differences to single pattern stim: A,B,C patterns have similar statistics (data from B-E; Mood's median test, N=3 stim, $p=0.15$).

Patterns corresponding to natural visual inputs produce larger responses

To study sequential modulation in the visual context, we turned to dynamic natural visual inputs. One of the most common kinds of sequential input that the visual cortex receives comes in response to natural motion. Because of the size and scatter of cortical receptive fields,[55,56] and because axons carrying inputs contact many cells, the cortical network receives a changing set of patterned inputs during dynamic vision. As an example, when we see a person or animal running, the cortex receives a sequence of patterns of input.

In principle, the cortical network could store, in its recurrent connections, information about these patterns and thus the temporal structure of the natural visual world. This could allow the network to preferentially respond to sequences of input corresponding to natural vision. In that case, naturally-occurring sequences would be expected to produce different responses in the network than sequences that do not correspond to natural vision. This could occur by earlier parts of a dynamic natural stimulus influencing responses to input arriving at later times, as we observed with random input.

We examined this by using two-photon stimulation to mimic responses from one part of a natural movie while we changed the prior visual context. We first measured responses to a

single frame taken from a movie (Fig. 3A, Supp. Fig. 3D). We constructed a two-photon input pattern from the cells activated by that single frame. We then showed animals the movie and played back the response to the frame with two-photon stimulation, replacing the visual frame with the optogenetic input pattern either at the correct time in the same movie (Fig. 3A,B: matched context), or in a different movie such that the preceding movie frames were not matched to the stimulation pattern (Fig. 3A,E: unmatched context). If V1 responses were dependent on the dynamic context of natural vision, neurons should produce different responses to the same, fixed, input pattern when it is presented in different contexts. We subtract the respective visual response from the visual plus optogenetic response to give the two-photon incremental response in each context (Fig. 3B-H). This reveals a context-dependent effect, where the input pattern produced a larger response at the appropriate time in the visual movie (matched context) than when presented in a different visual context (the unmatched context; Fig. 3D,G-K).

The average response in the matched case is similar to the response in the opto-only case (i.e. blue and red lines are similar in Fig. 3I-K; see also Supp Fig. 3A), not larger than the opto-only case, as a first impression of amplification might suggest. However, it could be that some patterns (in the matched case) are indeed selectively amplified, but the overall network response is smaller during complex natural vision than when animals view simple unchanging stimuli (gray screen: opto only). That is, the cortical state could change during natural vision, scaling both matched and unmatched responses down, so that the matched response is similar to the opto-only response. Thus, the key observation here is not similarity to the opto-only response, but the differential response in the matched and non-matched contexts.

An immediate question is about mechanism: how might responses to one pattern of input influence responses to later input? For a recurrent network interaction to modify responses to later inputs, it would seem that responses from one frame should be sustained for some time. Then, one input pattern could produce non-target responses in other local neurons, and a later

input could fall on these non-target neurons while their responses were changed by the previous input, allowing the two inputs to interact with each other. To determine if this kind of sustained response was present, we turned to Neuropixels electrophysiological recordings, due to their fidelity in measuring the times of spikes. We recorded V1 neuron responses to flashed single image frames (as in Fig. 3A) and measured the duration of their non-target responses. We sorted neurons into groups based on whether they showed activated or suppressed responses, and found that both activated and suppressed responses were sustained after visual stimulus offset, with suppressed responses lasting for several hundred milliseconds (Fig. 3L). This is sufficient for one frame of a natural visual movie to interact with the next. For comparison, a movie at 25 frames per second, a rate comfortably within the range where humans see smooth motion, has a 40 ms frame time. The mouse visual system likely supports even faster processing than the human, and yet here we found that natural images flashed for short times produced responses 70 ms long for activated neurons and 210 ms for suppressed cells (Fig. 3L). These data suggest that the dynamics of input during natural vision is sufficient to allow overlap between patterns, allowing recurrent influences from earlier input patterns to affect responses to later inputs.

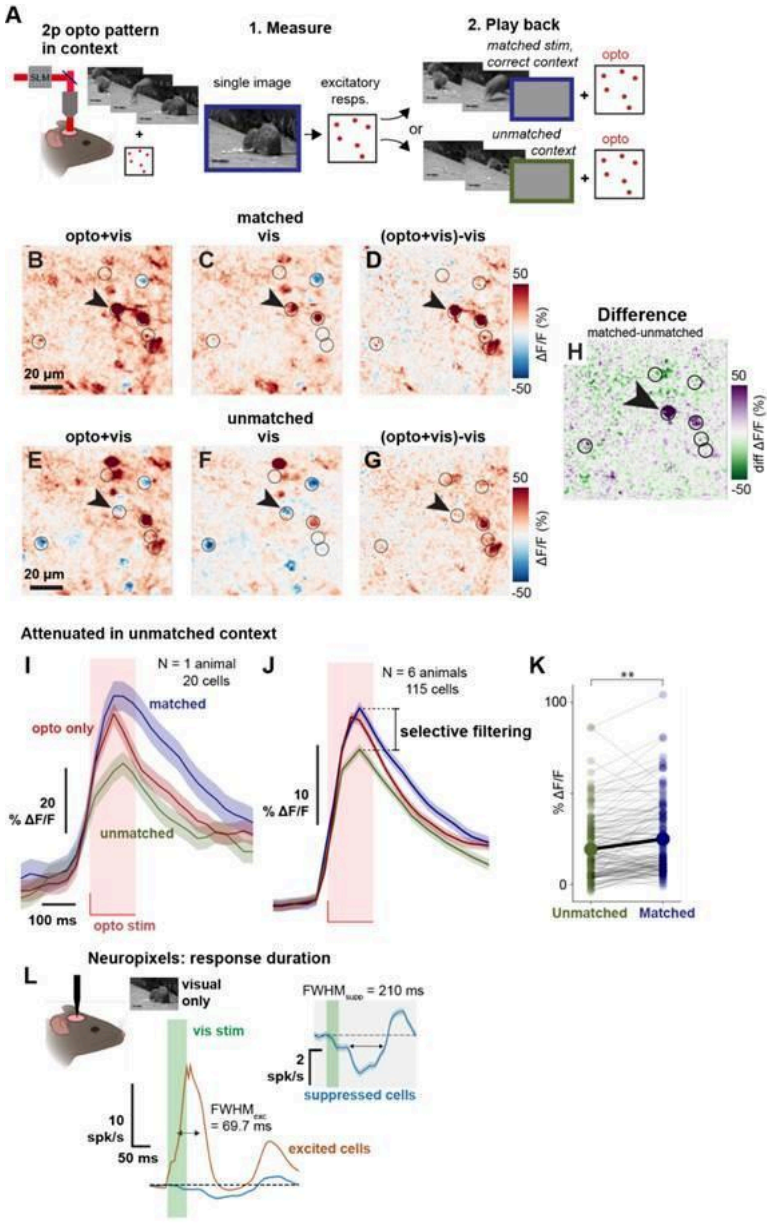


Figure 3: Sequence amplification for visual input, observed by simulating responses to a single frame of a movie.

(A) Schematic of dynamic context playback experiment: We recorded responses to a frame from a natural movie, and played back that pattern using stimulation during a visual movie. We did this at the correct time in the same movie (matched context) or in a different movie (unmatched). Three different temporal locations were tested in two movies (Methods). **(B-G)** Example cell responses to combined opto and visual stimulation, visual stim alone, and isolated opto for matched (B-D) and unmatched (E-G) contexts. (D,G) Responses in matched and unmatched contexts, computed by subtracting vis response from opto+vis combined response. Black circles: optogenetic stimulation targets. **(H)** Example larger response in matched context (arrow), difference of responses shown in (D) and (G). **(I)** Mean responses averaged across cells in matched (blue), unmatched (green), and opto only (red) trials (N=1 expt, pattern duration: 120ms). **(J)** Population time courses, N=6 animals. Responses are larger in matched context. **(K)** Points are cell responses in matched (blue) and unmatched (green) contexts from data in J. Mean optogenetic responses are significantly larger in the matched context (Mann-Whitney U test, N=115 cells, $p<0.01$). **(L-M)** Electrophysiological responses to a flashed natural image show responses sustained for up to hundreds of

milliseconds (N=5 animals, N=11 expts). **(L)** Suppressed cells (N=3,569 cells, blue; errorbar SEM). **(M)** Excited cells (N=14,528 cells, orange; errorbar, SEM, hidden by mean). FWHM: Full width at half max.

A recurrent neural network model that boosts specific sequences produces the context-dependent responses we see in our data

The observation that responses to single input patterns are modulated by context — by earlier patterns or frames (Fig. 3A-K) — implies that longer sequences of input are also selectively filtered, as each movie frame response interacts with later movie frames. That is, the pairwise sequential interactions we found, where single frames of input are treated differently based on prior input, can also enable processing of longer sequences of dynamic input where longer natural sequences of input are selectively filtered. Biological motion, like observing a predator moving to attack, can be ongoing for some time and generate changing sequences of input patterns to the cortex through that time.

To determine whether boosting of even longer sequences can result from the temporal interactions we studied, we examined a recurrent neural network (RNN) trained to preferentially generate stronger responses for some seconds-long input sequences and not others (Fig. 4A-C). The target output was a stronger, sparser scaled version of the input (in the target, for 35% of units output was 2x input; for the remaining, output was 0.9x the input), to mimic natural movie response characteristics (Fig. 4A). The model preferentially boosted the trained sequences and attenuated scrambled versions of those sequences (Fig. 4B). We asked if the context-dependent effects seen in our Fig. 3 experiment could explain this selective modulation. To test this, we chose a single input pattern as in the experiment, reflecting responses to a single frame of a sequence. We measured the response amplitude of that pattern when in the matched context versus the unmatched context. The two states are driven by presenting either the first 4 patterns of the trained sequence (matched) or of the untrained sequence (unmatched). These prior inputs in the model may be seen as simulating the L4 or LGN inputs to L2/3 visual cortex with the final input pattern simulating the optogenetic stimulus. Indeed, with an RNN trained to boost a sequence of inputs, a single pattern extracted from the sequence and

played back in the correct context produced a larger response, compared to when it was presented in the incorrect context (Fig. 4C). In contrast, a control network without amplification does not exhibit this context dependent effect. The input pattern produces the same response magnitude in both matched and unmatched contexts (Fig. 4C inset, Supp. Fig. 4).

Thus, an RNN trained to boost a sequence of input mirrors the context-dependent stimulation effect seen in our experiments (Fig. 3J, Fig. 4C). Together with the finding that neurons' responses to input are dependent on earlier inputs (Fig. 2), these observations support the idea that the visual cortex recurrent network is filtering extended sequences of input, specifically boosting input sequences corresponding to natural vision.

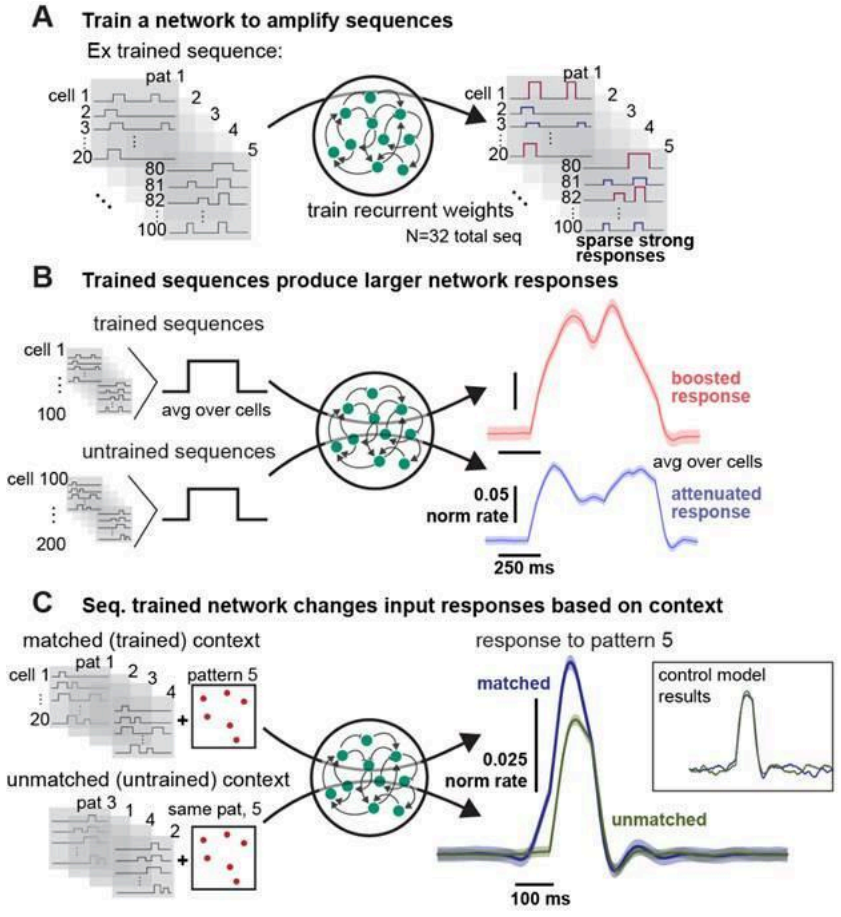


Figure 4: RNN trained to amplify patterns produces context dependent modulation.

(A-C) RNN model trained to amplify specific sequences shows responses that match our data. Pattern duration: 120 ms. **(A)** Network is trained to amplify sequences of input, resulting in **(B)** Boosted output timecourses for trained input sequences. **(C)** Simulating experiment of (Fig. 3A-K) recapitulates the data: single patterns taken from a natural

sequence produce larger responses in the correct (blue) context. Inset: control model does not show context-dependent responses. Error bars: SEM.

Recurrent network mechanism: one input pattern creates responses in other, non-targeted local neurons, which interact with later input patterns

If sequence modulation was created by local, recurrent interactions, we should be able to see signs of these interactions in the network activity. To look for specific influences — to see how particular input patterns filter later patterns — we separated stimulation patterns in time, inserting delays of several seconds (8 seconds, Methods) between one or more patterns in a sequence (Fig. 5A). We first inserted a delay between pattern B and C in an ABC sequence (A,B,C patterns randomly chosen, Methods).

The reduced sequence produces striking non-target responses (Fig. 5A-C) in cells not receiving stimulation, with some non-stimulated cells activated, and others suppressed. These non-target neural responses are likely to be primarily due to local recurrent interactions, not due to axonal or dendrite activation, as here we used a somatically-targeted opsin (stChrimsonR).[49]

These non-target responses induced by the AB sequence can occur in neurons that will later be stimulated in the C pattern. This could be a mechanism for sequential modulation: that is, a single input pattern produces non-target responses in other local neurons, and these responses interact with later patterns of input to modify responses to those later patterns.

To see how earlier patterns impacted later patterns, we examined how C cells changed their responses when preceded by the AB pattern, and whether this change was predictable from the AB pattern responses. In prior work we have found attenuation-by-suppression,[57] where V1 neurons, when suppressed, produce smaller (sublinear) responses to input due to the shape of neurons' activation functions. Such a nonlinearity could be a mechanism that creates preferential suppression of some patterns and sequences. To understand if sequence selectivity could depend on such nonlinearities, we examined non-target responses in C cells generated by the AB sequence. We compared the ABC response to a prediction computed by summing

the C response and the AB response. We found that excited non-target cells did produce linear responses when they were stimulated next in sequence, with ABC responses well-predicted by the sum of non-target responses to AB and C (Fig. 5D,F, Supp. Fig. 5A). And consistent with prior findings on activation-by-suppression, suppressed non-target cells showed nonlinear responses (Fig. 5E,F, Supp. Fig. 5A).

These responses suggest that the sequential modulation we observed does involve some network mechanisms. Non-target responses, induced by earlier patterns of input through the local recurrent network, generate a non-target response pattern that interacts with later patterns.

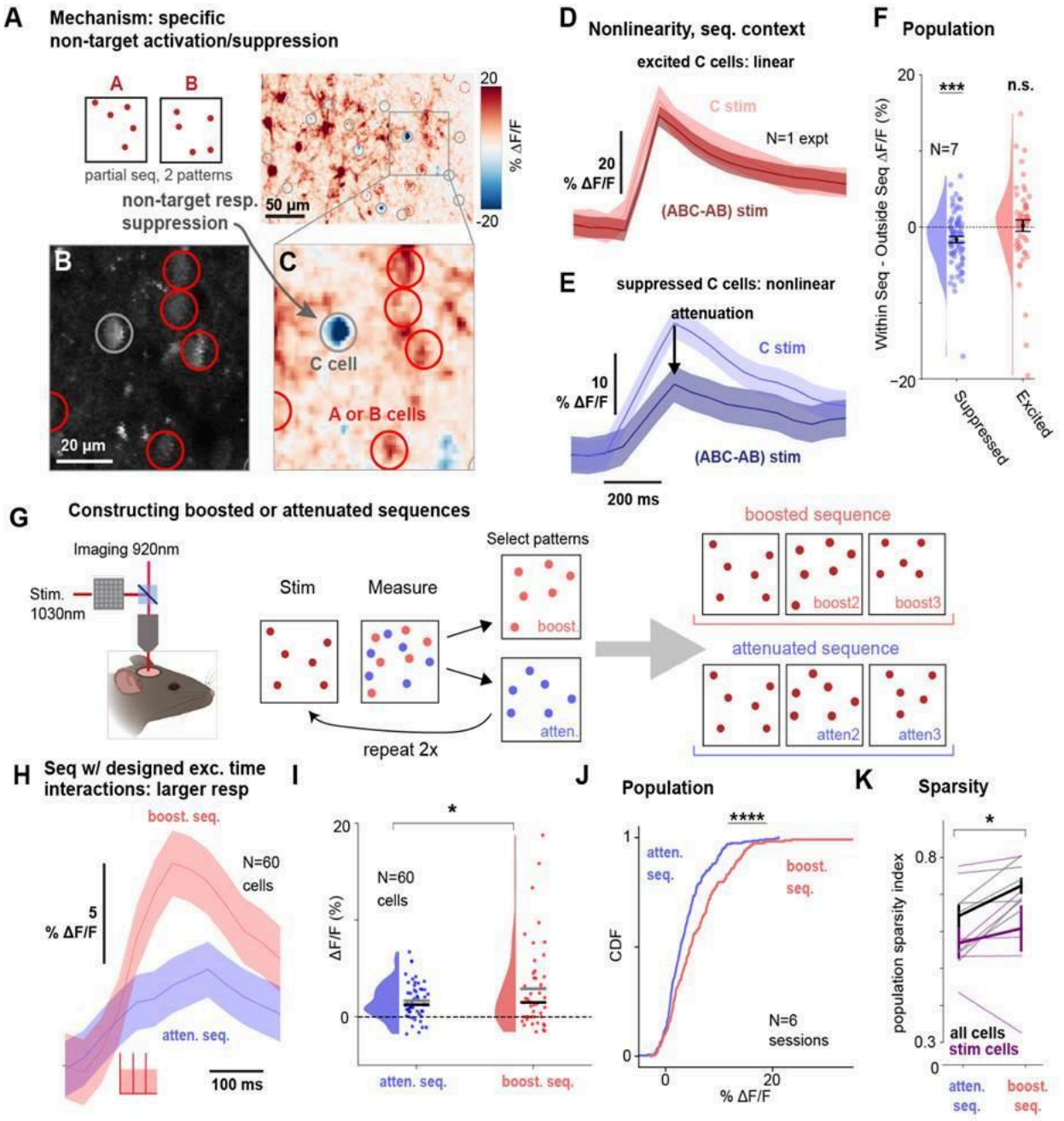


Figure 5: Sequences designed to be boosted produce strong and sparse responses, validating a recurrent network mechanism.

(A-C) Separating A, B, and C patterns in time shows non-target responses induced by earlier patterns. Red circles: stimulated cells (from A or B pattern). Gray circle: C cells, here unstimulated. 60 ms pattern duration. **(A)** Responses to AB sequence. Left, schematic; right, GCaMP responses. **(B)** Anatomical image (FOV; averaged over time). **(C)** Zoomed view shows suppression in unstimulated neuron (gray circle). **(D-F)** Excited non-target cells show linear summation, but suppressed cells show sublinear summation. **(D)** Excited cells from one example experiment (N=5 cells). Light red: response to C stimulus alone. Dark red: prediction, response to ABC sequence minus response to AB sequence. Response and prediction overlap. **(E)** Suppressed cells, with sublinear summation (N=5 cells; same expt). Same conventions as d. Response to C within the ABC sequence (ABC-AB, dark blue) is smaller than resp. to C alone (light blue; t-test, 300 ms period after stim onset, p=0.013). **(F)** Population data (N=7expts; N=130 cells, C cells only; compares ABC stim to AB+C stim as described in D-E). Blue: suppressed cells (response to AB stim<0;

t-test vs linear prediction, N=76 cells, p<0.001). Red: excited cells (response to AB stim>0; t-test, N=54 cells, p=0.84). These distributions are significantly different (K-S test, p<0.01). **(G)** Designing sequences to be boosted or attenuated based on non-target interaction mechanism: experimental schematic. We image responses to stimulation, and then select cells based on response to form the next pattern in the sequence. **(H)** Mean responses of boosted (pink) and attenuated (blue) sequences (N=20 cells per pattern, 3 patterns in the sequence; 30 ms pattern duration). **(I)** Distribution of responses of stimulated cells. Mean (gray horizontal line), median (black), means are sig. different (t-test, p=0.033; N=60 cells from N=3 patterns). **(J)** Population data (K-S test, p<0.00001, N=277 exc cells, N=270 supp cells, N=6 expts). **(K)** Sparsity is greater for boosted patterns. Each point: average across neurons. Sparsity in entire population is significantly greater in boosted sequences (t-test, p<0.05; N=6 expts; all cells, gray).

Designed input sequences produce predicted effects confirming the network mechanism

To provide an explicit test of this network mechanism, we constructed sequences that should be treated differently according to the mechanism described above.

To construct such sequences, we first measured responses to a single input pattern (Fig. 5G). We then created later stimulation patterns based on the non-target responses to those single patterns. We constructed sequences that should be ‘boosted’ by adding a pattern of stimulated cells that were excited by the first pattern, and constructed ‘attenuated’ sequences by adding a pattern of stimulated cells suppressed by the first pattern (Supp. Fig. 5G-J). Any cell within a 10 μm radius from a previously stimulated cell was excluded from selection to avoid pattern overlap. We repeated this step a second time to create boosted and attenuated sequences that were three patterns long.

If our constructed sequences, when delivered to the network, produce the expected excitation or suppression, this would be strong evidence that this recurrent mechanism underlies sequence modulation.

We studied this by playing back these sequences onto the cortical population. Confirming our hypothesis, we found that the sequence designed to be boosted produced a larger response than the sequence designed to be attenuated (Fig. 5H-J). This interaction between non-target responses induced by one pattern affecting later patterns is therefore a mechanism where some sequences can be enhanced and others suppressed.

However, one additional nonlinear phenomenon was present (Fig. 5A-E). Patterns designed to be boosted produced responses that were sparser than suppressed sequences — a number of outlier cells produced large responses that elevated the mean (Fig. 5I,K). This is a

kind of nonlinear amplification effect produced by later patterns falling on non-target cells that were excited by earlier patterns. This sparsity is a hallmark of responses to natural movies,[32,33,58] and supports the idea that the cortical network can selectively boost sequences of responses associated with natural vision. Thus we observe at least two kinds of nonlinearities that can contribute to sequence filtering, boosting some sequences and suppressing others: first, the attenuation-by-suppression effect that suppresses patterns falling on earlier non-target suppressed cells, and second, the increased strength and sparsity effect that can boost patterns that fall on earlier non-target excited cells.

In sum, experiments separating sequences in time (Fig. 5A-F), and constructing sequences designed to be excited or attenuated (Fig. 5G-K), together support a recurrent network mechanism for sequence tuning. Responses to an earlier pattern in a sequence create non-target responses in other local neurons that influence responses to later patterns. This modulation results in particular patterns, such as natural patterns (Fig. 3), producing larger responses compared to other responses that are attenuated (Fig. 6A).

Discussion

Here we find that individual V1 neurons' responses are modulated by inputs arriving to other V1 neurons earlier in time (Fig. 2). When we vary those prior inputs by changing prior visual stimuli, we find that earlier inputs boost responses to later inputs when the input sequence is derived from input sequences found in natural vision (Fig. 3). We show that a given flashed natural scene produces a response that lasts more than 100 ms (Fig. 3L), meaning the integration window for the interaction of one pattern with a future pattern can be at least this long. We then confirm that patterns interact across time via recurrent influences. To do this we design sequences of input patterns that should be boosted or attenuated if non-target responses to prior patterns interact with later patterns. We find that the resulting sequences are indeed selectively boosted or attenuated, as predicted. One intriguing result is that the boosted

sequences produce sparser responses than the attenuated sequences. This is consistent with the sparse responses to natural visual input observed by Gallant et al[32], supporting the idea that recurrent interactions do shape responses to natural visual input.

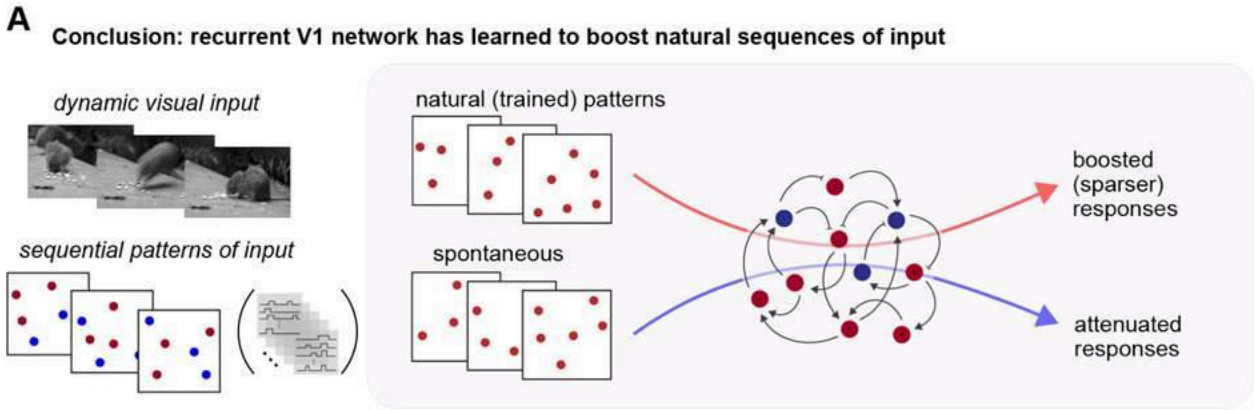


Figure 6: Mechanism and conceptual model from this work
(A) The recurrent network filters sequential inputs, boosting some and attenuating others. Errorbars: SEM.

Our results, taken together, show a computational role for the dense recurrent network in layer 2/3 of sensory (visual) cortex. In several cortical regions this dense recurrent connectivity is thought to generate ongoing or slow dynamics: for example, in motor cortex, recurrent dynamics are thought to be associated with muscle movement, and in prefrontal cortex, delay dynamics are associated with short-term memory[59,60]. In contrast, V1 does not generate complex sustained or ongoing dynamics. Here we find the V1 network does in fact support complex temporal processing, not by generating sequences but by filtering sequences of input, boosting natural sequences of patterns while attenuating non-natural sequences (Fig. 6A).

One feature of our results is matched responses and opto-only responses are similar in average magnitude, while unmatched responses are smaller than both on average (Fig. 3). One might have expected the cortex to explicitly amplify the response in the correct natural or sequence context, and instead, it appears at first glance that the cortical network may be

selectively suppressing non-matched input patterns. But this similarity to the opto-only responses could still be a signature of selective amplification, combined with a change in cortical state during natural vision. That is, the cortical network may produce responses that are on average smaller during complex visual input than during simple or static vision, so that both the matched and nonmatched responses are reduced relative to the opto-only case.

This would mean some sequences of patterns are indeed selectively boosted or amplified by the network based on the patterns that precede them, while natural vision also results in an overall, average or blanket suppression of incremental neural responses (gain reduction) that combines with that selective amplification. Supporting the idea that natural vision may lead to a blanket gain reduction compared to spontaneous activity states with simple or static vision, some data and models suggest that incremental responses decrease as cortical networks receive more input ([50,61,62], see also *Prior work* section below). This proposal of amplification paired with blanket response reduction would also be consistent with the evidence we found (Fig. 5) that sublinearity can be produced by suppressing neurons (see also [57]) — a blanket suppression of many neurons would suppress their responses to input while leaving responses to natural input unchanged, and thus larger than others. A possible alternative to this proposal could be a mechanism of selective suppression of some patterns, which then is deployed to suppress a large number of non-natural patterns — not in a blanket way, but for each non-natural pattern specifically. Given the larger space of non-natural patterns of activity compared to natural patterns of activity, selective suppression seems less likely than selective amplification combined with blanket suppression, but this is likely to be an interesting direction for future work.

Past work may not have identified this sequence filtering phenomenon because physiological recording methods are limited by the curse of dimensionality. That is, it is impractical, in finite experimental time, to show sufficient stimuli to fully estimate models of receptive fields that encode complex features of natural vision. The other relevant constraint

imposed by recording methods is the inability to easily determine which parts of neural responses to a stimulus come from feedforward input, and which come from local, recurrent interactions. Prior recording studies did observe that natural responses were stronger and sparser on average than expected from simple receptive field combinations.[32,33] But they could not identify that response gain varies depending on prior visual stimuli; this would have required either a complex receptive field model impractical to estimate, or another way to identify how a group of neurons was affected by different patterns of input to other local neurons. Here, two-photon stimulation allows us to match one specific excitatory pattern of input to the visual context in which it occurs. In that way we can test whether a single pattern influences another, and the direction of its effect, without needing to identify interacting patterns via sampling large numbers of patterns.

Relation to prior work: structure in spontaneous activity, straightening, cortical suppression, and hippocampal sequences

Prior observations that an echo of natural input can be seen in spontaneous activity[63–67] are consistent with what we find. During spontaneous activity, neurons in the cortex fluctuate[68–70] with measurable, but weak, correlation to natural visual responses. Our data suggest that while sequences of neural responses during spontaneous activity states are largely suppressed by the cortical network, some spontaneous patterns which partially match natural patterns can be weakly amplified by the recurrent network. However, because responses during spontaneous activity lack the structured feedforward driving inputs that occur during vision, spontaneous patterns are not amplified to the same extent as visual inputs, resulting in the relatively weak spontaneous correlations previously reported.

Martin and Douglas[3,4] noted the large number of local recurrent synapses in the cortex and hypothesized that they could be used for amplification of input. Our work confirms that hypothesis, and extends their ideas by showing the kinds of inputs that can be amplified, including complex temporal statistics.

Another related observation is the straightening of perceptual and neural responses seen in natural vision[71,72]. Straightening refers to the relationship between response patterns at different moments of a natural movie. Prior work has found that cortical responses at one moment are more geometrically similar to later responses than is seen in the cortical inputs. The sequence filtering by the recurrent network we find could well be the mechanism for the straightening effect.

There has also been some evidence for nonlinear spatiotemporal processing in V1 at the level of single neuron receptive fields. David, Reid, and a number of other investigators[34,44,73], (reviewed by [45]) have found in primate or carnivore V1 some non-separable or nonlinear interactions in single neuron receptive field centers. This nonlinearity, like the complex 'non-classical surround' in single neuron responses, could be a signature of nonlinear tuning at the population level for many interacting features, as occurs in natural vision when many RFs are stimulated together.

Driving one cell in visual cortex creates an average suppressive effect that falls off with distance from the stimulated cell (also seen for small input patterns,[74], [16], confirmed for balanced networks in simulation[50]). This suppression effect due to stimulating one cortical neuron is not at odds with our findings. The broad suppressive effect seen in these studies is observed by averaging across many neurons, and the average effect can be suppressive even as individual neurons respond in heterogeneous ways — sometimes excited and sometimes suppressed. We observe this heterogeneity in cells' responses, as patterned optogenetic stimulation excites some non-target, non-driven cells, and suppresses others (Fig. 5A, Supp. Fig. 5E). Moreover, simulation of balanced, strongly-coupled networks shows average suppression, but in some states and not others. In particular, single-cell stimulation combined with visual input produces average suppression, but single-cell stimulation in the absence of visual stimulation has a nearly zero average effect[50,75].

An additional related prior observation is the pervasive suppressive response at the end of external cortical stimulation, which is widely observed in electrical microstimulation studies.[76,77] This effect is that stimulation — input drive applied to the cortex — is followed by an average suppression that lasts for approximately 150 ms. This duration is consistent with the timescale of suppression we find following a flashed natural scene (Fig. 3L). Our results put forth a possible explanation for the previously-reported average suppression: that external stimulation triggers the non-target effects that we find for input patterns (Fig. 5). However, using cell-specific stimulation we find the firing rate of individual non-target cells can be either higher or lower based on the previous input pattern (Fig. 2). The average suppression previously reported reflects the average of many cells whose target (electrically stimulated) or non-target status is not known. Two-photon optogenetics allows us to separate the two groups of neurons since we can identify the cells receiving input. The post-stimulation suppressive modulation previously reported thus confirms the non-target effects we report here.

The two observations above, average suppression for single cell stimulation, and strong suppression for large-scale electrical stimulation, in fact argue that cortical incremental responses decrease as the cortical network receives more input, consistent with our proposal of selective amplification (matched vs. unmatched) combined with blanket suppression (matched/unmatched vs. opto-only).

Finally, our sequence effects are fundamentally different than the sequential replay that is observed in the hippocampus.[78] Our effect depends on firing rates over a few tens or hundreds of milliseconds, a distinct effect from the ordered spike sequences with millisecond-level precise timing that are generated during sharp-wave replay.[79,80]

The dense E-E cortical network in sensory cortex may encode the structure of natural sensation

Our data suggest the large number of excitatory-excitatory recurrent connections in the cortex are used for learning the structure of the natural world. Depending on which cells are

activated at a previous time, the recurrent network creates effects in other neurons that allow relative boosting (or suppression) of responses to particular input sequences.

The numerous excitatory recurrent connections in the cortex can, in principle, give the V1 network great capacity to selectively process the space of natural visual inputs.[36] The patterns of population responses to visual input are high-dimensional: any given frame from a natural movie can affect the firing of many thousands of neurons in the cortex. So there is a large space of neural population responses, forming the set of patterns of activity that result from many different frames or different visual stimuli. While some sorts of dendritic nonlinearities (discussion in[81]) and nonlinearities due to short-term synaptic plasticity[82] may also contribute, our data suggest that the high-dimensional space of visual input can be met with high-dimensional processing in the recurrent network. That is, the recurrent network has a large number of synapses which can act to set the non-target pattern produced by any input pattern. There are hundreds of millions of excitatory-excitatory recurrent synapses in a cubic millimeter of cortex,[1] and so the dense recurrent network of the cortex seems well-placed to support the space of transformations required for sequence filtering.

While specific inhibitory connections[83,84] may play a role in producing these non-target recurrent responses, no specific inhibitory population is needed to drive our effects. The suppression we observe could be generated via withdrawal of excitation. That is, the mechanisms we observe may be controlled primarily by excitatory-excitatory synapses, without being mapped to a set or subclass of inhibitory cells. In our experiments, we express opsin in excitatory cells only, so the interactions we observe must be caused by activation of excitatory cells. This excitatory activation can cause suppression in other excitatory cells in steady-state. For example, exciting excitatory cells may drive a few spikes in a number of inhibitory cells, which then increase their input to a large number of excitatory cells. In turn, the excitatory cells firing rates decrease, and that withdraws excitation from other connected excitatory cells. In this way the steady-state firing rates of the excitatory network can be primarily controlled by the

pattern of connections amongst the excitatory cells, not by specific inhibitory connections, and without subsets of inhibitory cells that are substantially activated (reviewed in [81], their Fig. 3). Computations that are not mapped in a 1:1 way to inhibitory subclasses are attractive for high-dimensional coding, since the space of excitatory-excitatory connections is larger than the space of inhibitory-excitatory connections, given the larger number of excitatory cells in the cortex. This type of mechanism, with excitatory activity in one population driving either excitation or suppression in other excitatory cells in the presence of balancing inhibition, is consistent with the balanced state or inhibition-stabilized state of cortical operation.[50,69,70,75,85,86] Future work should be able to determine the role of excitatory and inhibitory interconnectivity in recurrent cortical computations.

The sequence filtering computation is also a form of predictive processing.[27,87] Because responses to a given visual input are influenced by previous inputs, the natural, expected sequence of inputs is boosted. This predictive processing — changing the response to the next input based on the previous — is achieved by altering the network's input-output transformation, or network gain based, on prior inputs. This occurs via short-timescale non-target recurrent effects (Figs. 2,3), without large error or mismatch signals that might be expected from traditional predictive coding theories.[27,88–91]

Sequence filtering seems likely to occur in other sensory regions beyond visual cortex

While we have made these measurements in visual cortex, we might speculate that recurrent connections in other sensory regions — auditory, somatosensory, etc — also are used for active filtering.[81] Those sensory modalities also process natural inputs that vary in time, though often with faster changes than in vision, predicting different timescales of sequence filtering in other sensory regions. Additionally, associative brain areas downstream of primary sensory cortex can have tuning for complex sequences,[92] and may filter at longer timescales.[93]

The optimal rates of sequential input for greatest boosting, and the temporal precision of this filtering at different rates could be explored in future studies.

Beyond biology, recurrent networks in artificial systems also often are used to create temporally-structured computations. Our model findings (Fig. 4) are consistent with the idea that recurrent artificial networks can learn temporal statistics[94]. The results support the idea that when properly trained, densely-connected recurrent networks can perform sequence processing and learn temporal structure, not just in artificial systems but in biological brains. Current state-of-the art generative language models also learn temporal statistics, but using non-recurrent architectures [95]. Understanding the factors that favor different architectures for computation seems likely to be an important continuing question at the intersection of neuroscience and AI.

Conclusion

These data show a new and powerful purpose for recurrent connectivity in the sensory cerebral cortex: to confer sensitivity to sequential input. The visual cortical L2/3 network is sensitive to dynamic visual context, boosting responses to sequences of input corresponding to dynamic natural sensation, and suppressing others.

References

1. Braitenberg, V., and Schüz, A. (2013). Anatomy of the Cortex: Statistics and Geometry (Springer Science & Business Media).
2. Liu, Y.-J., Ehrengruber, M.U., Negwer, M., Shao, H.-J., Cetin, A.H., and Lyon, D.C. (2013). Tracing inputs to inhibitory or excitatory neurons of mouse and cat visual cortex with a targeted rabies virus. Curr. Biol. *23*, 1746–1755.
3. Douglas, R.J., Koch, C., Mahowald, M., Martin, K.A., and Suarez, H.H. (1995). Recurrent excitation in neocortical circuits. Science *269*, 981–985.
4. Douglas, R.J., and Martin, K.A.C. (2007). Recurrent neuronal circuits in the neocortex. Curr. Biol. *17*, R496–R500.
5. Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. *148*, 574–591.
6. Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.
7. Lien, A.D., and Scanziani, M. (2013). Tuned thalamic excitation is amplified by visual cortical circuits. Nat. Neurosci.
8. Chung, S., and Ferster, D. (1998). Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. Neuron *20*, 1177–1189.
9. Li, Y.-T., Ibrahim, L.A., Liu, B.-H., Zhang, L.I., and Tao, H.W. (2013). Linear transformation of thalamocortical input by intracortical excitation. Nat. Neurosci. *16*, 1324–1330.
10. Priebe, N.J., and Ferster, D. (2012). Mechanisms of neuronal computation in mammalian visual cortex. Neuron *75*, 194–208.
11. Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87–91.

12. Lee, W.-C.A., Bonin, V., Reed, M., Graham, B.J., Hood, G., Glattfelder, K., and Reid, R.C. (2016). Anatomy and function of an excitatory network in the visual cortex. Nature *532*, 370–374.

13. Rossi, L.F., Harris, K.D., and Carandini, M. (2020). Spatial connectivity matches direction selectivity in visual cortex. Nature *588*, 648–652.

14. Douglas, R.J., Martin, K.A.C., and Whitteridge, D. (1989). A canonical microcircuit for neocortex. Neural Comput. *1*, 480–488.

15. Marshel, J.H., Kim, Y.S., Machado, T.A., Quirin, S., Benson, B., Kadmon, J., Raja, C., Chibukhchyan, A., Ramakrishnan, C., Inoue, M., et al. (2019). Cortical layer–specific critical dynamics triggering perception. Science *365*, eaaw5202.

16. Oldenburg, I.A., Hendricks, W.D., Handy, G., Shamardani, K., Bounds, H.A., Doiron, B., and Adesnik, H. (2024). The logic of recurrent circuits in the primary visual cortex. Nat. Neurosci. *27*, 137–147.

17. Marr, D., and Hildreth, E. (1980). Theory of edge detection. Proc. R. Soc. Lond. B Biol. Sci. *207*, 187–217.

18. Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607–609.

19. Yoshida, T., and Ohki, K. (2020). Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. Nat. Commun. *11*, 872.

20. Ringach, D.L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. J. Neurophysiol. *88*, 455–463.

21. Olshausen, B.A., and Field, D.J. (2006). What is the other 85 percent of V1 doing. L. van Hemmen, & T. Sejnowski (Eds.) *23*, 182–211.

22. Kang, I., Talluri, B.C., Yates, J.L., Niell, C.M., and Nienborg, H. (2025). Is the impact of spontaneous movements on early visual cortex species specific? Trends Neurosci. *48*, 7–21.

23. Niell, C.M., and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. Neuron *65*, 472–479.

24. Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J.A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron *86*, 740–754.

25. Nienborg, H., and Cumming, B.G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature *459*, 89–92.

26. Krug, K. (2004). A common neuronal code for perceptual processes in visual cortex? Comparing choice and attentional correlates in V5/MT. Philos. Trans. R. Soc. Lond. B Biol. Sci. *359*, 929–941.

27. Keller, G.B., and Mrsic-Flogel, T.D. (2018). Predictive Processing: A Canonical Cortical Computation. Neuron *100*, 424–435.

28. Vetter, P., Smith, F.W., and Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. Curr. Biol. *24*, 1256–1262.

29. Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Benfenati, F., and Medini, P. (2012). Sound-driven synaptic inhibition in primary visual cortex. Neuron *73*, 814–828.

30. Ghazanfar, A.A., and Schroeder, C.E. (2006). Is neocortex essentially multisensory? Trends Cogn. Sci. *10*, 278–285.

31. Bimbard, C., Sit, T.P.H., Lebedeva, A., Reddy, C.B., Harris, K.D., and Carandini, M. (2023). Behavioral origin of sound-evoked activity in mouse visual cortex. Nat. Neurosci. *26*, 251–258.

32. Vinje, W.E., and Gallant, J.L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. Science *287*, 1273–1276.

33. Froudarakis, E., Berens, P., Ecker, A.S., Cotton, R.J., Sinz, F.H., Yatsenko, D., Saggau, P., Bethge, M., and Tolias, A.S. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. Nat. Neurosci. *17*, 851–857.

34. David, S.V., Vinje, W.E., and Gallant, J.L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. J. Neurosci. *24*, 6991–7006.

35. Cossell, L., Iacaruso, M.F., Muir, D.R., Houlton, R., Sader, E.N., Ko, H., Hofer, S.B., and Mrsic-Flogel, T.D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. Nature *518*, 399–403.

36. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K.D. (2019). High-dimensional geometry of population responses in visual cortex. Nature *571*, 361–365.

37. Li, W., and Gilbert, C.D. (2002). Global contour saliency and local colinear interactions. J. Neurophysiol. *88*, 2846–2856.

38. Olshausen, B.A., and Field, D.J. (2005). How close are we to understanding v1? Neural Comput. *17*, 1665–1699.

39. Benucci, A., Ringach, D.L., and Carandini, M. (2009). Coding of stimulus sequences by population responses in visual cortex. Nat. Neurosci. *12*, 1317–1324.

40. Tolhurst, D.J., Walker, N.S., Thompson, I.D., and Dean, A.F. (1980). Non-linearities of temporal summation in neurones in area 17 of the cat. Exp. Brain Res. *38*, 431–435.

41. Shamma, S. (2001). On the role of space and time in auditory processing. Trends Cogn. Sci. *5*, 340–348.

42. Theunissen, F.E., David, S.V., Singh, N.C., Hsu, A., Vinje, W.E., and Gallant, J.L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network *12*, 289–316.

43. Machens, C.K., Wehr, M.S., and Zador, A.M. (2004). Linearity of cortical receptive fields measured with natural sounds. J. Neurosci. *24*, 1089–1100.

44. DeAngelis, G.C., Ohzawa, I., and Freeman, R.D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. J. Neurophysiol. *69*, 1118–1135.

45. Ringach, D.L. (2004). Mapping receptive fields in primary visual cortex. J. Physiol. *558*, 717–728.

46. Ringach, D.L., Hawken, M.J., and Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. Nature *387*, 281–284.

47. Emiliani, V., Cohen, A.E., Deisseroth, K., and Häusser, M. (2015). All-Optical Interrogation of Neural Circuits. J. Neurosci. *35*, 13917–13926.

48. Packer, A.M., Peterka, D.S., Hirtz, J.J., Prakash, R., Deisseroth, K., and Yuste, R. (2012). Two-photon optogenetics of dendritic spines and neural circuits. Nat. Methods *9*, 1202–1205.

49. LaFosse, P.K., Zhou, Z., Friedman, N.G., Deng, Y., Li, A.J., Akitake, B., and Histed, M.H. (2023). Bicistronic Expression of a High-Performance Calcium Indicator and Opsin for All-Optical Stimulation and Imaging at Cellular Resolution. eNeuro *10*. https://doi.org/10.1523/ENEURO.0378-22.2023.

50. O'Rawe, J.F., Zhou, Z., Li, A.J., LaFosse, P.K., Goldbach, H.C., and Histed, M.H. (2023). Excitation creates a distributed pattern of cortical suppression due to varied recurrent input. Neuron *111*, 4086–4101.e5.

51. Bauer, J., Margrie, T.W., and Clopath, C. (2024). Movie reconstruction from mouse visual cortex activity. bioRxiv, 2024.06.19.599691. https://doi.org/10.1101/2024.06.19.599691.

52. Hecht, S., and Verrijp, C.D. (1933). The Influence of Intensity, Color and Retinal Location on the Fusion Frequency of Intermittent Illumination. Proc. Natl. Acad. Sci. U. S. A. *19*, 522–535.

53. Evaluation of critical flicker-fusion frequency measurement methods using a touchscreen-based visual temporal discrimination task in the behaving mouse (2019). Neurosci. Res. *148*, 28–33.

54. Umino, Y., Solessio, E., and Barlow, R.B. (2008). Speed, Spatial, and Temporal Tuning of Rod and Cone Vision in Mouse. J. Neurosci. *28*, 189.

55. Bonin, V., Histed, M.H., Yurgenson, S., and Reid, R.C. (2011). Local diversity and fine-scale organization of receptive fields in mouse visual cortex. J. Neurosci. *31*, 18506–18521.

56. Smith, S.L., and Häusser, M. (2010). Parallel processing of visual space by neighboring neurons in mouse visual cortex. Nat. Neurosci. *13*, 1144–1149.

57. LaFosse, P.K., Zhou, Z., O'Rawe, J.F., Friedman, N.G., Scott, V.M., Deng, Y., and Histed, M.H. (2024). Cellular-resolution optogenetics reveals attenuation-by-suppression in visual cortical neurons. Proc. Natl. Acad. Sci. U. S. A. *121*, e2318837121.

58. Sparse coding of sensory inputs (2004). Curr. Opin. Neurobiol. *14*, 481–487.

59. Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K.V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. Nat. Neurosci. *18*, 1025–1033.

60. Fujisawa, S., Amarasingham, A., Harrison, M.T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nat. Neurosci. *11*, 823–833.

61. Ahmadian, Y., Rubin, D.B., and Miller, K.D. (2013). Analysis of the stabilized supralinear network. Neural Comput. *25*, 1994–2037.

62. Sanzeni, A., Histed, M.H., and Brunel, N. (2020). Response nonlinearities in networks of spiking neurons. PLoS Comput. Biol. *16*, e1008165.

63. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. Nature *425*, 954–956.

64. Luczak, A., Barthó, P., and Harris, K.D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron *62*, 413–425.

65. Tsodyks, M., Kenet, T., Grinvald, A., and Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. Science *286*, 1943–1946.

66. Nguyen, N.D., Lutas, A., Amsalem, O., Fernando, J., Ahn, A.Y.-E., Hakim, R., Vergara, J., McMahon, J., Dimidschstein, J., Sabatini, B.L., et al. (2024). Cortical reactivations predict future sensory responses. Nature *625*, 110–118.

67. Yao, H., Shi, L., Han, F., Gao, H., and Dan, Y. (2007). Rapid learning in cortical coding of visual scenes. Nat. Neurosci. *10*, 772–778.

68. Shadlen, M.N., and Newsome, W.T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J. Neurosci. *18*, 3870–3896.

69. van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science *274*, 1724–1726.

70. Brunel, N. (2000). Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons. J. Comput. Neurosci. *8*, 183–208.

71. Hénaff, O.J., Goris, R.L.T., and Simoncelli, E.P. (2019). Perceptual straightening of natural videos. Nat. Neurosci. *22*, 984–991.

72. Hénaff, O.J., Bai, Y., Charlton, J.A., Nauhaus, I., Simoncelli, E.P., and Goris, R.L.T. (2021). Primary visual cortex straightens natural video trajectories. Nat. Commun. *12*, 5982.

73. Reid, R.C., Soodak, R.E., and Shapley, R.M. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. J. Neurophysiol. *66*, 505–529.

74. Chettih, S.N., and Harvey, C.D. (2019). Single-neuron perturbations reveal feature-specific competition in V1. Nature *567*, 334–340.

75. Sanzeni, A., Palmigiano, A., Nguyen, T.H., Luo, J., Nassi, J.J., Reynolds, J.H., Histed, M.H., Miller, K.D., and Brunel, N. (2023). Mechanisms underlying reshuffling of visual responses by optogenetic stimulation in mice and monkeys. Neuron *111*, 4102–4115.e9.

76. Butovas, S., and Schwarz, C. (2003). Spatiotemporal effects of microstimulation in rat neocortex: a parametric study using multielectrode recordings. J. Neurophysiol. *90*, 3024–3039.

77. Sombeck, J.T., Heye, J., Kumaravelu, K., Goetz, S.M., Peterchev, A.V., Grill, W.M., Bensmaia, S., and Miller, L.E. (2022). Characterizing the short-latency evoked response to intracortical microstimulation across a multi-electrode array. J. Neural Eng. *19*. https://doi.org/10.1088/1741-2552/ac63e8.

78. Davidson, T.J., Kloosterman, F., and Wilson, M.A. (2009). Hippocampal replay of extended experience. Neuron *63*, 497–507.

79. Roumis, D.K., and Frank, L.M. (2015). Hippocampal sharp-wave ripples in waking and sleeping states. Curr. Opin. Neurobiol. *35*, 6–12.

80. Buzsáki, G. (1986). Hippocampal sharp waves: their origin and significance. Brain Res. *398*, 242–252.

81. Histed, M.H. (2025). Active filtering: a predictive function of recurrent circuits of sensory cortex. arXiv [q-bio.NC].

82. Buonomano, D.V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. Nat. Rev. Neurosci. *10*, 113–125.

83. Sadeh, S., and Clopath, C. (2020). Patterned perturbation of inhibition can reveal the dynamical structure of neural processing. Elife *9*. https://doi.org/10.7554/eLife.52757.

84. Znamenskiy, P., Kim, M.-H., Muir, D.R., Iacaruso, M.F., Hofer, S.B., and Mrsic-Flogel, T.D. (2024). Functional specificity of recurrent inhibition in visual cortex. Neuron *112*, 991–1000.e8.

85. Ahmadian, Y., and Miller, K.D. (2021). What is the dynamical regime of cerebral cortex? Neuron *109*, 3373–3391.

86. Sanzeni, A., Akitake, B., Goldbach, H.C., Leedy, C.E., Brunel, N., and Histed, M.H. (2020). Inhibition stabilization is a widespread property of cortical networks. Elife *9*. https://doi.org/10.7554/eLife.54875.

87. Rao, R.P., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. *2*, 79–87.

88. den Ouden, H.E.M., Kok, P., and de Lange, F.P. (2012). How prediction errors shape perception, attention, and motivation. Front. Psychol. *3*, 548.

89. Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. Annu. Rev. Neurosci. *23*, 473–500.

90. Lindsay, G. (2021). The Challenge of Proving Predictive Coding. Simons Foundation Global Brain.

91. Furutachi, S., Franklin, A.D., Aldea, A.M., Mrsic-Flogel, T.D., and Hofer, S.B. (2024). Cooperative thalamocortical circuit mechanism for sensory prediction errors. Nature *633*, 398–406.

92. Purandare, C., and Mehta, M. (2023). Mega-scale movie-fields in the mouse visuo-hippocampal network. Elife *12*. https://doi.org/10.7554/eLife.85069.

93. Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. Nat. Neurosci. *17*, 1661–1663.

94. Graves, A. (2013). Generating Sequences With Recurrent Neural Networks. arXiv [cs.NE].

95. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. arXiv [cs.CL].

96. Gorski, J.A., Talley, T., Qiu, M., Puelles, L., Rubenstein, J.L.R., and Jones, K.R. (2002). Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. J. Neurosci. *22*, 6309–6314.

97. Ye, Z., Bull, M.S., Li, A., Birman, D., Daigle, T.L., Tasic, B., Zeng, H., and Steinmetz, N.A. (2023). Brain-wide topographic coordination of traveling spiral waves. bioRxiv, 2023.12.07.570517. https://doi.org/10.1101/2023.12.07.570517.

98. Goldbach, H.C., Akitake, B., Leedy, C.E., and Histed, M.H. (2021). Performance in even a simple perceptual task depends on mouse secondary visual areas. Elife *10*. https://doi.org/10.7554/eLife.62156.

99. Giovannucci, A., Friedrich, J., Gunn, P., Kalfon, J., Brown, B.L., Koay, S.A., Taxidis, J., Najafi, F., Gauthier, J.L., Zhou, P., et al. (2019). CaImAn an open source tool for scalable calcium imaging data analysis. Elife *8*. https://doi.org/10.7554/eLife.38173.

100. Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L.F., Dalgleish, H., Carandini, M., and Harris, K.D. (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. bioRxiv, 30. https://doi.org/10.1101/061507.

101. Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data. PLoS Comput. Biol. *13*, e1005423.

102. Natural Movies https://observatory.brain-map.org/visualcoding/stimulus/natural_movies.

103. Burgess, C.P., Lak, A., Steinmetz, N.A., Zatka-Haas, P., Bai Reddy, C., Jacobs, E.A.K., Linden, J.F., Paton, J.J., Ranson, A., Schröder, S., et al. (2017). High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. Cell Rep. *20*, 2513–2524.

104. Steinmetz, N.A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. Science *372*. https://doi.org/10.1126/science.abf4588.

105. Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv [cs.LG]. https://doi.org/10.48550/ARXIV.1412.6980.

106. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. arXiv [cs.LG]. https://doi.org/10.48550/arXiv.1912.01703.

Acknowledgments

We thank B. Averbeck, M. Stopfer, A. Levine, S. David and members of the Histed Lab for comments on the manuscript and discussion.

Funding:

National Institutes of Health grant U19NS107464 (MH)

National Institutes of Health intramural project ZIAMH002956

Pew Biomedical Scholars Program (NAS)

Klingenstein-Simons Fellowship in Neuroscience (NAS)

Author contributions

Conceptualization: CD, ZZ, MH, PKL

Methodology: CD, ZZ, PKL, YD, SM, NS, MH

Investigation: CD, ZZ, SM, PKL

Formal Analysis: CD, ZZ

Visualization: CD, PKL, ZZ

Funding acquisition: MH, NS

Project administration: MH, NS

Supervision: MH, NS

Writing – original draft: CD, MH

Writing – review & editing: CD, MH, SM, NS

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data and code will be made available on DANDI at the time of publication.

Methods

Animals

All experiments were conducted in adherence to NIH and Department of Health and Human Services (HHS) guidelines for animal research and were approved by the Institutional Animal Care and Use Committee (IACUC) at the relevant institutions (optogenetic experiments: NIMH; Neuropixels experiments: University of Washington.)

Optogenetic experiments

Emx1-Cre animals (N=16, https://www.jax.org/strain/005628)[96] of both sexes were used. No systematic differences were observed between males and females. Viral injection and window implants were done at ages 2-7 months. After procedures, animals were singly housed on a reverse 12-hour dark/light cycle.

Neuropixels experiments

Male C57BL/6J mice (N=5), aged two to three months, were used. After procedures, animals were singly housed on a standard 12-hour dark/light cycle.

All animals were put on water schedule five or more days after head-plate surgery, and their weights were carefully monitored to ensure they remained above 85% of their baseline body weight.

Viral injection and cranial window implants

We performed injections and implants as described in [49] (optogenetic experiments; NIMH) or [97] (2023; Neuropixels; University of Washington). Minor differences in the two sets of approaches are not expected to impact results.

Optogenetic experiments

Briefly, mice were given dexamethasone (3.2 mg/kg) 30 minutes before surgery and anesthetized with 1-3% isoflurane (in 100% O^2). A titanium headplate was affixed using C&B Metabond (Parkell) and a 3 mm craniotomy was made over the primary visual cortex (-3.1 mm ML, +1.5 mm AP from lambda) and a glass optical window chronically inserted. Mice were injected with either a mixture of two AAV viruses to induce GCaMP and stChrimsonR into the cells (mixed virus injections), or a bicistronic virus expressing both GCaMP and stChrimsonR in each cell.[49] For the mixed injection, the fluorescent calcium indicator, (AAV9-syn-jGCaMP8s; titers: 5.0-10x10^{12} genome copies (GC)/ml) and the soma-targeted opsin (AAV9-hsyn-DIO-stChrimsonR-mRuby2; titer: 3.0x10^{12} GC/ml) were mixed in phosphate-buffered saline. In the case of the bicistronic virus injections, we used AAV9-hsyn-DIO-jGCaMP8s-p2a-stChrimsonR (titers: 2.9-5.9x10^{12} GC/mL). For each mouse, 3 to 5 injections (100 nL/min, 200 μm depth) were made to cover a wider area of the cortex. A custom made light-blocking cap was fixed onto the implant to limit ambient light exposure and prevent debris from contacting the window. Experiments began three weeks after these procedures.

Neuropixels experiments

Briefly, mice were induced into anesthesia with 5% isoflurane, and maintained at 2-3%. Carprofen (5 mg/kg) and Lidocaine (2 mg/kg) were administered for analgesia. A titanium head plate and a 3D-printed chamber were affixed using Metabond. Carprofen was administered at 0.05 mg/ml in the water post-surgery for three days. Five days following recovery, mice were put on water restriction. After two days, they were habituated to head fixation for two days, during which they received up to thirty random 5μL sucrose water 10% rewards. Subsequently, animals were exposed to a set of 20 natural images presented over 1000 trials, followed by a 5μL sucrose water reward for five days. This step was performed to provide a control training condition for another experiment. None of the 20 images in this step were used as part of the experiment mentioned in this work. One or two days before recordings, a 2 x 2 mm craniotomy was performed over the primary visual cortex using a dental drill and stereotaxic techniques. The area was then covered with silicone gel. This process was conducted under anesthesia procedures used for the head plate implant, and Carprofen (5 mg/kg) was given for analgesia. A cap was placed on the chamber to prevent debris from getting into the area around the craniotomy.

Retinotopic mapping

Before optogenetic experiments, we determined the location of V1 in the cranial window using a hemodynamic intrinsic imaging protocol previously described in [98]. Briefly, small visual stimuli were presented while 530 nm light was delivered using a fiber-coupled LED (M350F2; Thorlabs, Newton NJ). Hemodynamic response was calculated as the change in reflectance of the cortical surface between the baseline period and a response window starting 3 ms after stimulus onset. Imaging was done on a Zeiss Discovery stereo microscope with a 1x widefield objective through a green long-pass emission filter, acquired at 2Hz. An average retinotopic map was fit to the cortical responses based on the centroids of the hemodynamic response for each stimulus location.

Two-photon holographic imaging and stimulation

Two-photon imaging and stimulation procedures are described in detail in [49]. Briefly, animals were awake and alert under a 16x water-immersion objective (Nikon; Tokyo, Japan; NA=0.8) manually positioned over the implanted optical window. Imaging was done with a custom-built microscope and controlled by ScanImage software in MATLAB (The Mathworks, Natick, MA). Calcium responses were measured ~100-200 μm below the surface of the pia (L2/3 of V1) with an imaging field of view of 414 x 414 μm. Imaging was performed using 920 nm wavelength light at 15-20 mW and frames acquired at 30 Hz.

Holographic stimulation was performed using a femtosecond pulsed laser (1030 nm, Satsuma, Amplitude Laser, or 1040nm, Monaco, Coherent Inc.) A spatial light modulator (SLM) shaped the laser wavefront to create stimulation patterns (10 μm diameter disks; 2.5-5.3mW/target, 500 kHz pulse rate). Timing of sequential pattern presentation was selected to mimic the temporal structure of natural visual inputs: pattern duration was 30, 60, 91, or 121 ms (1-4 imaging frames; frame duration approx 30.1 ms), and overlap between patterns (0-1 frames, 0-30 ms) was sometimes used to simulate the effect of a dynamic visual stimulus, which can change more frequently than pattern period Overlap in patterns: Fig. 1IJ, 30ms; Fig. 2, 0 ms; Fig 3 AB x ms, Fig. 4 (simulation) 0 ms ETC The response to a 50 ms flashed single visual input lasts around 70 ms (Fig. 3L), overlapping with the next input from a natural visual stimulus. The radial point-spread function (PSF) of diffraction limited spots generated by the SLM was 9.4 μm and the axial PSF was 54 μm.The laser was gated on during horizontal flyback periods and off during the imaging pixel acquisition to allow for approximately simultaneous stimulation and imaging. Reported stimulation power is the average power over on and off periods (i.e. reduced by a factor of 0.3 from the laser power measured without this gating.) To minimize crosstalk we chose stChrimsonR opsin. Previous work with the bicistronic virus used here has not shown evidence of crosstalk activation[49]. This work additionally characterized off-target stimulation and found the majority of off-target stimulation occurs in cells within this 10 μm radius when

stimulating ~20 or fewer cells. Thus, we generally avoid selecting cells whose centroid is within a 10 μm of another selected cell.

Visual responses were measured in awake, head-fixed mice viewing flashed single natural image frames (40° circular mask with neutral gray background; 120 ms) or full natural movie stimuli (full-field, 2 seconds, 30Hz frame rate) presented on an LCD monitor. Mice were given small volumes of water on 20% of trials (3 μL).

Two-photon data analysis

We performed motion correction using the CaImAn toolbox[99] and cell segmentation using Suite2p.[100] All data analysis was done in Python (https://www.python.org). For pixel based analyses, we computed $\Delta F/F^0$ ($\Delta F/F^0$; F: raw fluorescence intensity; F^0: average fluorescence across the 45 imaging timepoints, 1.5 s, prior to stimulus presentation) at every pixel of the image stacks for display across the FOV. The $\Delta F/F^0$ for all time courses was calculated from

fluorescent traces output from Suite2p. Neuropil is subtracted during the Suite2p preprocessing to account for extraneous signals. However by design, because the stimulations occur at different time points, changes in pixel response due to cell body or dendrites would be modulated by temporal context thus the conclusion would be unchanged. Off-target suppression measured in Fig. 5 would only be diluted by dendritic activation.

For full sequence trials (ABC, CBA), averaging windows for each pattern (A, B, C) started at stimulation onset of that pattern and lasted 300 ms; shifted for pattern order so the response window always began on the first frame of stimulation of that pattern (raw time course data without shift is shown in Fig. 2B). To isolate the optogenetic response following visual input, we subtracted off the response to the visual stimulus alone, [(vis+opto)-vis] (Fig. 3). Averaged response interval for individual cells began at optogenetic stimulus onset and ended at stimulus offset (120 ms, Fig. 3K). Activity of the C pattern in ABC sequence was calculated as [ABC - AB] (Fig. 5D,E). In Figure 5I-K, the averaging interval for individual cell responses started at stimulus

onset and lasted 200 ms. Response intervals for all pixel based analyses started at stimulus onset and lasted 300 ms (Fig. 3B-H, 5A). Deconvolution (Fig. 1J) was done using OASIS with an autoregressive constant of 1.[101] Population sparsity[32] (Fig. 5K) was calculated as:

$$S = \{1 - [(\sum r/n)^2 / \sum (r^2/n)]\}/[1 - (1/n)]$$

Where r is the cell responses in the frame following the last stimulation and n is the number of cells.

Statistics

All statistical analyses were performed using Python. Statistical tests used include t-tests (Fig. 2D, Fig. 5D-F,I,K), Kolmogorov-Smirnov and Anderson-Darling tests (Fig. 2E,N-O, Fig. 5F,J), Mann-Whitney U tests (Fig. 2G-L, Fig. 3K) and Mood's median test (Fig. 2Q). Significance threshold was held at α=0.05; n.s., not significant (p>0.05); *p≤0.05, **p≤0.01, ***p≤0.001. All experiments were replicated in multiple animals.

Cell pattern selection

For analysis of visual response to flashed natural images, individual cell responses were averaged over a 300 ms interval starting from the stimulus onset. Cells activated by the visual stimulus were defined by responses exceeding a 5% $\Delta F/F^0$ threshold. All cells that met this threshold were included in the optogenetic pattern (11-15 cells per pattern, Supp. Fig. 3D,E). Data includes three randomly selected time points across two natural movies (Mouse movie: Supp. Video 1; Touch of Evil Movie[102]).

To construct the boosted and attenuated patterns (Fig. 5), we first selected at random (not based on visual responses) the first 15-20 cell pattern in the sequence. Average $\Delta F/F^0$ response to stimulation of that pattern was calculated across a 300 ms time interval starting at stimulation onset. To select the latter patterns, cells within a 10 μm radius of a previously

stimulated cell were excluded to control for off-target effects. Excited non-target cells meeting a minimum threshold of 5% $\Delta F/F^0$ were included in the next pattern for the boosted sequence (15-20 cells per pattern). Conversely, suppressed non-target cell responses below -5% $\Delta F/F^0$ (15-20 cells per pattern) were selected for the attenuated sequence pattern. If more than 20 cells met these criteria, the 20 cells with the largest magnitude response were used. This process was repeated for the two sets of first and second patterns to generate the third pattern for each sequence. Each pattern in both sequences had an equal number of cells. Any cells included in previous patterns were excluded from selection for the subsequent patterns (Supp. Fig. 5G-J).

Electrophysiology

For Neuropixels electrophysiology the mice were head-fixed for the recordings while seated in plastic support with forepaws on a wheel,[103] with the ability to move body parts other than the head but no ability to locomote. The behavioral state of the animals was monitored using two Basler acA2440-75um cameras at 560x560 resolution. One camera focused on the eye area was used to monitor the pupil diameter, and the other focused on the face was used to detect licks and ensure the animal was not in distress. Animals were awake during the duration of the experiment. Recordings were conducted from all layers of the primary visual cortex using Neuropixel 2.4 probes[104] with a sampling rate of 30 kHz. The recording sites were targeted based on the stereotaxic coordinates in the CCF Allen atlas using the Pinpoint system.[97] The reference for the recordings was set at the tip of the electrode. Three recordings were performed: one from the first animal and two (on subsequent days) from the second animal. Spike sorting was executed using Kilosort 2.5 (https://github.com/MouseLand/Kilosort; RRID: SCR_016422)[104] on each of the four shanks separately.

Visual stimuli were presented on three 60 Hz screens (LG LP097QX1), surrounding the mice and covering 270 x 70 degrees (azimuth x elevation) visual angle. Each trial featured a

natural image from a set of two, sourced from the Allen Institute Brain Observatory, and repeated across all screens. These images were equiluminant and underwent histogram equalization for contrast consistency. Images were warped to appear rectilinear from the animals' viewpoint. Each image was displayed 30 times for 50 ms. A reward of 5 µL of 10% sucrose water was administered at 1150 ms post-stimulus offset to maintain engagement. The trials were randomly interleaved with each other and with additional trials from another experiment, which also included natural image presentations and sucrose rewards. The inter-trial interval followed an exponential distribution with a minimum and mean of 2 and 2.6 seconds, respectively.

All electrophysiology analysis was done in Python. Baseline value computed in a 40 ms window ending 5 ms prior to stimulus presentation. Response value was calculated across a 50 ms interval beginning 5 ms after stimulus presentation. Excited cells are those with average response greater than baseline; suppressed cells those with response less than baseline.

RNN Model

We trained a recurrent neural network (RNN) consisting of N=500 units, whose input dynamics for the i-th neuron are given by:

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} W_{ij}^{rec} \phi(x_i) + I_{im}(t) + \eta_i$$

The input structure is defined by:

$$I_{i,m}(t) = w_i^{in} \cdot \sum_{k=1}^{N_k} P_{k,m}^{wf}(t) \cdot P_{i,k}^{pat}$$

Where P_{km}^{wf} is a matrix specifying the input waveform for each pattern, *k*, and sequence, *m*. P_{ik}^{pat} = {0,1} is an indicator matrix denoting which cell i is in each pattern *k*. The weights of

the input pattern w_{in} are positive and exponentially distributed for a fraction p = 0.3 of units, and zero otherwise.

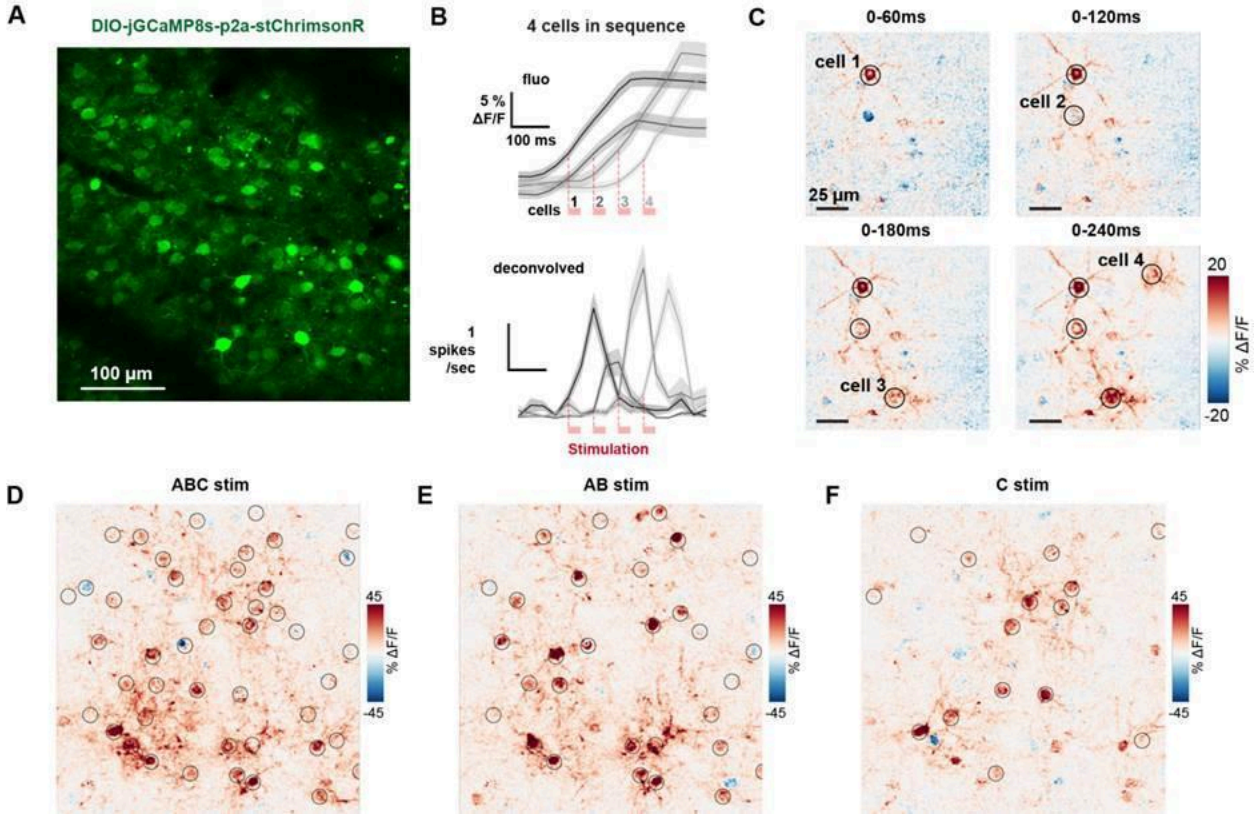
The readout of the network is defined as:

$$Z_i(t) = w_i^{out}\phi(x_i(t))$$

The transfer function of single units is φ(x) = tanh(x). The readout weights, w_{out}, are an Identity matrix. The initial recurrent weights W_{ij}^{rec}, before any training, are independently sampled from a random Gaussian distribution with a mean zero and standard deviation $g_0/\sqrt{N}$. The noise term η_i is randomly sampled from a zero mean distribution with standard deviation 0.0005 at every time step.

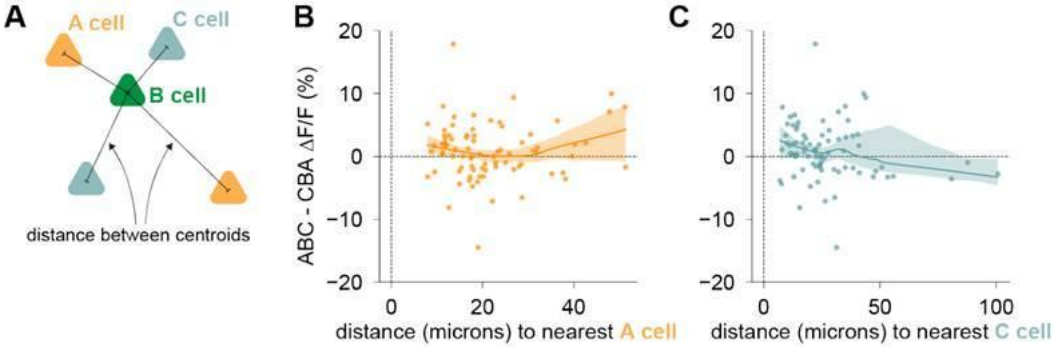
We trained the recurrent weights W_{ij}^{rec} of the RNN using backpropagation-through-time (ADAM optimizer in PyTorch)[105,106] such that the network readout Z for designated "trained sequences" matches a sparse scaled version of the time-varying input $I_{im}(t)$. To mimic the strong sparse activity seen in response to natural movies (and in Fig. 5G-K), we set the target output so 35% of the cells in the pattern (7 of 20) have a target level that is 2x the input and the other 65% (13/20) have a target that is 0.9x their input. The network readout Z for "control sequences" was set to match the time-varying input without any scaling. Sequences were generated by selecting 5 patterns from a set of 18 distinct input patterns. Each pattern had independently sampled weights from the exponential distribution. Untrained sequences were scrambled versions of the numerically ordered trained sequences. The input and output weights remained fixed. We trained the network on 500 epochs to produce these strong sparse responses. Control training had a target readout that exactly matched the time-varying input for all sequences. Parameters: τ = 60 ms, g_0 = 0.8, Euler integration timestep Δt = 1 ms, learning rate 0.01.

Supplemental figures



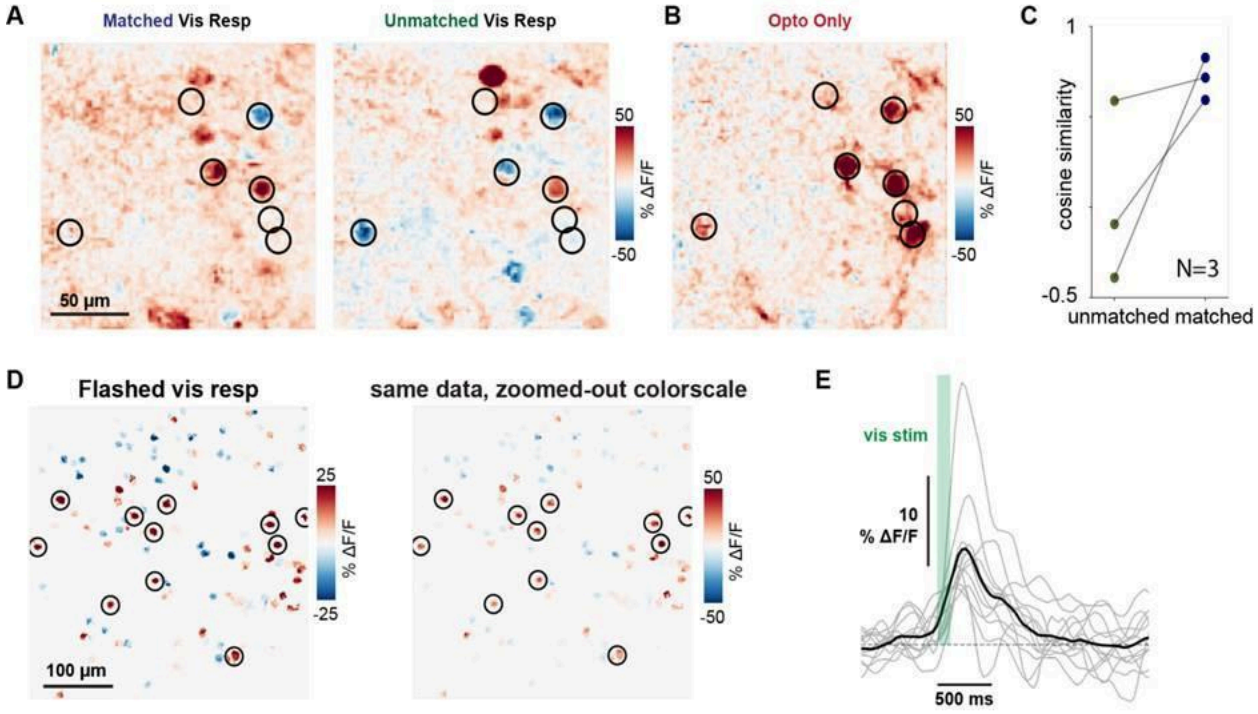
Supplementary Figure 1: Two-photon stimulation responses: timing and individual patterns.

(A) Example FOV image in L2/3 V1 with single virus injection, DIO-jGCaMP8s-p2a-stChrimsonR; this uses GCaMP8s which has even faster dynamics than GCaMP7s; more characterization in [49]. **(B)** Sequential stimulation of 4 individual cells (30 ms per stim, 30 ms between stim). Fluorescent calcium traces above; deconvolved traces below. **(C)** Response maps from data in A. Stimulation pattern lasts 30 ms each in sequence triggered at times: 0 ms, 60 ms, 120 ms, 180 ms. b-c use preparation with expression of indicator and opsin in two viruses, AAV9-syn-jGCaMP7s and AAV9-syn-DIO-stChrimsonR. **(D-F)** Response maps of ABC, AB, and C pattern stimulation, using single bicistronic virus (jGCaMP8s, stChrimsonR, Methods). Associated with Figure 1 and 2.



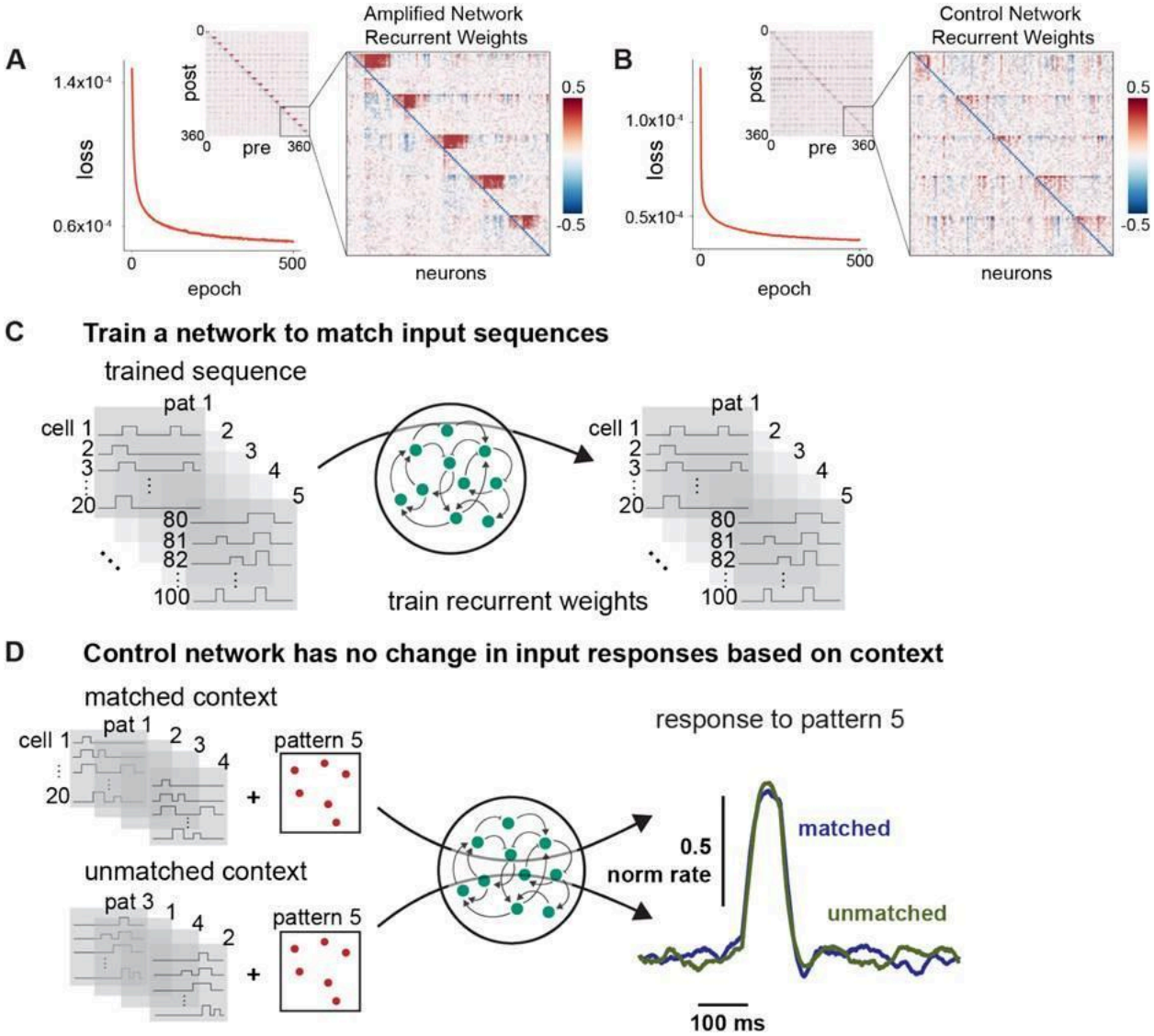
Supplementary Figure 2: Sequential context effects are not dependent on distance to previously stimulated cells.

(A) For each B cell, the euclidean distance between the centroids of that B cell and all A and C cells was calculated. The nearest A and C cells are selected for each B cell and plotted in the next panel. **(B)** The minimum distance between each B cell to A cell plotted (x-axis) against the difference in response of each B cell in ABC vs CBA (y-axis) **(C)** Same as panel (B) for B to C cells. No correlation is found between the distance from the nearest stimulated cell and the response within the sequence (Linear regression; B to A: r = 0.08, p-val = 0.5; B to C: r = -0.17, p-val = 0.13). Associated with Figure 2.



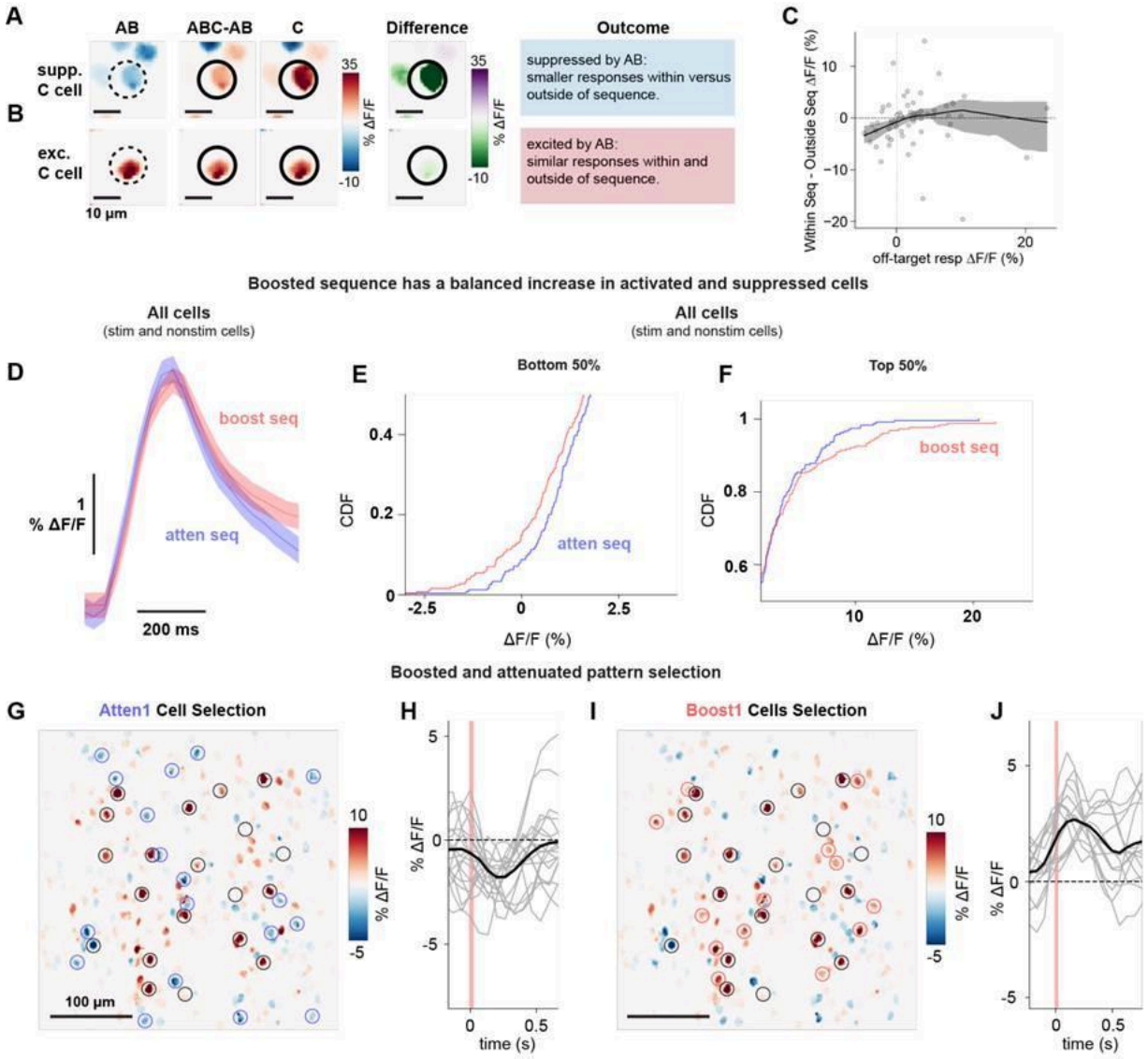
Supplementary Figure 3: Boosting is not due to spike or indicator saturation.

(A) Visual response alone of the same cells as shown in Fig. 3 for both matched and unmatched contexts at the time the optogenetic stim would be presented. **(B)** Optogenetic response of same cells without prior visual stimulus. **(C)** Greater cosine similarity between the optogenetic response alone and the visual responses alone at the time the optogenetic stim would be presented in the matched context than unmatched across experiments. **(D-E)** Cell selection for the natural-image-derived optogenetic pattern. **(D)** Response to flashed single frame; stimulated cells circled in black. Right: same data as left with un-zoomed color scale to show the larger cell responses that were chosen for optogenetic stimulation. **(E)** Timecourses of the circled cells in D**.** Only cells with a mean response above 5% are included in the pattern (Methods). Associated with Figure 3.



Supplementary Figure 4: RNN training for both amplified and control networks.

(A) Loss plot for amplified network training (left) and trained recurrent weights for neurons in the sequences (right). Inset highlighting a single sequence (5 patterns, 20 neurons per pattern). **(B)** Same as (A) for the control network. **(C-D)** Control training paradigm for the artificial recurrent neural network. **(C)** Target output matched the input dynamics, no boosting. **(D)** Context dependent changes are eliminated when the model is not trained to selectively amplify sequences. Associated with Figure 4.



Supplementary Figure 5: Constructed boosted and attenuated sequences have similar mean population responses.

(A) Prior suppression attenuates the stimulated response. Example non-stimulated suppressed cell in response to AB stim. Dotted black circle: nonstimulated cell; Solid black circle: stimulated cell. **(B)** Prior excitation leaves the incremental response largely unchanged. Example non-stimulated excited cell in response to AB stim with same comparison within versus outside the sequence. **(C)** More suppression from prior stim leads to stronger attenuation of the cell response within the sequence. Lowess fit (95% confidence interval) of C cell mean responses within vs. outside of the sequence ((ABC-AB) - C) depending on the non-target modulation from prior stimulation (AB). **(D)** While the stimulated cells show excitation/suppression (Fig. 5) as expected based on how they were selected, the

overall population response mean does not appreciably change. Both stimulated and non-target (non-stimulated) cells. Average response (mean +/- SEM) of all cells in the FOV for the attenuated sequence (blue) and boosted sequence (pink). The mean population response for the boosted and attenuated sequences is similar. **(E-F)** Increased responses in the boosted sequence are balanced by an increased number of suppressed cell responses. Bottom 50% of cell responses for all cells in the FOV (D) for the attenuated sequence (blue) and boosted sequence (pink). Top 50% of cell responses (E). **(G-J)** Non-target patterns of suppression and excitation are selected for subsequent sequence patterns. (G) First arbitrary pattern (N=20 cells; pattern A) that was used to initiate sequence design circled in black. Selected suppressed (attenuated) cells circled in blue. This pattern contains cells that are suppressed by prior stimulation. (H) Timecourses of individual suppressed cells (from f) to A pattern stimulation. Mean trace (black) shows the mean is suppressed; suppressed cells were selected. (I) Same data as in C with boosted pattern cells shown. Boosted pattern: selected cells (red circles) excited by A stimulation. This pattern contains cells that are activated by prior stim. (J) Response timecourses of excited cells (F) in response to A pattern stimulation. Mean trace in black. Associated with Figure 5.

Chapter 4: Prediction Content Shapes Population Representations in Mouse Primary Visual Cortex

Saghar Mirbagheri[1], Rajesh P. N. Rao[2], Nicholas A. Steinmetz[1]

[1] Department of Physiology and Biophysics, University of Washington, Seattle, WA, USA

[2] Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, USA

Abstract

The hierarchical predictive coding framework proposes that perception arises through interactions between top-down expectations and sensory input in early sensory areas. In sensory cortex, unexpected stimuli are often associated with enhanced neural responses, but it remains unclear whether these effects reflect prediction-specific computations or instead arise from other prediction-related signals such as local sensory adaptation, saliency, attention, novelty, temporal context, or nonspecific surprise. Here, we investigated how prediction alters neural representations in mouse primary visual cortex (V1) using a visual sequence paradigm combined with large-scale Neuropixels recordings across cortical layers. Mice viewed familiar image sequences containing omission and stimulus substitution trials, alongside matched control conditions designed to dissociate prediction from stimulus history and temporal context while controlling for novelty, rarity, saliency, and magnitude of expectation violation. Unexpected stimuli elicited response patterns consistent with previous reports, including omission responses and enhanced responses to deviant stimuli. However, these effects were not specific signatures of prediction, since omission response magnitude did not differ when sequence order was disrupted and enhanced deviant responses were not observed when the deviant image was not rare. Nevertheless, population decoding analyses revealed prediction-specific representational changes, such that identical unexpected stimuli evoked dissociable neural population states depending on prediction content when expectation violation magnitude was kept comparable.

Similar effects were not readily detectable at the single-neuron level, consistent with weak but distributed prediction-related signals across neural populations. Together, these findings suggest that predictive processes in mouse V1 drive distributed, prediction-specific population activity.

Significance statement

Predictive coding theory proposes that the brain anticipates sensory input, and unexpected stimuli often evoke enhanced cortical responses. Yet such enhancement can also arise from adaptation, novelty, saliency, or attention, making it an ambiguous signature of prediction. By presenting mice with structured image sequences in which identical unexpected events occurred under different predictions with matched sensory history, rarity, and the magnitude of expectation violation, we dissociated prediction content from these confounds. Although mismatch-related enhancement was heterogeneous and not uniquely diagnostic of prediction, population activity in primary visual cortex distinguished what had been predicted. This prediction-content signal was weak but reliable, distributed across many neurons, suggesting that predictions shape sensory representations through subtle, population-level changes.

Introduction

Perception requires the brain to infer the causes of noisy, incomplete, and ambiguous sensory inputs. Normative frameworks propose that the brain combines internal models with incoming sensory evidence to generate stable interpretations of the external world (Aitchison & Lengyel, 2017; Friston, 2005; Heeger, 2017). Hierarchical predictive proposes that cortical circuits perform inference through iterative interactions between top-down predictions and bottom-up sensory signals, where mismatches between expected and observed input are used to update latent-state estimates across cortical hierarchies (Rao & Ballard, 1999; Friston & Kiebel, 2009; Millidge et al., 2022; Walsh et al., 2020). Although predictive coding was initially developed to explain spatial and contextual effects in perception, it naturally extends to the temporal domain, where higher-order cortical representations operating over longer timescales may encode latent contextual variables and shape sensory responses beyond immediate stimulus transitions (Rao & Ballard, 1999; Murray et al., 2014).

A growing body of work has examined how temporal regularities shape sensory representations. Because predictive coding proposes that sensory processing depends on interactions between hierarchical brain regions (Rao & Ballard, 1999), with higher-order areas integrating information over longer timescales than early sensory cortex (Murray et al., 2014), experimental studies often probe prediction across systems or extended temporal structure. These include multisensory and sensorimotor contingencies (Audette et al., 2022; Fiser et al., 2016; Furutachi et al., 2024; Garner & Keller, 2022; Keller et al., 2012), structured sensory sequences (Finnie et al., 2021; Gavornik & Bear, 2014; Gillon et al., 2024; Homann et al., 2022; Price et al., 2023; Wyrick et al., 2023), and longer-timescale stimulus statistics, including oddball paradigms (Bastos et al., 2023; den Ouden et al., 2012; Hamm et al., 2021). Consistent with hierarchical predictive coding formulations, many expectation-related effects in such paradigms are modulated by interactions between higher-order and sensory areas, including distributed corticocortical and higher-order thalamocortical pathways (Audette et al., 2022; Bastos et al.,

2023; Fiser et al., 2016; Furutachi et al., 2024; Garner & Keller, 2022). In these paradigms, enhanced responses to unexpected sensory events are frequently interpreted as prediction-error signals, which are expected to scale with the degree of mismatch between predicted and observed input (Rao & Ballard, 1999).

However, mismatch enhanced responses that scale with expectation violation are not unique to predictive-coding. Similar expectation-related response modulation can arise from multiple interacting processes, including saliency-driven enhancement of unexpected events, attentional orienting, novelty processing, alerting, and other forms of behavioral relevance that influence sensory responses through local recurrent dynamics or long-range top-down interactions (Aitchison & Lengyel, 2017; Clark, 2016; Weber & Fairhall, 2019). Consequently, enhanced responses to violated expectations may reflect sensitivity to surprise or regularity violation rather than neural representations that depend on the specific predicted content. Moreover, mismatch-related processes are not mutually exclusive and may interact nonlinearly, such that different mechanisms can amplify, suppress, or mask one another. For example, local sensory history and adaptation can in some cases enhance responses to violated regularities, while in other contexts partially mask the effects of expectation-related signals (Homann et al., 2022; Peterka et al., 2026; Weber & Fairhall, 2019). The central challenge is therefore not determining whether expectation-related signals exist, but isolating prediction-content-specific effects from other sources of mismatch-related modulation while minimizing interactions that may mask or distort them.

To address this challenge, we used a visual sequence paradigm designed to manipulate prediction content while controlling sensory input, recent stimulus history, and broader expectation-related confounds. Mice were presented with structured image sequences containing omissions and stimulus substitutions, allowing identical unexpected sensory events to occur under different prediction contexts while maintaining matched recent sensory history. Under such conditions, mechanisms primarily sensitive to surprise magnitude, novelty, saliency,

or nonspecific deviance should not systematically distinguish between prediction contexts, whereas mechanisms generating content-specific neural representations based on latent-state estimates should still produce differentiable changes in sensory activity. Using large-scale Neuropixels recordings across cortical layers of the mouse primary visual cortex, we found that unexpected events elicited enhanced neural responses consistent with expectation-related modulation. However, changes in response magnitude alone did not distinguish between general expectation-related processes and mechanisms generating content-specific neural representations based on predicted stimulus identity. We therefore next tested whether identical sensory inputs became distinguishable under different prediction contexts while controlling sensory input, recent stimulus history, and broader task-related contextual variables, and how prediction organizes the geometry of sensory population representations. We further asked how these representational changes are implemented, whether they reuse the sensory representation of the predicted stimulus, and whether they emerge at the level of individual neurons or distributed neural populations. Together, these findings suggest that sensory responses in V1 reflect specific expectations about upcoming stimuli rather than only nonspecific enhancement associated with expectation violation, and characterize how prediction shapes sensory population representations.

Results

Visual sequence paradigm for probing prediction-dependent representations

To probe how prediction content shapes early sensory representations, we designed a visual sequence paradigm in which head-fixed mice viewed repeated presentations of a fixed image sequence while neural activity was recorded from the primary visual cortex (Fig. 1a) using 4-shank Neuropixels 2.0 electrode arrays (Steinmetz et al., 2021). We refer to the standard sequence as ABBCC where each letter denotes a unique full-field natural image. Repeated exposure to the fixed sequence allowed animals to encounter a stable sequential

structure from which predictions could emerge. Recording sessions additionally included probe and control conditions designed to test how prediction-dependent representations interact with sensory input and recent stimulus history, while dissociating prediction-related effects from confounds such as stimulus order.

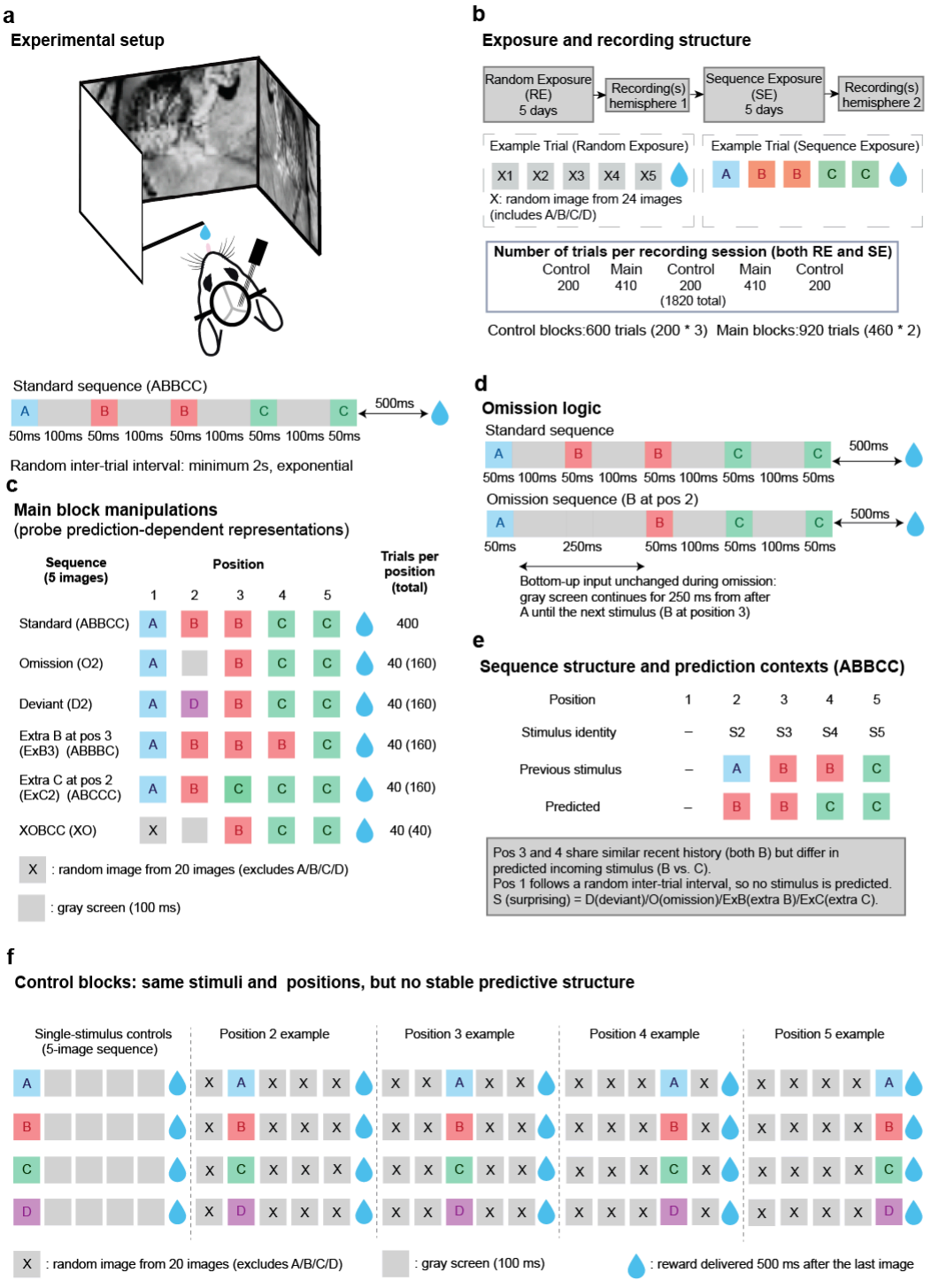


Figure 1: Experimental design for probing prediction-dependent representations in mouse visual cortex.
(a) Experimental setup. Head-fixed mice viewed fullfield natural images repeated across three LCD monitors surrounding the animal while neural activity was recorded from the primary visual cortex using 4-shank Neuropixels 2.0 probes. Animals observed many repetitions of a standard visual sequence ABBCC during recordings where each letter denotes a specific image. Other probe trials were presented as well to study the effect of prediction content on representation. Images were presented for 50 ms followed by a 100 ms gray screen. Inter-trial intervals were drawn

from an exponential distribution with a minimum of 2 s to maintain an approximately flat hazard rate. 10% sucrose water reward was delivered 0.5 s after the final image. **(b)** Exposure and recording structure. Mice first underwent Random Exposure (RE) for 5 days, during which they viewed randomly ordered sequences of five natural images followed by reward delivery. Recording sessions (0–2 sessions) were then performed, followed by 5 days of Sequence Exposure (SE) to the fixed sequence ABBCC and additional recording sessions from the opposite hemisphere. Example trials for RE and SE are shown. During recording sessions, control and main blocks were interleaved (Control–Main–Control–Main–Control). Control blocks contained 200 trials each, whereas main blocks contained 460 trials each. **(c)** Main block manipulations used to probe prediction-dependent representations. Standard ABBCC sequences were interleaved with surprising (S) manipulations including omissions (O), deviant substitutions (D), and substitution with additional familiar stimuli (ExB, ExC). In the XOBCC condition, the initial sequence context was replaced with a random image (X) while the second element in sequence was omitted. Standard ABBCC trials occurred 400 times per session, whereas each manipulation occurred 40 times per position unless otherwise indicated. **(d)** Timing and omission logic. In standard trials, images were separated by gray-screen intervals. During omission trials, the expected stimulus was replaced by continued gray screen, such that bottom-up visual input remained unchanged from the offset of the stimulus preceding the omitted image until presentation of the next expected stimulus. **(e)** Sequence structure and prediction contexts. Sequence positions were indexed relative to A, which was treated as position 1. Positions 3 and 4 shared similar recent sensory history because both followed stimulus B, but differed in the predicted upcoming stimulus (B vs. C). Therefore, surprising stimuli at these positions were used to partially dissociate prediction context from recent sensory history. Letters denote bottom-up stimulus identity, whereas numbers denote sequence position and prediction context. **(f)** Control block structure. Control blocks preserved stimulus identities and sequence positions while disrupting stable higher-order predictive structure. Stimuli were presented either alone or embedded at different positions within otherwise random image sequences composed of random natural images (X). These controls tested whether stimulus order or recent sensory history alone could generate discriminable population states in the absence of stable sequence predictions.

Stimulus content and timing were chosen to facilitate probing how prediction content shapes sensory representations. Full-field natural images were used because they contain richer combinations of visual features than simple parametric stimuli, evoking distributed population responses well suited for representational analyses. Images were presented for 50 ms and separated by a 100 ms gray-screen interval (Fig. 1a). This timing was designed to preserve transient sensory dynamics while reducing overlap between offset and onset responses of successive sequence elements. Moreover, because visual adaptation can persist across intervening stimuli even following brief stimulus presentations (250 ms), we used short image durations separated by gray intervals to reduce long-timescale stimulus-history effects (Fritsche et al., 2022).

To minimize motor and cognitive signals while maintaining engagement, mice viewed sequences under passive conditions and received reward independent of sequence content. Delivery of 10% sucrose water reward occurred 0.5 s after the final image in each sequence (Fig. 1a). Inter-trial intervals were drawn from an exponential distribution with a minimum of 2 s,

allowing reward consumption between trials and separating successive sequences. Anticipatory licking additionally provided a behavioral readout of engagement and reward expectation.

Because prediction-related effects can emerge relatively rapidly within recording sessions (Hamm et al., 2021; Homann et al., 2022), we additionally examined whether prolonged exposure to the fixed sequence was necessary for altered neural responses (Fig. 1b). Mice first underwent 5 days of Random Exposure (RE), during which they viewed 1,000 trials per day consisting of randomly ordered five-image sequences drawn from a set of 24 natural images, each followed by reward delivery. The same animals subsequently underwent 5 days of Sequence Exposure (SE) to the fixed ABBCC sequence. Recording sessions in the primary visual cortex were then performed following RE and SE using identical recording protocols, consisting of standard ABBCC trials interleaved with sequence perturbations and randomized control blocks. For brevity, we refer to recording sessions performed after Random Exposure as RE sessions and those performed after Sequence Exposure as SE sessions. Thus, animals encountered the ABBCC sequence for the first time during RE sessions, allowing comparison of neural responses to the same sequence structure before and after prolonged exposure. Prior RE familiarized animals with the general task structure, reward timing, and stimulus set, reducing the likelihood that responses observed during RE sessions primarily reflected novelty or acquisition of reward associations rather than expectation formed based on sequential relationships among stimuli. In a subset of cases, recordings following RE were not obtained due to technical issues, and these animals contributed only SE data.

Recording sessions were organized into alternating main and control blocks (control–main–control–main–control), interleaved throughout the experiment to minimize confounds arising from slow changes in behavioral state, recording stability, or time-dependent adaptation (Fig. 1b) Main blocks consisted of repeated presentations of the fixed ABBCC sequence interleaved with perturbation trials designed to probe prediction-dependent representations. Specific perturbation and control conditions are described below.

In the main block, prediction-dependent representational structure was probed through three primary perturbations of the fixed ABBCC sequence (Fig. 1c). Expected sequence elements were either (1) omitted (O), (2) replaced with a deviant image (D), or (3) substituted with an extra B or C image (ExB or ExC). To test whether expected but absent sensory events evoked prediction-related activity in the absence of bottom-up sensory input, omission trials replaced the expected image with continued gray screen (O). To test whether responses to unexpected sensory events were modulated by prediction context, deviant trials replaced the expected element with a natural image (D) selected from the set of 24 images encountered during RE. Unlike sequence reordering manipulations, ExB and ExC introduced an additional instance of a B or C element by replacing another expected sequence element while preserving overall sequence length. Because deviant images violated both global expectations (e.g., rarity or frequency of events) and local content-specific sequence expectations, ExB and ExC trials additionally tested whether prediction-dependent changes generalized to violations of learned sequence expectations alone while more closely matching stimulus familiarity and salience. To test whether prediction-dependent activity depended on preservation of the broader learned sequence structure, an additional manipulation, XOBCC, replaced the initial sequence stimulus with a random image (X) drawn from the pool of 20 non-ABCD images encountered during RE, followed by omission of the expected B stimulus (O). This manipulation allowed us to test whether responses to omitted or unexpected events depended on the broader learned sequence context rather than only local stimulus transitions immediately preceding the perturbation.

Sequence timing was designed to facilitate interpretation of omission-related activity at the expected stimulus time. Each image was followed by a 100 ms gray-screen interval, so by the time an expected stimulus was omitted, visual input was already gray (Fig 1d). Thus, omission trials replaced the expected image onset with continued gray screen rather than introducing a new image-to-gray transition, reducing the likelihood that activity at omission time

simply reflected an immediate offset response to the preceding stimulus. However, offset responses can also exhibit delayed rebound dynamics that occur near the time of the subsequent expected stimulus. Pilot experiments using slower presentation rates revealed rebound activity aligned to expected stimulus timing, motivating the faster cadence adopted here. Prior characterization of post-stimulus dynamics in the same dataset further demonstrated rebound activity occurring approximately 250 ms following stimulus offset (Deveau et al., 2026). By shortening the interval between successive sequence elements, the faster cadence reduced temporal overlap between delayed rebound peaks and the expected stimulus time, allowing omission-related activity to be more readily dissociated from rebound responses to preceding sensory events.

To facilitate comparison of manipulations in terms of sensory input, sensory history, sequence position, and prediction content, we use a notation that indexes each condition according to both stimulus identity and sequence position. Letters denote bottom-up sensory identity, whereas numbers denote position within the sequence. The label S denotes surprising events, including omissions (O), deviant substitutions (D), and extra B or C substitutions (ExB, ExC). Sequence positions are indexed relative to the first sequence element, which is treated as position 1 and contains either stimulus A or, in XOBCC trials, a random image X. For example, B2 and B3 denote the same sensory input (B) occurring as the second and third items of the standard ABBCC sequence, whereas C4 and C5 denote the fourth and fifth sequence elements. ExC manipulations are denoted C2 and C3 because C replaces B2 or B3, respectively, whereas ExB manipulations are denoted B4 and B5 because B replaces C4 or C5. Similarly, D2–D5 denote the same deviant image presented in place of the expected B2–B3 or C4–C5 stimulus, whereas O2–O5 denote omission of the expected B2–B3 or C4–C5 stimulus.

The ABBCC sequence structure was designed to partially dissociate prediction context from recent sensory history and adaptation (Fig. 1e). Items in positions 3 and 4 shared substantial overlap in recent sensory history because both followed stimulus B but differed in the

predicted upcoming stimulus (B vs. C). For example, D3 and D4 involved the same bottom-up sensory input presented in different learned sequence contexts while sharing the same immediately preceding stimulus, whereas O3 and O4 probed responses to omission of different expected stimuli at those sequence positions. More generally, this design enabled tests of whether population representations became organized according to prediction context across distinct sensory inputs and recent sensory histories, including omissions and deviant stimulus.

Control blocks were included to test whether sensory adaptation and stimulus order alone could account for prediction-related effects observed in the main sequence blocks (Fig. 1f). In these blocks, stimuli were presented either alone or embedded at different sequence positions within otherwise random image sequences (e.g., XAXXX, XXAXX, XBXXX), where X was drawn from the same pool of 20 non-ABCD images used in XOBCC trials. By disrupting stable higher-order sequence structure while preserving many lower-order sensory and temporal features of the task, these conditions provided matched comparisons for assessing prediction-dependent modulation of sensory responses and testing whether stimulus order or recent sensory history alone could generate discriminable population states. Control blocks additionally provided an empirical basis for characterizing sensory- and adaptation-related population response structure independent of stable sequence predictions. As described above, control and main blocks were interleaved throughout recording sessions to minimize systematic differences arising from slow changes in behavioral state, recording stability, or time-dependent adaptation.

Sequence-locked V1 activity with matched behavior and arousal across conditions

We recorded extracellularly from primary visual cortex (V1) while animals viewed the repeated ABBCC image sequence interleaved with surprising trials (omissions and

substitutions) and randomized control blocks. Across 13 recording sessions from 6 mice, we isolated 1,779 single units and 2,485 multi-units spanning cortical layers L2/3, L4, and L5/6 (Table S1). A representative 50 s excerpt from one session (Fig. 2a) shows the layer-resolved spiking together with the simultaneously acquired pupil trace, trial events, and licking behavior. Spike-depth raster maps across recording sessions were used to assess recording stability and informed session trimming when necessary (Fig. S1).

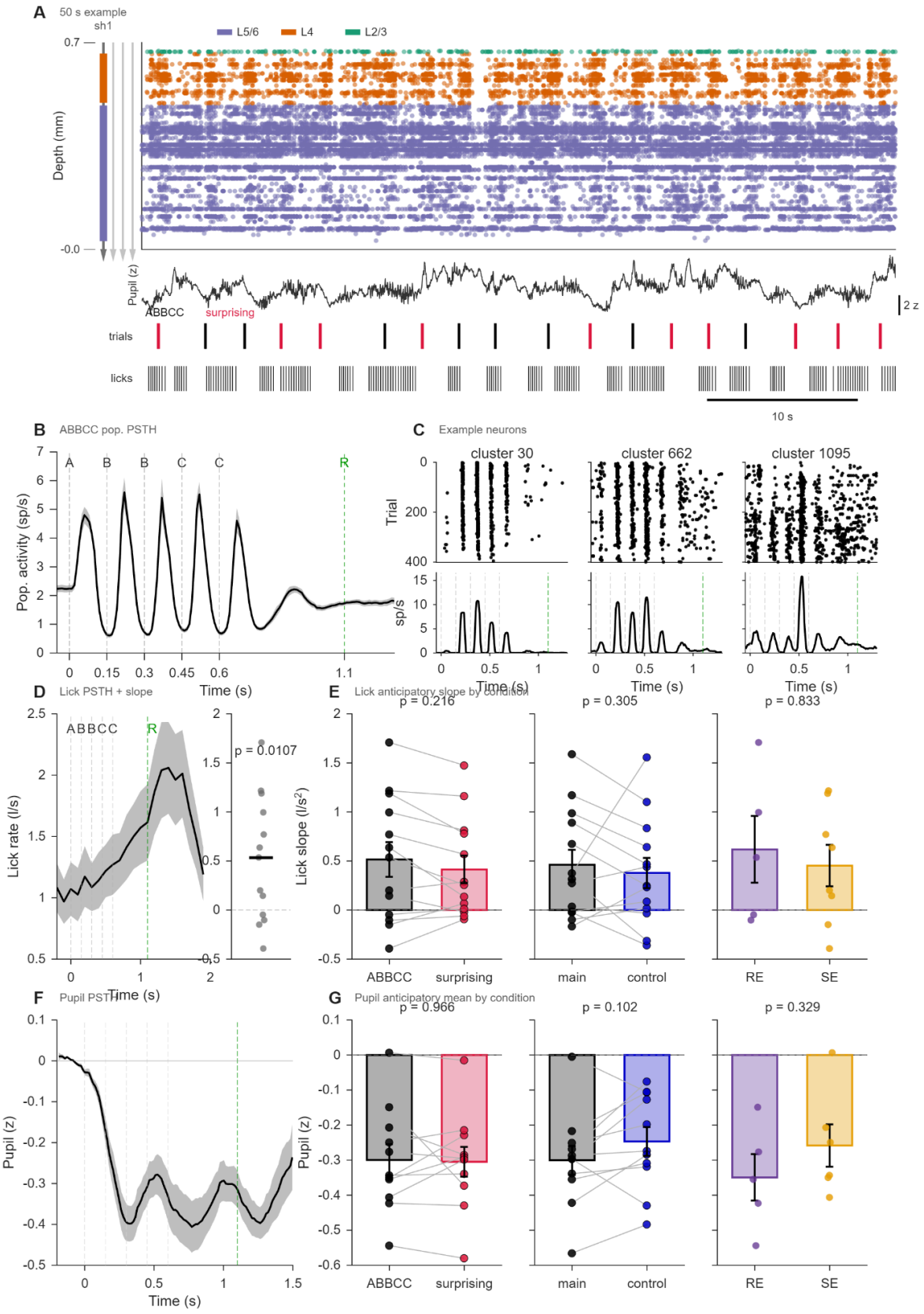

A
50 s example
sh1
L5/6
L4
L2/3
0.7
-0.0
Depth (mm)
Pupil (z)
2 z
ABBCC
surprising
trials
licks
10 s
B
ABBCC pop. PSTH
Pop. activity (sp/s)
Time (s)
C
Example neurons
cluster 30
cluster 662
cluster 1095
Trial
sp/s
Time (s)
D
Lick PSTH + slope
Lick rate (l/s)
p = 0.0107
Lick slope (l/s²)
E
Lick anticipatory slope by condition
p = 0.216
p = 0.305
p = 0.833
ABBCC
surprising
main
control
RE
SE
F
Pupil PSTH
Pupil (z)
Time (s)
G
Pupil anticipatory mean by condition
p = 0.966
p = 0.102
p = 0.329

Figure 2: Sequence-locked V1 population activity with matched behavior and arousal across conditions.
(a) Example 50 s excerpt from one recording session showing layer-resolved spike raster (colored by cortical layer: L2/3, L4, L5/6), z-scored pupil area, trial onsets (ABBCC, black; surprising trials, red), and lick raster. Spike-depth raster maps used to assess recording stability across sessions are shown in Fig. S1.**(b)** Population PSTH for ABBCC sequences, averaged across units within a session and then across sessions (mean ± SEM). Letters indicate stimulus identities and vertical dashed lines indicate stimulus onsets; **R** indicates reward delivery. **(c)** Example single-unit spike rasters (top) and PSTHs (bottom) for representative neurons showing sequence-locked responses to different elements of the ABBCC sequence. **(d)** Lick PSTH for ABBCC trials (mean ± SEM across sessions). Inset: per-session linear slope of lick rate from the first stimulus to reward delivery, used as a summary measure of anticipatory licking and tested against zero (Wilcoxon signed-rank, one-sided). **(e)** Per-session anticipatory lick slopes by condition: ABBCC versus surprising trials and main versus control blocks (paired, Wilcoxon signed-rank), and Random Exposure (RE) versus Sequence Exposure (SE) sessions (unpaired, Mann–Whitney U test). Bars indicate mean ± SEM across sessions; dots indicate individual sessions; connecting lines indicate paired comparisons. **(f)** Pupil PSTH across sessions (mean ± SEM), z-scored relative to the pretrial baseline period (−0.2 to 0 s). **(g)** Per-session anticipatory pupil responses by condition, plotted as in (e). Pupil responses did not differ across sequence conditions despite a small common constriction following sequence onset. Neural panels, n = 13 sessions. Behavioral panels (d, e), n = 12 sessions (SM_0009 excluded, no behavioral video). Pupil panels (f, g) include sessions with usable eye tracking (n = 11).

At the population level, V1 activity was strongly entrained to the sequence. The mean firing rate, averaged across units within each session and then across sessions, showed a clear response to each of the five sequence elements, with some adaptation visible as reduced responses to the repeated B and C elements compared to first B and C (Fig. 2b). Individual neurons were heterogeneous but reliably sequence-locked: example units showed sharp, stimulus-aligned responses tiling the sequence (Fig. 2c), consistent with the population PSTH.

Licking behavior indicated that animals generally learned the task's reward structure and engaged with the stimuli. Mice tended to increase licking across the sequence in anticipation of reward (Fig. 2d). To quantify anticipatory licking, we fit a linear slope to lick rate from the first stimulus to reward delivery within each session, providing a simple summary measure of progressive change in licking across the sequence and allowing a direct test of whether anticipatory licking increased over time. Session-wise slopes were significantly greater than zero (signed-rank test, one-sided, p = 0.0046, n = 12 sessions), indicating overall anticipatory ramping (Fig. 2d; Fig. S1). Although several sessions showed weaker or minimal anticipatory licking (e.g., SM12 d1, SM13 d1), the majority of sessions exhibited positive ramping, and this pattern was observed across both Random Exposure (RE) and Sequence Exposure (SE) conditions (Fig. S1). Anticipatory licking also declined gradually over the course of each session, consistent with satiation: per-session slopes of anticipatory lick rate against trial

number were reliably negative (median = −0.083 licks/s per trial; signed-rank test, $p = 5 \times 10^{-4}$, $n = 12$; Fig. S2a–b).

Licking depended primarily on trial timing rather than sequence content or condition. Anticipatory lick slopes did not differ significantly between familiar ABBCC trials and surprising trials containing omissions or substitutions (signed-rank test, $p = 0.077$), between main and control blocks (signed-rank test, $p = 0.18$), or between Random Exposure (RE) and Sequence Exposure (SE) sessions (rank-sum test, $p = 1.0$; Fig. 2e). The ABBCC-versus-surprising comparison showed a small nonsignificant trend ($\Delta = 0.135$ licks/s²), and lick dynamics were visually similar across individual surprising trial types (Fig. S2).

Pupil size likewise showed no trial-aligned modulation that distinguished conditions despite a small common constriction following sequence onset. Anticipatory pupil area (z-scored relative to the pretrial baseline period) did not differ between ABBCC and surprising trials (signed-rank test, $p = 0.97$), between main and control blocks (signed-rank test, $p = 0.10$), or between RE and SE sessions (rank-sum test, $p = 0.33$; Fig. 2g). Pupil time courses were similar across individual trial types (Fig. S2), and across sessions we observed a small, consistent constriction following sequence onset that did not vary by condition (Fig. 2f).

Together, these results indicate that differences in V1 activity between familiar and modified sequences, between control and main blocks, or between RE and SE conditions are unlikely to be attributable to differences in overt behavior or arousal as indexed by pupil.

Responses at Times of Omitted Stimuli Reflect Rebound Dynamics with an Additional Experience and Expectation-Related Component.

To test whether omitted stimuli elicited responses beyond nonspecific post-stimulus dynamics, we compared population-averaged activity following omission of an expected sequence element with activity following the final stimulus of the sequence, when no further

stimulus was expected. Population activity following omission (Om) exhibited a transient increase followed by a decrease (Fig. 3a). Similar dynamics, however, also followed the final sequence element at a time when no stimulus appearance was expected (Fig. 3a,b). Such rebound has previously been reported in the mouse primary visual cortex following stimulus offset (Tang et al., 2023; Zhang et al., 2016).

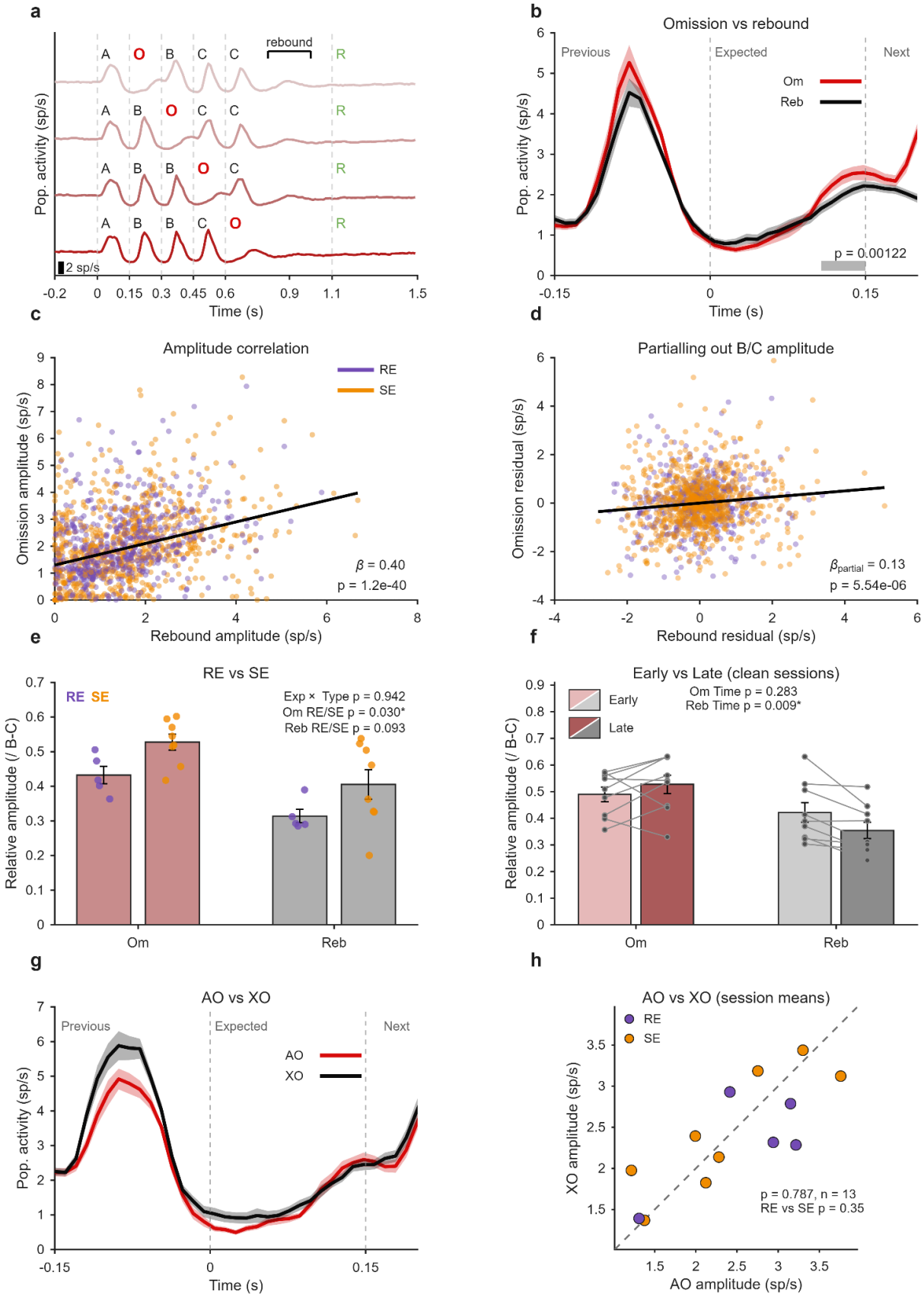

a
A O B C C rebound R
A B O C C R
A B B O C R
A B B C O R
2 sp/s
Pop. activity (sp/s)
-0.2 0 0.15 0.3 0.45 0.6 0.9 1.1 1.5
Time (s)
b
Omission vs rebound
Previous
Expected
Next
Om
Reb
p = 0.00122
Pop. activity (sp/s)
-0.15 0 0.15
Time (s)
c
Amplitude correlation
RE
SE
β = 0.40
p = 1.2e-40
Omission amplitude (sp/s)
Rebound amplitude (sp/s)
d
Partialling out B/C amplitude
β_partial = 0.13
p = 5.54e-06
Omission residual (sp/s)
Rebound residual (sp/s)
e
RE vs SE
RE SE
Exp × Type p = 0.942
Om RE/SE p = 0.030*
Reb RE/SE p = 0.093
Relative amplitude (/ B-C)
Om
Reb
f
Early vs Late (clean sessions)
Early
Late
Om Time p = 0.283
Reb Time p = 0.009*
Relative amplitude (/ B-C)
Om
Reb
g
AO vs XO
Previous
Expected
Next
AO
XO
Pop. activity (sp/s)
-0.15 0 0.15
Time (s)
h
AO vs XO (session means)
RE
SE
p = 0.787, n = 13
RE vs SE p = 0.35
XO amplitude (sp/s)
AO amplitude (sp/s)

Figure 3: Responses at omission times reflect rebound dynamics with additional experience and expectation-dependent components.
(a) Population-averaged activity for omission trials at each position within the ABBCC sequence (AOBCC, ABOCC, ABBOC, ABBCO). Vertical dashed lines indicate stimulus onsets; the green "R" indicates reward delivery. Responses are aligned to sequence onset. The bracket marks the rebound period following the final element. Scale bar, 2 sp/s. **(b)** Population-averaged omission (Om, red) and rebound (Osc, black) responses aligned to the onset of the stimulus preceding omission (or sequence end for Osc; 0 s). Osc was computed from the same trials as O1–O3, excluding O4, which lacked a matched end-of-trial rebound period. Shaded regions, s.e.m. across sessions. The gray bar marks the significant cluster (cluster-based permutation test); the signed-rank p across sessions is shown. **(c)** Trial-by-trial omission versus rebound amplitude in the same trials. Each point is one trial; color denotes exposure (RE, purple; SE, gold). The black line is the marginal fit of a linear mixed-effects model with a session random intercept. β and p are shown. **(d)** As in (c), after partialling out the within-trial B/C amplitude from both omission and rebound (residual versus residual). The omission–rebound relationship persists. β_partial and p are shown. **(e)** Relative omission and rebound amplitudes (normalized to B/C responses in the same trials), shown for RE and SE sessions. Dots, individual sessions; bars and error bars, mean ± s.e.m. Omission amplitude was higher in SE than RE (ranksum p = 0.030); rebound did not differ by exposure (p = 0.093), however, no interaction effect was present (Exposure*Type p = 0.942). **(f)** Relative omission and rebound amplitudes for early versus late trial halves (sessions with intact trial counts; see Methods). Lines connect halves within a session. Bars and error bars, mean ± s.e.m. Omission was stable across the session (Time p = 0.283); rebound declined from early to late (Time p = 0.009). **(g)** Population-averaged omission responses following a familiar element (AO, after A; red) and following a random image (XO, after X; black), aligned to the onset of the stimulus preceding omission (0 s). Shaded regions, s.e.m. across sessions. No significant time-resolved differences (cluster-based permutation test). **(h)** Session-level AO versus XO omission amplitudes. Dots, individual sessions (RE, purple; SE, gold); dashed line, unity. AO and XO were similar across sessions (signed-rank p = 0.787, n = 13), with no exposure difference in the AO–XO contrast (ranksum p = 0.35).

To characterize the relationship between omission and rebound responses, we compared their amplitudes and tested whether they covaried trial by trial (O1–O3 only, as O4 lacked a matched rebound period; Fig. 3b–d). Although the two responses shared similar overall dynamics, omission responses were significantly larger than rebound (Fig. 3b; signed-rank p = 0.0012; the difference formed a significant cluster spanning 0.054–0.150 s, cluster-based permutation test, see Methods), i.e. during the rising phase. Across trials, omission amplitude was positively correlated with rebound amplitude measured in the same trials (linear mixed-effects model with a session random intercept, $\beta = 0.40$, $p = 1.2 \times 10^{-40}$; Fig. 3c). This difference was not a consequence of the omission occurring earlier in the sequence than the rebound: across the main-block omission positions (2–5), relative omission amplitude did not vary systematically with position (per-session slope, signed-rank p = 0.64), indicating that the larger omission response reflects a difference in kind from the rebound rather than a gradual decline in responsiveness across the trial. To test whether this coupling simply reflected shared trial-level components in the visually evoked B/C responses — an index of how excitable the population was during the trial — we asked whether omission and rebound amplitudes

remained related after accounting for the within-trial B/C response in both measures. Specifically, we first regressed omission and rebound amplitudes against B/C amplitude and then tested the relationship between the resulting residuals. re-estimated the relationship after partialling out the within-trial B/C amplitude from both measures. The omission–rebound relationship persisted, though it weakened (partial $\beta = 0.13$, down from $\beta = 0.40$; $p = 5.5 \times 10^{-6}$; Fig. 3d), indicating that omission responses share a component with cortical rebound dynamics that is not fully explained by trial-to-trial fluctuations in population excitability. This coupling was reproducible at the session level (per-session regression of omission on rebound amplitude: median $\beta = 0.10$, signed-rank $p = 9.8 \times 10^{-4}$, $n = 11$) and was unchanged when pre-stimulus pupil diameter and lick rate were included as covariates (median $\beta = 0.11$), indicating that it does not reflect shared modulation by arousal or engagement. Because the visually evoked B/C response thus indexes trial-to-trial and session-to-session changes in population excitability, we expressed omission and rebound amplitudes relative to the within-trial B/C response for the across-session comparisons that follow, which are otherwise confounded by differences in overall excitability between sessions.

Because expectation-related responses can potentially emerge rapidly within a session, independent of prolonged exposure to the specific sequence, we asked whether the amplitude of omission responses additionally carried a sequence-exposure-dependent component. We compared relative omission amplitude between Random Exposure (RE) and Sequence Exposure (SE) sessions (Fig. 3e). Relative omission amplitude was higher during SE than RE sessions (ranksum $p = 0.030$), whereas relative rebound amplitude did not differ significantly between exposure conditions (ranksum $p = 0.093$). However, the two responses were not differentially affected by exposure (Exposure × response-type interaction $p = 0.94$), so we treat this as a modest general exposure effect rather than an omission-specific one. This exposure effect on the relative measure was not a consequence of a change in the B/C normalizer, which did not differ between RE and SE sessions (ranksum $p = 0.28$). Although pooled omission

amplitude showed no systematic positional gradient (above), splitting by exposure revealed position-specific structure. Increase of omission response by exposure was not uniform across the sequence: at position 2 or 3, relative omission amplitude did not differ between RE and SE (ranksum $p = 0.72$ and 0.52, respectively), whereas it increased from RE to SE at positions 4 and 5 (ranksum $p = 0.019$ each). Under random exposure, relative omission amplitude showed a modest but nonsignificant decline across position (positions 2–5) that was absent after sequence exposure (per-session slope, signed-rank $p = 0.125$, $n=5$; Fig. S3a). However, a formal Position × Exposure interaction was not supported by a session-level slope contrast (ranksum $p = 0.22$). Thus the exposure-dependent enhancement is not a uniform rescaling but is concentrated at the later positions: under random exposure the relative omission response was largest at the earliest omission position and declined across the trial, whereas after sequence exposure it remained elevated through the final positions, so that SE exceeded RE specifically at positions 4–5. This is consistent with an omission response scaling with the expectation of an upcoming stimulus — strongest early, when more elements remain — whose maintenance across later positions continues to consolidate with exposure rather than saturating during random exposure.

We next asked whether omission and rebound amplitudes evolved differently within a session, both to characterize their within-session dynamics and to test whether the two responses are dissociable (Fig. 3f). Within sessions, relative omission amplitude did not change significantly from early to late trials (mixed ANOVA, main effect of Time $F(1,11) = 1.35$, $p = 0.283$; Exposure × Time interaction $F(1,11) = 1.07$, $p = 0.334$), indicating stability across the recording session. In contrast, relative rebound amplitude declined from early to late trials (main effect of Time $F(1,11) = 12.87$, $p = 0.009$; no significant interaction with exposure type, $F(1,11) = 3.16$, $p = 0.119$), an effect that was robust to defining early and late as the first versus last quartile of trials (paired signed-rank $p = 0.027$). Thus omission and rebound responses, while

sharing common dynamics and amplitude covariation, behaved differently within the session: omission amplitude was stable while rebound amplitude declined.

This within-session dissociation was not an artifact of the relative-amplitude normalization or of a shift in behavioral state across the session. The visually evoked B/C responses used for normalization were themselves stable across the session (mixed ANOVA on raw B/C amplitude: no effect of early-versus-late half, Time $F(1,11) = 0.47$, $p = 0.51$; no exposure effect, $F(1,11) = 0.63$, $p = 0.44$; no interaction, $F(1,11) = 0.09$, $p = 0.77$), so the divergence between omission and rebound is not a consequence of drift in the normalizer. Pre-stimulus pupil diameter, an index of arousal, did not differ between the early and late halves used in panel f (signed-rank $p = 0.83$, $n = 11$). Lick rate declined from early to late trials (signed-rank $p = 9.8 \times 10^{-4}$), consistent with satiation, but lick rate was not correlated with omission or rebound amplitude on a trial-by-trial basis (omission–lick and rebound–lick both $p > 0.5$; pupil correlations likewise non-significant, omission–pupil median $r = -0.13$, $p = 0.21$; rebound–pupil median $r = -0.05$, $p = 0.17$; $n = 11$ sessions). Because a variable that changes across the session but is unrelated to the response is not likely to account for the response's within-session change, neither measure can explain the divergence between omission and rebound. Confirming this directly, when omission and rebound amplitudes were residualized on pupil and lick within each session and the panel-f early/late contrast was recomputed on the residuals (same session set as panel f, $n = 8$), the dissociation was preserved: residual rebound amplitude still differed between early and late trials (signed-rank $p = 0.016$) whereas residual omission amplitude did not ($p = 0.46$). These findings argue against a simple account in which within-session changes in rebound are driven primarily by differences in arousal or engagement across the session.

To test whether the omission response is triggered by the prediction made based on the preceding familiar cue or by the broader sequence context, we compared omission responses following stimulus A (AO; AOBCC) with those following a random image X (XO; XOBCC), with

the remainder of the sequence held constant (Fig. 3g,h). Population dynamics following AO and XO were similar (Fig. 3g), with no significant time-resolved differences (cluster-based permutation test). At the session level, AO and XO amplitudes were distributed closely around the unity line (Fig. 3h), with no significant difference (paired signed-rank $p = 0.787$, $n = 13$) and no difference in the AO–XO contrast between exposure conditions (ranksum $p = 0.35$). Thus, omission responses did not depend on whether the immediately preceding stimulus belonged to the familiar sequence. One interpretation is that the augmentation reflects a broader learned sequential context rather than a prediction initiated specifically by the immediately preceding cue: once the animal is engaged in an ongoing structured-sequence context, the absence of an expected event may be sufficient to enhance the response even when the preceding cue does not uniquely identify the familiar sequence.

To further confirm whether these omission and rebound responses reflect the learned ABBCC sequence specifically or a more general property of sequence structure, we examined the interleaved control blocks, in which the same images appeared either in isolation or embedded in random sequences. Responses with the same dynamics as the main-block omission and rebound were present in the control block: an omission-like response followed a single isolated stimulus (Fig. S3b), and a rebound followed the end of a five-image control sequence (Fig. S3c). Quantifying these responses relative to the first sequence element, the response following a single isolated stimulus was statistically indistinguishable from the main-block omission (signed-rank all $p > 0.45$) and significantly larger than the rebound ($p = 2\times10^{-4}$), whereas the rebound at the end of a five-image control sequence was indistinguishable from the main-block rebound ($p = 0.094$) and correlated with it across sessions (Spearman $r = 0.69$). The same physical event — a gap following stimuli — was therefore omission-like when it followed a single image but rebound-like when it followed a complete five-image sequence (signed-rank $p = 2\times10^{-4}$), indicating that the omission-like enhancement depends on whether an ongoing sequence is incomplete rather than on the identity of the learned sequence. Finally, this

response did not depend on sequence exposure: unlike the main-block omission, which increased from RE to SE (Fig. 3e), the control single-stimulus response was unchanged across exposure conditions (pooled rank sum $p = 0.62$), consistent with the main-block omission response at position 2, which likewise did not differ between RE and SE (Fig. 3e and above). Together these results indicate that omission and rebound responses are not specific to the learned ABBCC sequence but reflect a general expectation that a stimulus sequence should continue, with a separable, exposure-dependent component emerging specifically for the learned sequence that further augment the omission at positions 4 and 5.

Together, these analyses reveal a double dissociation between omission and rebound responses across two independent axes. Omission responses were consistently larger than rebound (Fig. 3b), and Across sessions, omission amplitude increased with sequence exposure (SE > RE) whereas rebound amplitude did not; within sessions, rebound amplitude declined from early to late trials whereas omission amplitude remained stable. Because the two responses are modulated by different variables — e.g. omission by sequence exposure, rebound by within-session time — they cannot reflect a single underlying process, even though they share similar dynamics and covary trial-to-trial. This dissociationThisNeither dissociation was not explained by the relative-amplitude normalization or by changes in arousal or engagement across the session.

The control-block analyses further indicate that the omission response is not unitary but combines a generic rebound with a separable, expectation-related enhancement. The rebound was indistinguishable between the learned sequence and random control sequences, consistent with a generic end-of-sequence offset response. The expectation-related enhancement appeared whenever a trial was interrupted before its expected complement of stimuli — including after a single isolated stimulus — and was indistinguishable in magnitude from the main-block omission, indicating that it does not require the learned ABBCC identities. This general enhancement was exposure-independent at the earliest position (Fig. 3e); what

changed with exposure was its persistence across the trial — declining toward the later positions under random exposure, but remaining elevated after sequence exposure. The omission response is therefore better described as a generic rebound with an additional, expectation-related component superimposed on it than as a single, unitary signal.

Expectation Violations Modulate Responses to Deviant but Not Extra Familiar Elements

We next investigated whether responses to unexpected stimuli were enhanced relative to expected stimuli and whether such modulation generalized across different forms of expectation violation. Interleaved with the standard ABBCC sequence, trials were presented in which one sequence element after the first image (A) was replaced by a deviant image (D) at different positions within the sequence (Fig. 1c). We first asked whether responses to D were enhanced relative to expected B/C stimuli presented at comparable, minimally adapted positions (B2 and C4), where the same stimulus identity was not presented immediately beforehand (Fig. 4a). Although population-averaged responses to D and B/C were broadly similar in magnitude, deviant stimuli exhibited a modest increase in amplitude, most evident during the rising-to-peak phase of the response (Fig. 4a). At the session level, the amplitude of D responses was modestly but significantly larger than that of B/C responses (median session difference = 0.73 spikes/s, signed-rank $p = 0.021$, $n = 13$; Fig. 4b), indicating that unexpected deviant stimuli elicited modest response enhancement in the repeating sequence context.

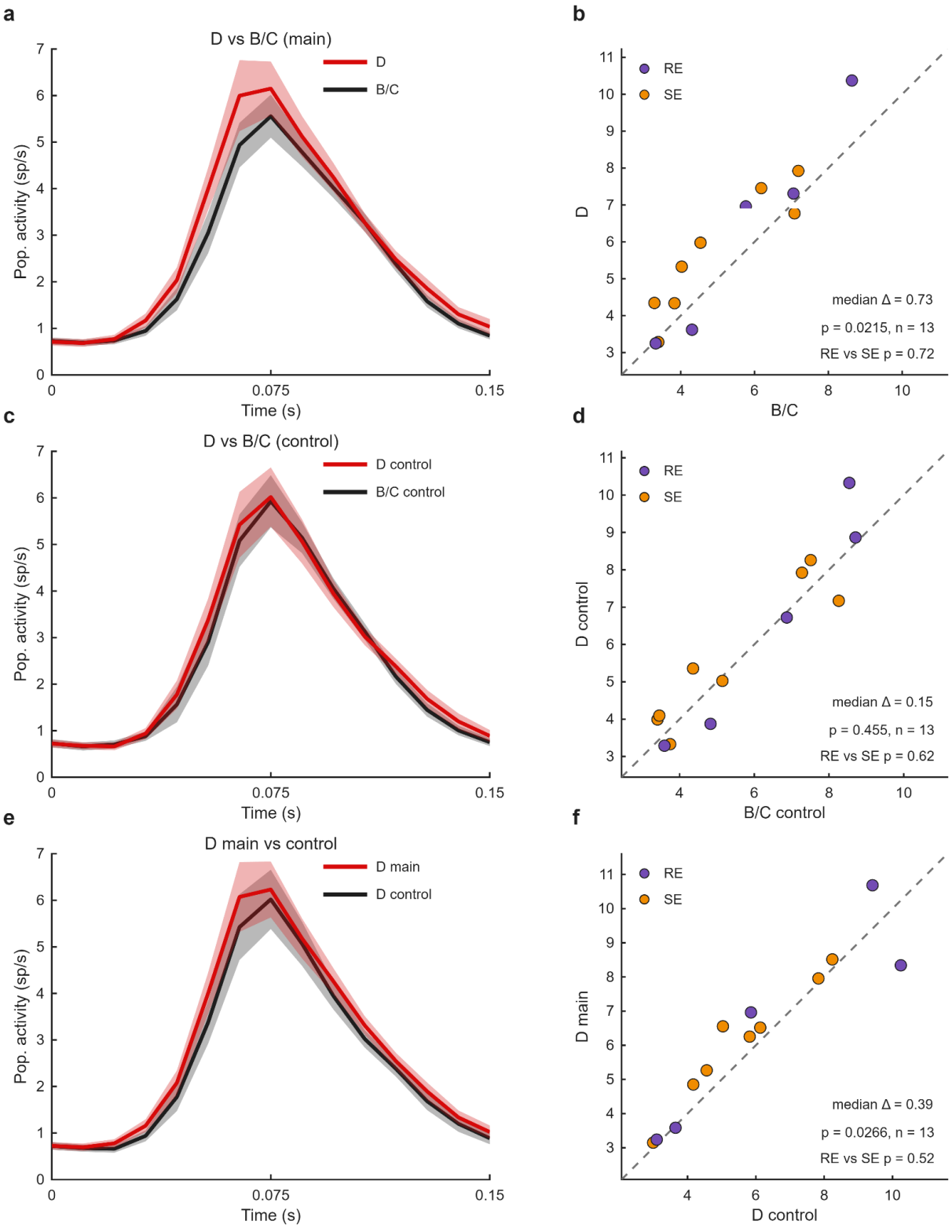

a
D vs B/C (main)
D
B/C
Pop. activity (sp/s)
Time (s)
b
RE
SE
D
B/C
median Δ = 0.73
p = 0.0215, n = 13
RE vs SE p = 0.72
c
D vs B/C (control)
D control
B/C control
Pop. activity (sp/s)
Time (s)
d
RE
SE
D control
B/C control
median Δ = 0.15
p = 0.455, n = 13
RE vs SE p = 0.62
e
D main vs control
D main
D control
Pop. activity (sp/s)
Time (s)
f
RE
SE
D main
D control
median Δ = 0.39
p = 0.0266, n = 13
RE vs SE p = 0.52

Figure 4: Deviant stimuli elicit context-dependent response enhancement specific to the learned sequence.
(a) Population-averaged activity for the deviant element D (red) and for expected B/C elements at minimally adapted positions (B1, C3; black) in the main block, aligned to stimulus onset (0 s). Shaded regions, s.e.m. across sessions. **(b)** Session-level D versus B/C amplitudes. Dots, individual sessions (RE, purple; SE, gold); dashed line, unity. D exceeded B/C across sessions (median Δ = 0.73 sp/s, signed-rank p = 0.021, n = 13), with no exposure difference (rank-sum p = 0.72). **(c)** Population-averaged D (red) and B/C (black) responses in the control block, aligned to stimulus onset. Shaded regions, s.e.m. across sessions. **(d)** Session-level D versus B/C amplitudes in the control block, formatted as in (b). D and B/C did not differ (median Δ = 0.15 sp/s, signed-rank p = 0.45, n = 13; rank-sum p = 0.62). **(e)** Population-averaged responses to D in the main block (red) and to the same stimulus at matched positions in randomized control sequences (black), aligned to stimulus onset. Shaded regions, s.e.m. across sessions. **(f)** Session-level D amplitudes, main versus control block, formatted as in (b). D was larger in the main block (median Δ = 0.39 sp/s, signed-rank p = 0.027, n = 13; rank-sum p = 0.52).

We next asked whether this enhancement reflected properties of the deviant stimulus itself or instead depended on expectation context. If enhancement of D responses relative to B/C were driven primarily by the sensory properties of D, it should also be present when the same images appeared in the randomized control context. We therefore first compared responses to D with the corresponding B/C baseline in the interleaved control block, where identical images appeared at matched positions embedded in randomized sequences (Fig. 4c,d). In this context, D responses did not differ from B/C responses at the session level (median session difference = 0.15 spikes/s, signed-rank p = 0.45, n = 13; Fig. 4d). We next directly compared responses to D presented within the standard ABBCC sequence (main block) to responses evoked by the same stimulus at matched positions in the randomized control block (Fig. 4e,f). Responses to D were larger in the main than control block (median session difference = 0.39 spikes/s, signed-rank p = 0.027, n = 13; Fig. 4f). Together, the absence of D-versus-B/C enhancement in the control block and the larger D response in the main than control block indicate that deviant enhancement depended on expectation context established by the repeating sequence rather than stimulus identity alone.

We next asked whether deviant enhancement depended on prior sequence exposure, accumulated within-session experience, or progression through the sequence itself (Fig. S4a-d). Deviant enhancement was comparable across Random Exposure (RE) and Sequence Exposure (SE) sessions (rank-sum p = 0.72; Fig. 4b), indicating that prolonged exposure to the fixed sequence was not required for the effect. Likewise, deviant enhancement remained stable

across the recording session, showing no detectable change between early and late trial halves (main effect of Time $p = 0.055$; Fig. S4a). We also asked whether deviant enhancement varied across sequence position, motivated by omission responses, which showed a trend of position-dependent changes across the trial resulting in difference with exposure. In contrast, relative D amplitude remained broadly consistent across positions following the first image, with no significant exposure-related differences at any position (all rank-sum $p > 0.19$; position 5 uncorrected $p = 0.063$; Fig. S4c). Importantly, these findings were not explained by dividing the amplitude by differences in the corresponding B/C responses measured from the same trials, which likewise remained stable across exposure condition, recording time, and sequence position (Fig. S4b and d). Together, these findings indicate that deviant enhancement exhibited a distinct profile from omission-related enhancement: whereas omission responses varied across sequence position and exposure history, deviant enhancement remained relatively stable across experience, within-session time, and sequence position consistent with distinct expectation-related mechanisms contributing to their enhancement.

Although the average enhancement was modest, its magnitude varied across recording sessions (Fig. 4b, distance to unity line). To illustrate this variability, we show representative sessions with high and intermediate effect sizes (Fig. S4e-h), both exhibiting larger responses to D within the learned sequence relative to matched B/C responses and to D responses in the control block. These examples demonstrate that the population-level effect reflected an average across individual sessions that exhibited some significant variability.

We next asked whether suppression of expected B/C responses within the repeating sequence additionally contributed to the larger D-versus-B/C difference observed in the main block. Although D responses were enhanced relative to the randomized control context (Fig. 4e,f), a complementary reduction of expected responses in the repeating sequence could further amplify the observed D–B/C difference. If expectation suppressed responses to predicted sequence elements, expected B/C responses should be smaller in the standard

ABBCC sequence than in the randomized control context. We therefore compared responses to minimally adapted B/C elements (B2 and C4) between the main and control blocks (Fig. 5a,b). Population-averaged responses were highly similar across contexts (Fig. 5a), and session-level amplitudes did not differ between the main and control blocks (median session difference = −0.01 spikes/s, signed-rank $p = 0.38$, $n = 13$; Fig. 5b). Likewise, context-related differences did not differ between Random Exposure (RE) and Sequence Exposure (SE) sessions (rank-sum $p = 1.0$; Fig. 5b). Thus, enhancement of deviant responses was not accompanied by detectable suppression of expected B/C responses in the repeating sequence, arguing against a simple prediction-error account of the observed modulation.

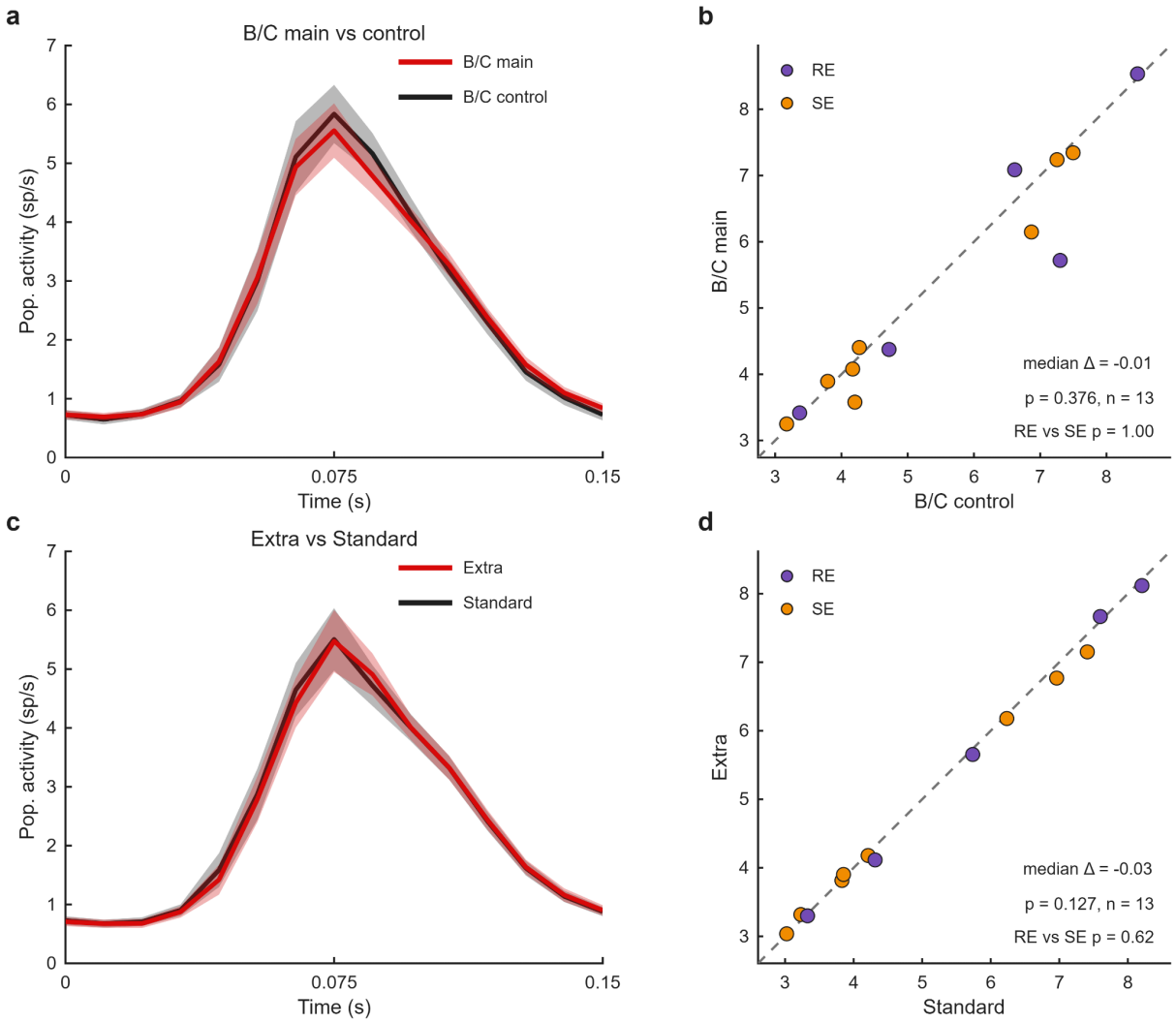


Figure 5: Response modulation in structured sequences is selective to deviant stimuli.
(a) Population-averaged B/C responses in the main block (red) and at matched positions in randomized control sequences (black), aligned to stimulus onset (0 s). Shaded regions, s.e.m. across sessions. **(b)** Session-level B/C amplitudes, main versus control block. Dots, individual sessions (RE, purple; SE, gold); dashed line, unity. B/C did not differ between contexts (median Δ = −0.01 sp/s, signed-rank p = 0.38, n = 13; rank-sum p = 1.0). **(c)** Population-averaged responses to familiar B/C elements at unexpected positions (Extra B/C; red) and to matched minimally adapted standards (B1, C3; black), aligned to stimulus onset. Comparisons were restricted to unexpected positions not immediately following the same stimulus identity (Extra B4, Extra C1/C2). Shaded regions, s.e.m. across sessions. **(d)** Session-level Extra versus Standard amplitudes, formatted as in (b). Extra and Standard did not differ (median Δ = −0.03 sp/s, signed-rank p = 0.13, n = 13; rank-sum p = 0.62).

We next asked whether enhancement generalized to familiar sequence elements appearing at unexpected positions within the sequence. If the modulation reflected a general consequence of violating predictions established by regularities in sequence content, familiar B/C elements presented out of place should likewise exhibit enhanced responses. To minimize

confounds introduced by short-timescale repetition adaptation, comparisons were restricted to minimally adapted standard positions (B2 and C4) and unexpected positions not immediately following the same stimulus identity (Extra B5 and Extra C2/C3). Unlike deviant stimuli, unexpected presentations of familiar B/C elements did not exhibit detectable enhancement relative to their matched standard responses (median session difference = −0.03 spikes/s, signed-rank $p = 0.13$, $n = 13$; Fig. 5c,d), and the effect did not differ between Random Exposure (RE) and Sequence Exposure (SE) sessions (rank-sum $p = 0.62$; Fig. 5d). Thus, violations of learned sequence content alone were insufficient to produce the enhancement observed for deviant stimuli, further arguing against a simple prediction-error account of the observed modulation.

Together, these findings indicate that the response enhancement observed for deviant stimuli was not a specific consequence of prediction violation tied to exact temporal sequence content. Expected B/C responses were not detectably suppressed in the repeating sequence, and familiar stimuli presented at unexpected positions did not show comparable enhancement despite violating learned sequence content. Instead, enhancement was observed specifically for deviant stimuli that occurred infrequently relative to the repeating B/C elements in the main block, but not when the same stimuli appeared in the randomized control context where their relative frequency was matched to surrounding elements. Thus, the observed modulation could not be explained by a simple prediction-error account in which predicted stimuli are suppressed and all expectation violations are enhanced, while it is consistent with a broader form of expectation-related modulation due to stimulus likelihood or environmental statistics. This distinction motivated analyses that more directly test whether expectation alters neural representations beyond overall response magnitude in a more temporally precise content-specific manner.

Expectation-related response modulation is broadly distributed across cortical layers

Predictive-coding frameworks often propose laminar specialization of expectation-related computations, with superficial cortical layers proposed to preferentially carry ascending prediction-error-related signals, layer 4 primarily reflecting feedforward sensory input, and deeper layers contributing to contextual or internally generated representations (Rao & Ballard, 1999; Keller & Mrsic-Flogel, 2018; Shipp, 2016). We therefore asked whether expectation-related response modulation differed across cortical depth by quantifying response amplitudes separately in layers 2/3, 4, and 5/6 for omission, deviant, and extra-stimulus conditions (Fig. S5a-d). Overall condition effects were assessed using signed-rank tests on session-averaged paired differences, whereas layer dependence was tested using linear mixed-effects models with session as a random intercept fit to paired condition differences across layers.

Despite substantial differences in overall response magnitude across cortical depth, expectation-related response modulation was broadly distributed across layers. Omission responses exceeded post-sequence rebound activity within all layers (Fig. S5a; condition $p = 0.0034$), with no significant layer-dependent modulation (layer × condition $p = 0.48$). Deviant responses in the standard sequence exceeded responses to the same stimuli in matched randomized control sequences across layers (Fig. S5b; condition $p = 0.027$), with no evidence for layer dependence (layer × condition $p = 0.16$). Enhancement of deviant responses relative to minimally adapted B/C stimuli was likewise observed across layers without significant layer dependence (Fig. S5c; condition $p = 0.048$, layer × condition $p = 0.82$). In contrast, responses to additional familiar stimuli presented at unexpected sequence positions did not differ from matched standard responses (Fig. S5d; condition $p = 0.22$, layer × condition $p = 0.85$).

Together, these findings indicate that expectation-related changes in response magnitude were broadly distributed across cortical layers and did not exhibit strong laminar specificity. Moreover, response enhancement differed across forms of expectation violation: deviant stimuli showed modest enhancement, whereas familiar stimuli presented at unexpected

positions did not, and omission-related enhancement showed distinct dependence on sequence position and exposure history. These observations suggest that increases in average response magnitude following expectation violation do not reflect a single underlying mechanism, motivating analyses that more directly test whether prediction shapes neural activity in a stimulus-specific manner beyond overall response magnitude.

Prediction context alters neural representations beyond nonspecific response enhancement

The analyses above show that unexpected events can change the magnitude of V1 responses. A change in response magnitude is, however, not by itself diagnostic of prediction-specific processing, because several mechanisms can produce one. Predictive-processing mechanisms that depend on the specific content of a violated prediction would be expected to alter activity in a prediction-dependent manner, such that the subset of neurons modulated by an expectation violation depends on what stimulus was expected. In contrast, expectation-related modulation can also arise from mechanisms sensitive to broader stimulus statistics or sequence structure without requiring moment-to-moment specificity about the predicted item. For example, responses to infrequent stimuli may be enhanced relative to a familiar and repetitive sensory context, such that enhancement depends on broader environmental statistics or average stimulus prevalence rather than the precise identity of the locally expected stimulus. All of these mechanisms can increase the overall response to an unexpected event, for example through additive mismatch signals or gain-like scaling of sensory responses.

Figure 6a–c illustrates these candidate mechanisms schematically in a simplified representational space. For visualization, only two arbitrary dimensions of the underlying high-dimensional population response are shown; these axes do not correspond to specific neurons or measured quantities. In each schematic, the same sensory event (“Stim”) is transformed into a neural population representation under two sequence contexts: S3,

corresponding to the representation of the stimulus at element 3, and S4, corresponding to the representation of the stimulus at element 4 of the sequence. Although the sensory event is matched, the predicted identity differs across these positions (B expected at element 3, C at element 4). Under additive mismatch signals (Fig. 6a), modulation adds a similar response component independent of prediction content, producing similar shifts from the sensory representation to S3 and S4. Under gain-like modulation (Fig. 6b), responses are scaled while preserving their underlying representational structure, again yielding largely overlapping S3 and S4 states when modulation magnitude is matched. In both cases, matched sensory events should occupy similar population states despite different prediction contexts. In contrast, prediction-specific modulation (Fig. 6c) alters neural activity according to the specific content of the prediction, such that the same sensory event is represented differently at element 3 (S3) versus element 4 (S4) depending on what was expected. The critical diagnostic is therefore not simply the size of the response but whether the temporally precise prediction content reshapes the representation of an otherwise matched event.

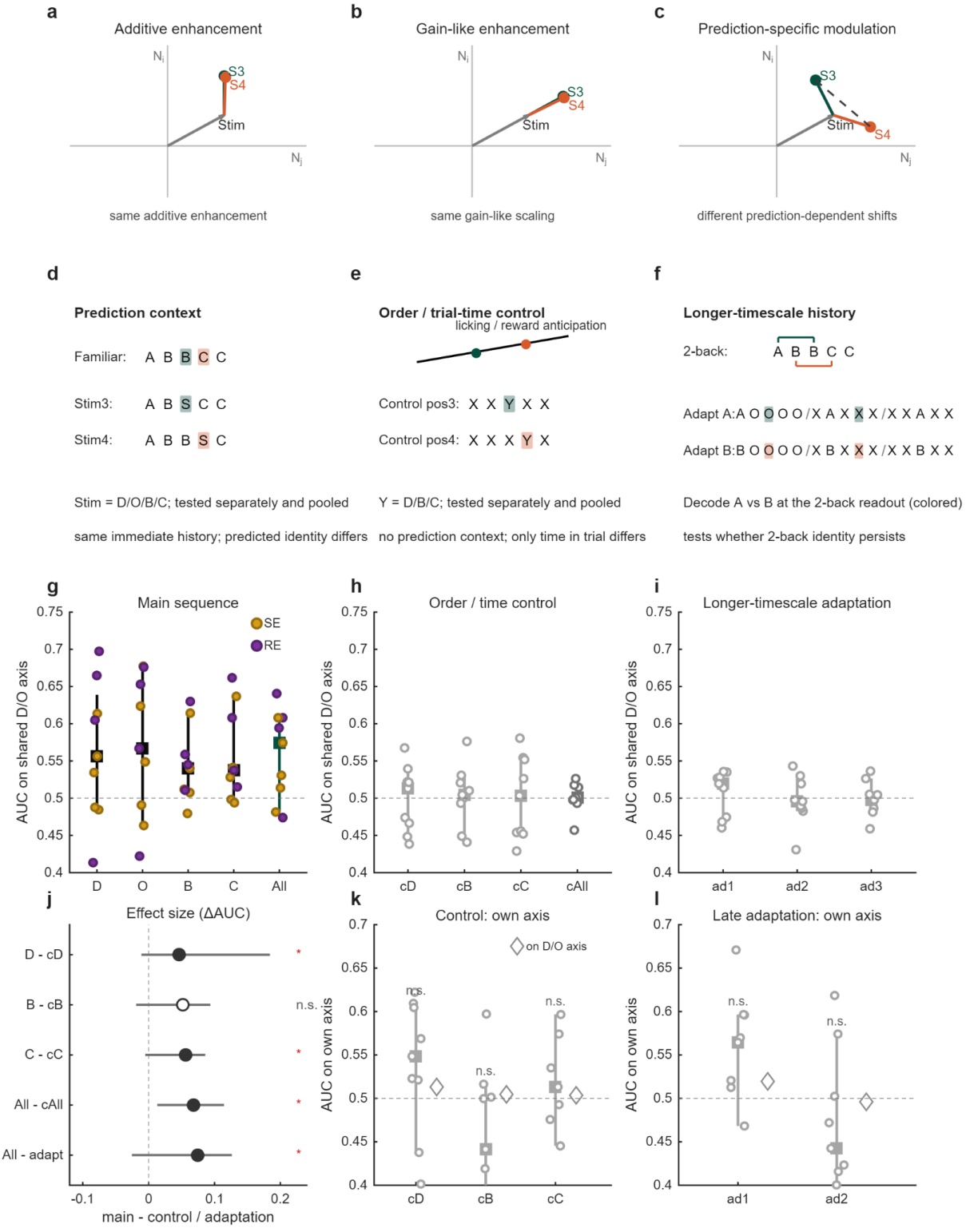


Figure 6: Prediction context alters neural population representations beyond nonspecific response enhancement.

(a–c) Candidate mechanisms in a low-dimensional representational space (two of many population dimensions shown). The gray vector ("Stim") represents the population response to the **same sensory event** before

prediction-dependent modulation. Colored endpoints indicate the resulting neural population representation at element 3 (S3) and element 4 (S4) of the sequence, where immediate sensory history is matched but predicted identity differs (B expected at S3, C at S4). **(a)** Additive enhancement: modulation adds a similar component independent of prediction content, producing similar shifts from the sensory representation to S3 and S4. **(b)** Gain-like enhancement: responses are multiplicatively scaled while preserving representational structure. Under both mechanisms, responses to the same sensory event remain largely overlapping despite different prediction contexts. **(c)** Prediction-specific modulation: prediction-dependent shifts reshape the representation differently at element 3 versus element 4, such that the same sensory event occupies distinct neural states (**S3 versus S4**) depending on what was predicted. **(d)** Prediction-context comparison. Element 3 (S3) and element 4 (S4) of the familiar ABBCC sequence share immediate sensory history (both preceded by B) but differ in predicted identity (B expected at S3, C at S4). Tested for deviant (D), omission (O), and familiar/extra B and C stimuli, and pooled across event types ("All"). **(e)** Order/trial-time control: matched control sequences present the same stimulus identities at the two positions without predictive structure, controlling for reward anticipation, licking, and trial-time signals. **(f)** Longer-timescale (two-back) history control: the stimulus two positions before element 3 versus element 4 differs (A vs B). A-versus-B identity is decoded at the two-back readout position (coloured token, two elements after A/B) in the control sequences AOOOO / XAXXX / XXAXX and BOOOO / XBXXX / XXBXX. **(g)** Main sequence, AUC on the shared element-3-vs-4 axis for D, O, B, C, and pooled (All); points, sessions coloured by exposure (RE, purple; SE, gold); square, median; bars, bootstrap 95% CI of the median; dashed line, chance. **(h)** Order/time controls (cD, cB, cC, cAll) projected on the same axis. **(i)** Two-back adaptation projected on the same axis, for the O-anchored (AdO[O]), X-embedded (AdX[X]), and pooled (AdO[O]+AdX[X]) conditions. **(j)** Effect sizes (ΔAUC): main minus matched control (D−cD, B−cB, C−cC, All−cAll) and main minus adaptation (All−adapt); filled marker, signed-rank $p < 0.05$; open, n.s.; CI = bootstrap median. **(k)** Each control decoded on its own cross-validated element-3-vs-4 axis (open diamond, the same control on the shared D/O axis). **(l)** Two-back adaptation decoded on its own A-versus-B axis, for the O-anchored (AdO[O]) and X-embedded (AdX[X]) conditions (open diamond, the same condition on the shared D/O axis); the pooled condition is omitted here because decoding pooled sub-conditions on a single own axis inflates AUC through sub-condition identity. Statistics are signed-rank against the within-session permutation null; n = 9 sessions (primary). Asterisks, $p < 0.05$.

To realize this comparison, we required two sequence positions that were matched in immediate sensory history, since differences in recent sensory input could themselves alter population responses, but differed in the stimulus identity predicted by the sequence. Elements 3 and 4 of the standard ABBCC sequence provide exactly this: both are immediately preceded by B, so their recent sensory history is identical, but the learned sequence predicts a different upcoming identity at each (B expected at element 3, C at element 4; Fig. 6d). The cleanest version of the test uses the deviant (D) and omission (O) conditions, in which the physical sensory event at the two positions is identical and only the predicted identity differs. We additionally examined familiar B and C stimuli at the two positions, where the bottom-up stimulus is again identical but one occurrence is expected and the other an "extra" presentation; these provide a less specific test because they partially conflate predicted identity with expected-versus-unexpected status. Critically, response amplitudes did not differ systematically between the two positions for any comparison (all signed-rank $p > 0.2$; Fig. S6a-d), indicating that response magnitude was approximately matched and that any population differences

primarily reflect how a matched event is represented rather than how strongly it drives the population.

To ask whether prediction content shaped neural representations, we built a single cross-validated decoding axis separating the two prediction contexts (element 3 versus element 4) and measured, per session, how well responses separated along it using the area under the ROC curve (AUC; 0.5 = chance, two-sided). The axis was trained on the deviant and omission conditions — the cleanest comparisons, in which the sensory event is matched across positions. In these conditions, nonspecific expectation-violation signals, such as rarity-related or sequence-structure-related mismatch, are also expected to be broadly similar across positions. By contrast, the familiar B and C conditions provide a less specific comparison because expectedness may differ across positions (expected versus extra presentation), partially conflating predicted identity with expectation violation despite the absence of systematic enhancement for extra stimuli. Principal-component projection followed by linear discriminant analysis was used to define the axis, and the familiar B and C conditions were then projected onto this same axis out of sample. We report a primary set of nine sessions that passed drift and trial-count quality control and confirm key effects in the full set of thirteen sessions (below). On this shared axis, the same sensory event was represented differently depending on prediction context: pooling across event types, the two contexts separated reliably above chance (All, median AUC = 0.57; signed-rank versus the within-session permutation null, $p = 0.039$), with the individual deviant, omission, B, and C comparisons each contributing in the same direction. The effect was reliable but modest in size — roughly 0.07 above chance — consistent with a weak but distributed influence of prediction context rather than a large reorganization of activity (see single-neuron analyses below).

We next asked whether this separability could instead reflect sequence order, trial time, reward-related signals, or other behavioral correlates rather than prediction content. In the interleaved control blocks, the same stimulus identities appeared at matched positions within

otherwise random sequences lacking the standard ABBCC structure; if separability reflected gradual within-trial changes — such as reward anticipation, licking, or trial-time signals — it should also appear here (Fig. 6e). This comparison is particularly important because licking behavior was broadly similar between the standard and control blocks, making behavioral differences an unlikely explanation for any separation effect. Projected onto the same axis, the matched controls sat at chance (cAll median AUC = 0.50), and the pooled prediction-context effect significantly exceeded the matched control (All − control = +0.068, bootstrap CI [+0.017, +0.114]; signed-rank p = 0.020; Fig. 6h,j). The per-comparison contrasts pointed in the same direction (C − cC = +0.056, p = 0.027; D − cD = +0.046, p = 0.039; B − cB = +0.052, p = 0.16; Fig. 6j). To verify that the controls' low values were not simply a consequence of reading them on an axis defined by the deviant and omission conditions, we also decoded each control on its own cross-validated element-3-versus-4 axis; they remained at chance there as well (cD = 0.55, cB = 0.44, cC = 0.51; all p > 0.16; Fig. 6k). The controls therefore carried little decodable structure of this kind even when given their own best axis, indicating that sequence order, trial time, and reward-related signals do not account for the prediction-context effect.

Finally, we asked whether longer-timescale sensory history could explain the effect. Although immediate sensory history was matched, the stimulus two positions back differed between the two contexts (A versus B; Fig. 6f). We therefore tested whether this two-back identity signal persisted by decoding A versus B at the matched two-back delay, using control sequences in which A or B were presented either alone or followed by intervening stimuli (AOOOO, XAXXX, XXAXX versus their B counterparts) and quantifying responses two positions downstream. Projected onto the prediction axis, this two-back adaptation signal was at chance (adaptation median AUC ≈ 0.50; Fig. 6i), and the pooled prediction effect exceeded it (All − adaptation = +0.074; signed-rank p = 0.039; Fig. 6j). Decoded on its own cross-validated axis, the two-back signal was likewise not significant (ad1 = 0.57, ad2 = 0.44; both p > 0.20; Fig. 6l), indicating that any adaptation signal had decayed by this delay rather than being obscured by

an unfavorable readout axis. That the decoder could detect this type of identity signal when present was confirmed using the same conditions evaluated one position earlier (one-back delay), where A-versus-B identity was clearly decodable on its own axis but fell to chance one position later (Fig. S7a), the expected signature of a transient adaptation effect. Two-back stimulus history therefore cannot account for the prediction-context separability. This interpretation was further supported by a complementary control using the main-block prediction axis itself. At positions 2 and 3, prediction is matched (B expected at both positions) while immediate sensory history differs (A versus B), such that the corresponding deviant/omission element-2-versus-3 axis provides a history-sensitive readout. One-back A-versus-B identity remained decodable on this axis (pooled median AUC = 0.57, signed-rank $p = 0.004$; Supplementary Fig. S7b), confirming that the main-block-defined decoder could recover sensory-history signals when present rather than systematically failing to read out control conditions. Together, these controls indicate that the chance-level two-back result reflects genuine decay of the sensory-history signal rather than insufficient decoder sensitivity.

An exploratory comparison by exposure did not reveal a clear sequence-exposure dependence: the prediction-versus-control effect was present in both Random- and Sequence-Exposure sessions and, if anything, trended larger in RE (All − control: RE median +0.091, SE +0.029; rank-sum $p = 0.29$). However, with only four and five sessions per group this comparison was underpowered and is not interpreted further.

Together, these analyses indicate that prediction context shapes population activity beyond what immediate sensory history, sequence order, trial time, reward-related signals, or lingering two-back adaptation can explain. The same sensory event occupied different population states depending on what the sequence predicted at that position, indicating that neural representations in V1 are influenced not only by stimulus identity but also by prediction content. We next asked whether this population-level effect was visible at the single-neuron level.

Robustness (full 13-session set). The pooled prediction-versus-control effect remained significant in the full session set (All − control $p = 0.017$). The C − cC and B − cB contrasts strengthened ($p = 0.0034$ and $p = 0.048$, respectively), whereas the deviant and adaptation contrasts weakened (D − cD $p = 0.11$; All − adaptation $p = 0.13$), likely reflecting the inclusion of four sessions with reduced main-block trial counts. Importantly, the control and two-back adaptation conditions remained at chance on both the shared prediction axis and their own optimized decoding axes.

Prediction content systematically organizes the population geometry

Having identified a prediction-content signal at the boundary between the two interior elements, we next asked whether this effect reflected a local comparison or a more general organization of the population representation. Representational similarity analysis (RSA) characterizes the geometry of the four within-family sequence positions through their six pairwise distances. Across these relationships, adaptation and prediction content can exert overlapping and, for some pairs, opposing effects (Fig. 7a). Positions 2 and 3 have different preceding stimuli (A versus B), causing adaptation to separate their representations, but share the same predicted stimulus (B), causing prediction content to draw them together. Positions 3 and 4 show the opposite pattern: both are preceded by B, causing adaptation to draw them together, but they predict different stimuli (B versus C), causing prediction content to separate them. Positions 4 and 5 repeat the logic of positions 2 and 3. Across the complete geometry, prediction content therefore groups positions 2–3 and 4–5, whereas adaptation groups positions according to preceding-stimulus identity and sequence order can produce additional separations according to ordinal position. Distinguishing these competing influences requires decomposing the observed geometry into adaptation-, prediction-, and order-related components measured within a common discriminant frame (Fig. 7a,b; Methods).

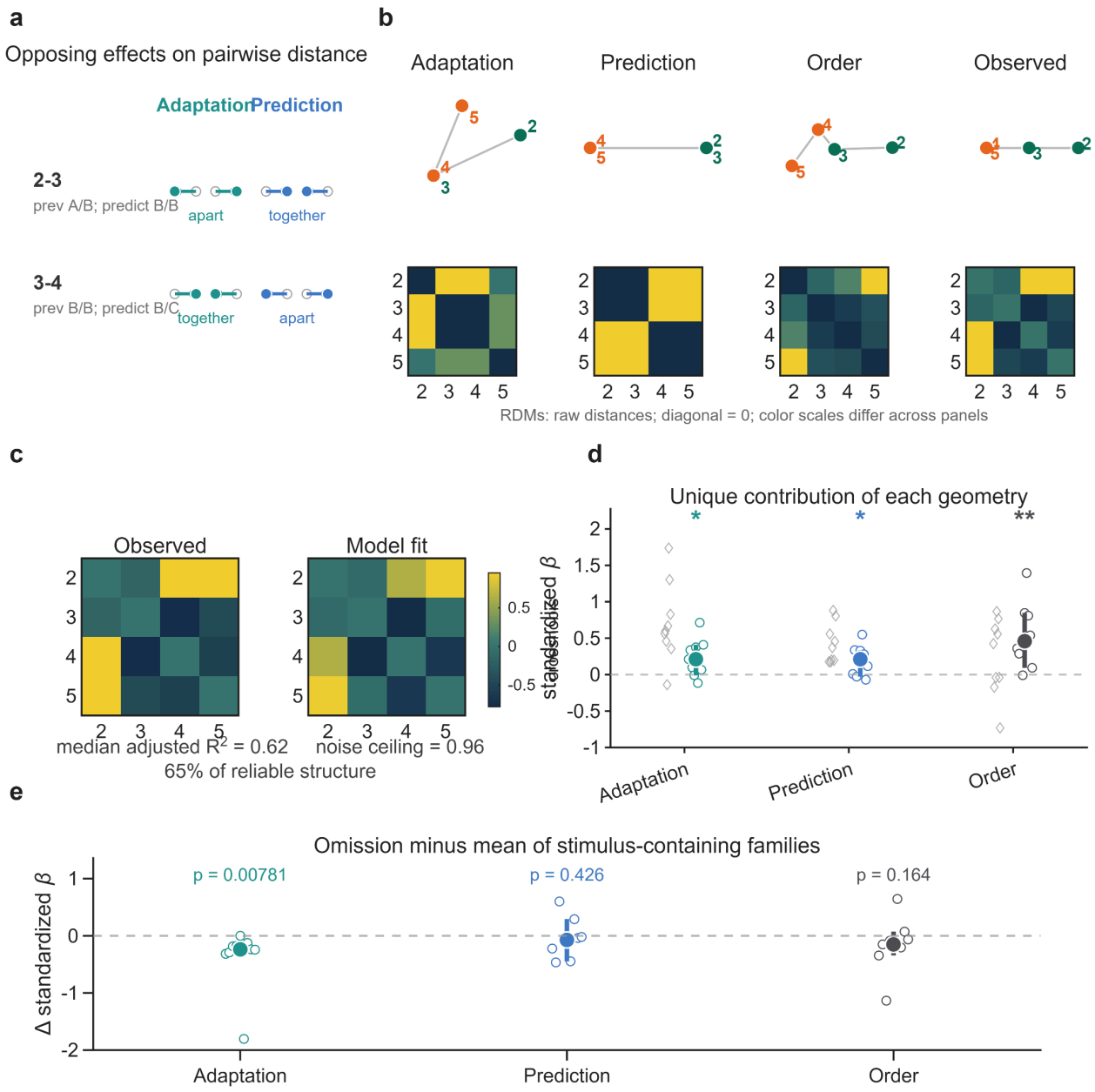


Figure 7: Prediction content shapes the within-family representational geometry.
(a) Opposing effects of adaptation and prediction content on selected pairwise distances. Positions 2 and 3 have different preceding stimuli (A versus B), so adaptation tends to separate their representations, whereas both positions predict B, so prediction tends to draw them together. Conversely, positions 3 and 4 have the same preceding stimulus (B), so adaptation tends to draw their representations together, whereas they predict different stimuli (B versus C), so prediction tends to separate them. Positions 4 and 5 follow the same logic as positions 2 and 3. Arrows indicate whether each component increases or decreases the expected pairwise distance. **(b)** Multidimensional-scaling (MDS) visualizations of the adaptation, prediction, sequence-order, and observed geometries, with the corresponding RDMs shown beneath. Green and orange points indicate positions predicting B and C, respectively; thin gray lines connect positions in sequence order. Adaptation organizes positions according to the preceding stimulus, prediction groups positions 2–3 and 4–5 into two clusters, and order is dominated by an onset-to-offset axis with the two interior positions relatively close. The observed geometry combines features of all three component geometries. MDS configurations were centered, scaled, and Procrustes-aligned to the observed configuration for visualization only; regression was performed on the original RDM entries. RDMs show raw distances with diagonals fixed at zero; the

same color palette is used throughout, but color limits differ across panels. Component geometries were grand-averaged across sessions, and the observed geometry was grand-averaged across sessions and sequence families. **(c)** Observed RDM and additive-model reconstruction, displayed using a shared color scale. The model reproduces the major features of the observed geometry, accounting for 65% of the reliable structure (median session-wise adjusted (R^2 = 0.62); median split-half noise ceiling = 0.96) in the discriminant space. **(d)** Standardized regression coefficients for adaptation (teal), prediction content (blue), and sequence order (charcoal). Colored open circles denote individual sessions in the discriminant space; filled circles and vertical bars indicate the median and bootstrap 95% confidence interval. Gray diamonds show the corresponding coefficients estimated in the unsupervised 50-dimensional principal-component space. Stars indicate signed-rank tests of the primary discriminant-space coefficients against zero: (*p < 0.05), (**p < 0.01), and (***p < 0.001). **(e)** Difference between the coefficient in the omission family and the mean coefficient across the stimulus-containing families (deviant, extra-B, and extra-C). Negative values indicate a weaker contribution in the omission family. Open circles denote individual sessions; filled circles and vertical bars indicate the median and bootstrap 95% confidence interval. Adaptation was reduced in the omission family ((\Delta\beta = -0.24), (p = 0.008)), whereas prediction content ((p = 0.43)) and sequence order ((p = 0.16)) did not differ detectably between omission and stimulus-containing families.

To separate these competing contributions, we estimated the representational geometry associated with the familiar sequence and with each candidate component as representational dissimilarity matrices (RDMs). An RDM summarizes the representational geometry by recording the pairwise distances between every pair of sequence positions. The observed, adaptation, and sequence-order RDMs were estimated empirically from cross-validated Mahalanobis (crossnobis) distances, whereas the prediction-content RDM was specified directly from the experimental design (Methods). Because an RDM specifies the distances between every pair of sequence positions, these distances define the underlying representational shape independently of its orientation in neural state space, allowing the geometry to be visualized using multidimensional scaling (MDS; Fig. 7b). We therefore asked whether the observed representational shape could be reconstructed as a weighted linear combination of the three candidate geometries using multiple regression (Methods). In this framework, each regression coefficient quantifies how strongly one candidate geometry contributes to the observed representational organization after accounting for the others. The resulting model accurately reproduced the major features of the observed representational geometry (Fig. 7c), accounting for 62% of the explainable variance (adjusted R^2=0.62 relative to a split-half noise ceiling of 0.96).

All three candidate geometries contributed significantly to the observed representation (Fig. 7D). Prediction content contributed reliably (standardized β = 0.21, *p* = 0.039, signed-rank

test across sessions), indicating that positions sharing the same predicted stimulus were represented more similarly than expected from recent sensory adaptation and sequence order alone. Adaptation also contributed significantly (β = 0.21, p = 0.039), as did sequence order (β = 0.46, p = 0.008). Thus, prediction content explained variance in the observed representational geometry beyond that accounted for by these empirically estimated components**.** Importantly, this contribution reflected the organization of the entire representational geometry rather than a single pairwise comparison. The same prediction geometry that grouped positions 2 with 3 and positions 4 with 5 simultaneously accounted for the pattern of relationships among all six position pairs. Thus, the prediction-dependent difference previously identified at the interior boundary (Fig. 6) emerges here as one consequence of a broader representational organization spanning the entire sequence.

Because the primary analysis was performed in a discriminant subspace optimized for the familiar-sequence geometry, we next asked whether the results depended on this representational frame. Prediction remained significant in the unsupervised 50-dimensional principal-component space (β = 0.37, p = 0.004), whereas the sequence-order component was no longer reliable (p = 0.30). Thus, irrespective of the contribution of sequence order, prediction content continued to explain variance in the observed representational geometry.

Prediction contributed consistently across sequence families despite their distinct sensory perturbations (Fig. 7e). The prediction coefficient did not differ between the omission family and the stimulus-containing families (omission minus the mean of deviant, extra-B, and extra-C: Δβ = −0.07, p = 0.43, signed-rank test), indicating that prediction organized the representational geometry to a similar extent regardless of how the expectation was violated. The sequence-order coefficient likewise did not differ detectably between omission and the stimulus-containing families (p = 0.16). In contrast, adaptation was significantly reduced in the omission family (Δβ = −0.24, p = 0.008), as expected because, in the absence of bottom-up

sensory input, there is little stimulus-driven response for adaptation to modulate. Thus, the analysis was sufficiently sensitive to detect a family-dependent change when one occurred, indicating that the absence of a corresponding effect for prediction reflects genuine consistency rather than limited statistical sensitivity. This family-invariant prediction contribution suggests a common representational transformation across sensory contexts. We therefore next asked whether this transformation corresponded to a shared geometric axis—that is, whether prediction shifted population representations in a parallel direction across sequence families.

Predicted context is represented as a common additive transformation

The family-invariant regression contribution suggests that prediction content may act as a shared representational transformation across sensory contexts. A strictly additive contribution would translate each family-specific state by the same vector, preserving the relative configuration of the families. By contrast, a family-dependent, non-additive transformation—including multiplicative or normalization-like modulation—could change the direction or magnitude of the shift as a function of the underlying sensory response (Fig. 8a). The similar prediction coefficients across families establish comparable contribution strength; we next asked whether prediction was also expressed along a shared geometric direction.

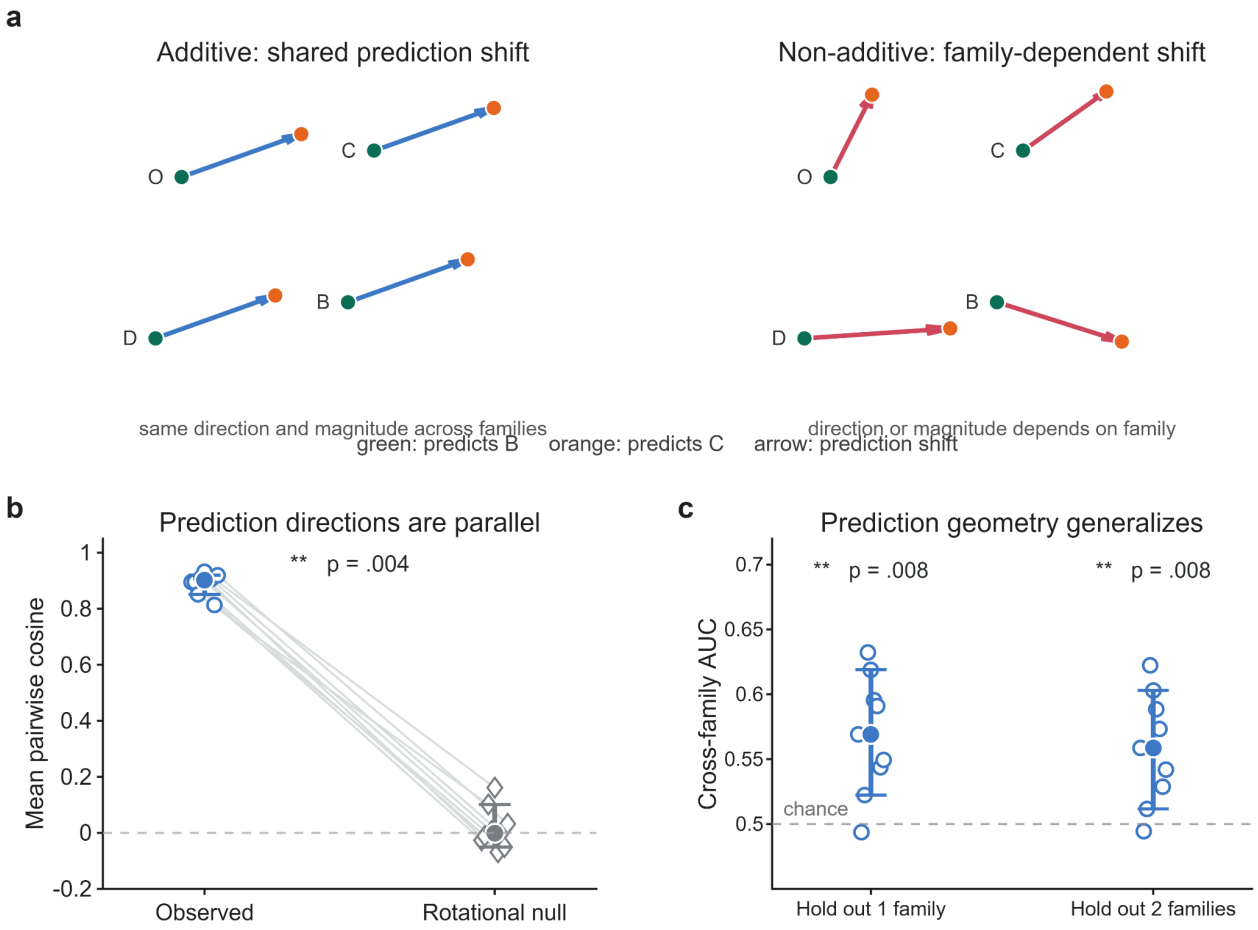


Figure 8: Prediction content defines a shared representational direction across sequence families.

(a) Schematic hypotheses for how prediction content could alter population geometry. Under an additive scheme (left), changing from a predict-B context to a predict-C context translates every family-specific representation by the same vector, preserving the relative configuration of the families. Under a non-additive, family-dependent scheme (right), the direction or magnitude of this shift depends on the starting state, deforming the configuration. Green and orange points denote predict-B and predict-C states, respectively; arrows denote the prediction-content shift. **(b)** Parallelism of the four family-specific prediction directions, quantified as the mean pairwise cosine similarity of the position-4 minus position-3 centroid differences in the discriminant space. Median observed parallelism was 0.90, substantially exceeding a matched-norm rotational null centered near zero (two-sided signed-rank test across nine sessions, ($p = 0.004$)). Lines connect observed and null values from the same session. **(c)** Cross-family generalization of the prediction-content axis. A linear boundary separating positions 3 and 4 was trained on a subset of sequence families and evaluated on held-out families; the position-discriminant subspace was refit using the training families only. Generalization exceeded chance when either one family was held out (median AUC = 0.57) or two families were held out (median AUC = 0.56; both ($p = 0.008$) versus chance and versus a label-permutation null, two-sided signed-rank tests). The dashed line indicates chance performance (AUC = 0.5). In **b** and **c**, open markers denote individual sessions; filled markers and bars indicate the median and bootstrap 95% confidence interval.

We therefore tested this second prediction directly using two complementary geometric analyses. Parallelism measures whether the prediction directions are aligned across families, whereas cross-condition generalization performance (CCGP) measures whether a decision boundary learned from one set of sequence families generalizes to different held-out families.

The prediction directions were highly parallel, with a mean pairwise cosine of 0.90, substantially exceeding a rotational null distribution (p = 0.004; Fig. 8b). Likewise, prediction generalized across families: a classifier trained to discriminate predicted context in one set of families accurately decoded held-out families (leave-one-family-out AUC = 0.57; leave-two-families-out AUC = 0.56; both p = 0.008 against chance and a label-permutation null; Fig. 8c). These analyses were performed in the primary discriminant space, with the representational space estimated using only the training families to prevent information leakage (Methods). Comparable performance was obtained in the unsupervised principal-component space (AUC = 0.56–0.57), demonstrating that the result does not depend on the choice of representational basis. Together, these findings indicate that predicted context is represented as a common additive transformation rather than a sensory-dependent multiplicative modulation.

The prediction axis is independent of the sensory representation

The finding that prediction is represented as a common additive transformation establishes how prediction enters the representation, but not the representational basis on which it acts. In the original linear formulation of predictive coding (Rao & Ballard, 1999), the symmetry between feedforward and feedback pathways predicts that prediction should be represented along the same neural dimensions as the sensory representation of the predicted stimulus, although the two may be aligned or anti-aligned depending on the cortical layer.

We therefore asked whether the prediction axis shares a representational axis with the sensory representation of the predicted stimulus. It did not. The prediction axis was no more aligned with the sensory direction separating the two predicted stimuli (B vs. C) than expected under a rotational null (null $z \approx 0$), nor was it more aligned than a control sensory direction separating two unrelated stimuli (A vs. D; null $z \approx 0$); the two alignment measures did not differ (paired signed-rank, p = 0.9). The same lack of alignment was observed in the denoised discriminant space and in the 50-dimensional principal-component space.

Because opposite alignment in different cortical layers could cancel in the population average, we next asked whether the relationship between prediction and sensory coding differed across cortical depth. It did not. The prediction axis was not significantly aligned with either sensory direction in layers 2/3, 4, or 5/6 (all null $z \approx 0$), and the degree of alignment did not differ between layers (all paired between-layer comparisons, n.s.). Repeating the analysis in a 20-dimensional principal-component space to match the reduced neuronal population available within individual layers yielded the same result.

Together, these results indicate that the prediction transformation identified above does not share a common representational axis with the sensory representation of the predicted stimulus. Thus, predicted context is not encoded by simply recruiting the sensory representation of the expected stimulus, but instead is represented along a neural direction that is distinct from it.

Prediction-content information is weakly distributed across individual units

Then we asked whether the prediction-content information identified by the population decoder (Fig. 6) reflected a small subset of highly prediction-specific units, as might be expected under specialized prediction-error or latent-estimate accounts, or instead a more distributed modulation across the V1 population. To answer this question, we computed for each unit the AUC discriminating sequence position 3 versus 4 (S3 versus S4) within each trial type and compared $|AUC - 0.5|$ to a per-unit trial-label permutation null (1,000 shuffles per unit). Because this analysis does not require estimating within-session covariance, all 13 usable sessions were included.

Across all four contrasts, stimulus-position information at the level of individual units was at most faint (Fig. 9). Mean $|AUC - 0.5|$ did not exceed the permutation null for the deviant contrast (D3 versus D4: 0.036 versus 0.036, signed-rank $p = 0.68$), showed only a small elevation for omission (O3 versus O4: 0.0335 versus 0.0316, $p = 0.034$), and was modestly but

significantly elevated for the standard-stimulus comparisons (B3 versus B4: 0.0255 versus 0.0222, $p = 1.2 \times 10^{-4}$; C3 versus C4: 0.0247 versus 0.0227, $p = 8.5 \times 10^{-3}$; Fig. 9, first column). Critically, however, no contrast exhibited a robust excess of significant units above the 5% chance level at the session level (D, O, and C all $p \geq 0.99$; B: 7.2%, $p = 0.046$; Fig. 9, second column), and the AUC distribution showed no heavier-than-null tail in any contrast (all $p \geq 0.13$; Fig. 9, third column). Thus, where a weak aggregate signal existed, it was broadly distributed across the population rather than concentrated in a strongly selective subset of neurons.

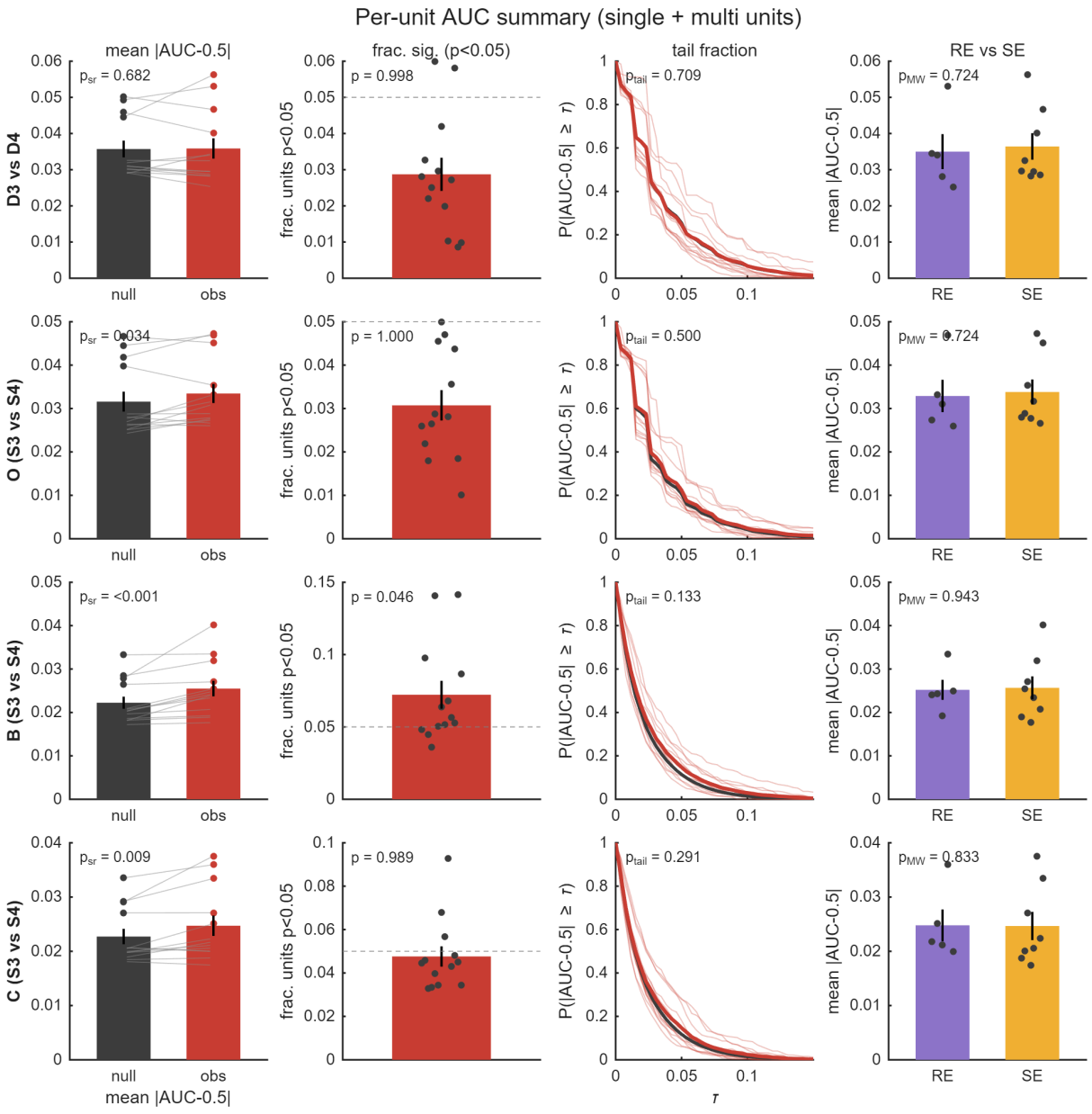

Per-unit AUC summary (single + multi units)
mean |AUC-0.5|
frac. sig. (p<0.05)
tail fraction
RE vs SE
D3 vs D4
O (S3 vs S4)
B (S3 vs S4)
C (S3 vs S4)
psr = 0.682
p = 0.998
ptail = 0.709
pMW = 0.724
psr = 0.034
p = 1.000
ptail = 0.500
pMW = 0.724
psr = <0.001
p = 0.046
ptail = 0.133
pMW = 0.943
psr = 0.009
p = 0.989
ptail = 0.291
pMW = 0.833
frac. units p<0.05
P(|AUC-0.5| ≥ τ)
mean |AUC-0.5|
null
obs
RE
SE
τ

Figure 9: Stimulus-position (S3 vs S4) information is not resolved by individual units.

Per-unit AUC discriminating the stimulus at position 3 from position 4 within four trial types (rows: deviant D3 vs D4; omission, B, and C standard, each S3 vs S4), single- and multi-unit pooled, versus a per-unit trial-label permutation null (1,000 shuffles per unit; all 13 usable sessions). Column 1, mean |AUC−0.5|, observed (red) versus null (grey); one point per session, lines pairing within session; p, one-sided signed-rank (observed > null). Column 2, per-session fraction of units significant at $p < 0.05$ (dots), bar = mean, dashed line = 5% chance; p, one-sided sign-test (fraction > 0.05). Column 3, tail fraction $P(|AUC-0.5| \geq \tau)$: thin lines, sessions; thick red, session mean; black, pooled null; p, sign-test on the per-session tail fraction. Column 4, mean |AUC−0.5| by exposure group (RE, purple; SE, gold); p, rank-sum. Error bars, SEM across sessions. No contrast shows a significant-unit excess or a heavy tail, indicating a broadly distributed rather than unit-concentrated code. These analyses use all 13 sessions; the population decoder (Fig. 6) uses 9, as it requires a stable covariance estimate that the per-unit test does not.

The magnitude of the per-unit signal did not differ between Random- and Sequence-Exposure sessions (all rank-sum $p \geq 0.72$; Fig. 9, fourth column) and showed no consistent laminar concentration (all $p \geq 0.10$; Fig. S8). Together, these findings suggest that prediction-content information in V1 is represented as a weak, distributed population code rather than by a small subset of strongly selective neurons.

To confirm that the weak single-unit effects in Fig. 9 did not reflect insensitivity of the analysis, we applied the same per-unit framework to contrasts expected to contain genuine single-neuron modulation (Fig. S9). In contrast to the stimulus-position comparisons, response-difference contrasts exhibited clear per-unit structure. Omission-versus-rebound responses exceeded null in aggregate (0.0196 versus 0.0157, $p = 3.1 \times 10^{-3}$), although without a significant excess of individually informative units (8.0%, $p = 0.13$; Fig. S9, first row). The deviant-versus-control comparison showed the strongest effect, with elevated mean $|AUC - 0.5|$ (0.0336 versus 0.0180, signed-rank $p = 1.2 \times 10^{-4}$), a large excess of significant units (19.6%, sign-test $p = 1.2 \times 10^{-4}$), and a heavy-tailed distribution ($p = 1.2 \times 10^{-4}$; Fig. S9, second row). Familiar-versus-extra comparisons also showed strong per-unit discriminability for both B (third row) and C (fourth row) stimuli (both $p = 1.2 \times 10^{-4}$), accompanied by significant-unit excesses (11.2% and 9.0%, both $p = 1.7 \times 10^{-3}$) and heavier-than-null tails (both $p \leq 1.7 \times 10^{-3}$). Importantly, however, signed mean effects in the familiar-versus-extra contrasts remained near zero (B: +0.004, $p = 0.068$; C: +0.002, $p = 0.31$), indicating balanced up- and down-modulation across units rather than a net directional shift. Thus, individual neurons in V1 can carry robust

information in this task, but prediction-content effects are represented differently: weakly at the level of single units and distributed across the population.

Robustness to multi-unit inclusion. Because the population decoder operated on all isolated and multi-units, the analyses above include both. Restricting to well-isolated single units (median 87–131 per session) preserved every conclusion. The deviant contrast remained concentrated (13.0% significant, $p = 1.2\times10^{-4}$; tail $p = 1.2\times10^{-4}$), the familiar-vs-extra contrasts remained strongly discriminable with near-zero signed means (B: 0.0210 vs 0.0164, $p = 1.2\times10^{-4}$; C: 0.0147 vs 0.0121, $p = 1.2\times10^{-4}$), and all four stimulus-position contrasts remained without a significant-unit excess (all $p \geq 0.87$), with aggregate signal at or below the all-unit level as expected from the reduced unit count.

Prediction-content decoding is supported by distributed unit contributions

Although individual units carried only weak prediction-content information (Fig. 9), the population decoder of Fig. 6 nonetheless reliably separated the two prediction contexts. We therefore asked whether the weak individual unit tuning identified above systematically contributed to decoder performance or whether the decoded signal instead arose from a small subset of highly informative neurons. To do so, we decomposed the population decoding axis into individual unit contributions (all units, single + multi; 9 sessions, as in Fig. 6) and quantified how decoding accuracy accumulated as progressively more units were added according to their contribution to the held-out decoder.

Prediction-content decoding accrued gradually as units were added and did not saturate within the highest-weighted subset (Fig. 10a). The top 20% of units captured no more of the signal than expected under a concentrated architecture (high- versus low-weight-first over the top 20%: AUC 0.529 versus 0.505, signed-rank $p = 0.25$), whereas the full-population decoder reached AUC 0.573 (versus chance, $p = 0.027$), matching the decoder of Fig. 6. Despite this

absence of concentration, unit contribution reliably predicted single-unit tuning: within each session, the magnitude of a unit's decoder weight correlated positively with its held-out AUC (median Spearman ρ = +0.80, signed-rank versus 0, p = 0.0039; Fig. 10b). Because decoder weights were estimated on training trials and AUC on held-out trials, this relationship is non-circular. Thus, prediction-content decoding is supported by a graded but broadly distributed code in which many weakly informative units contribute and no small subset dominates decoding.

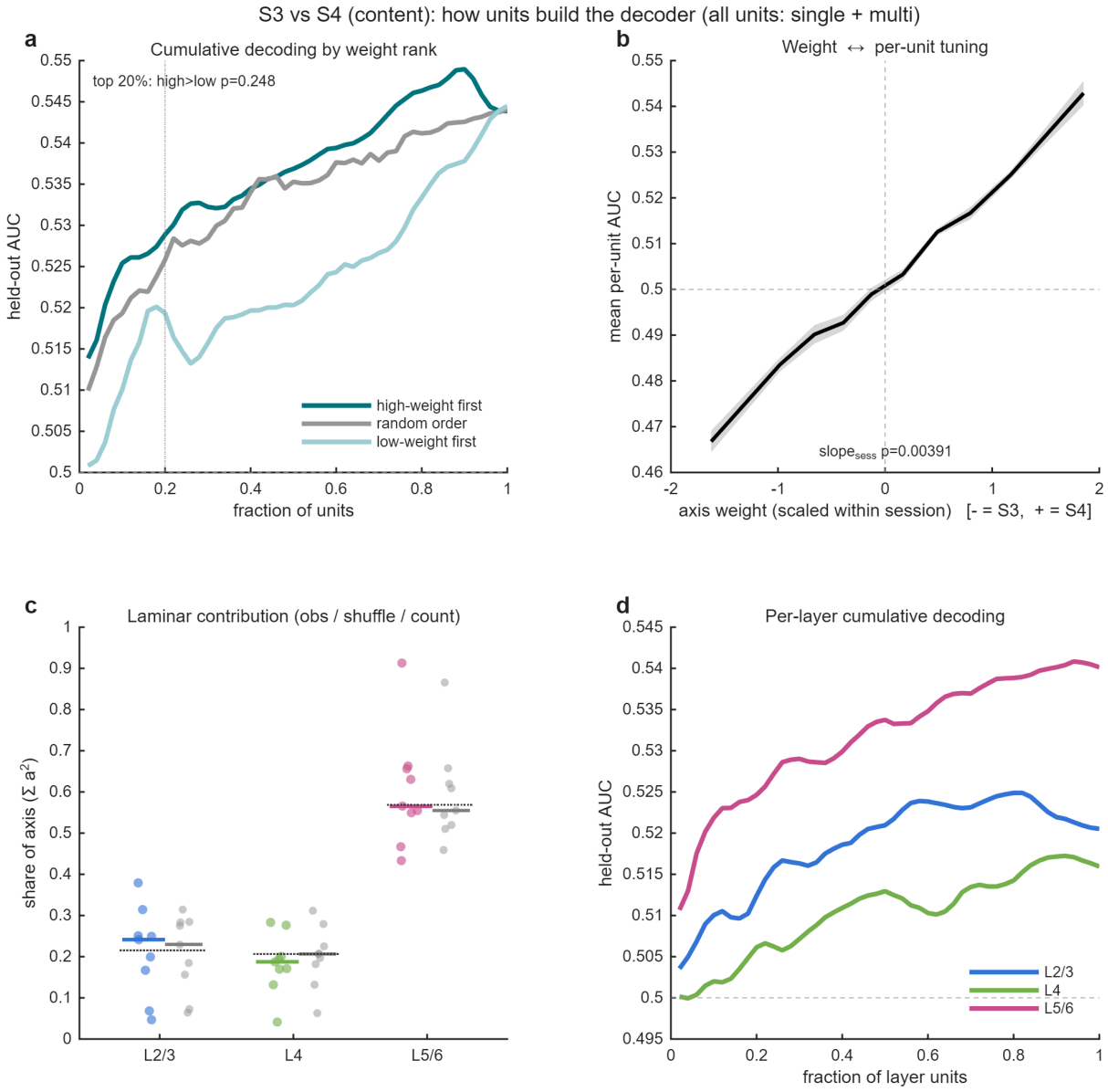


Figure 10: The content decoder is distributed across weakly tuned units.
Decomposition of the S3-vs-S4 population decoding axis into single-unit contributions (all units, single + multi; 9 sessions). (**a)** Held-out decoding accuracy as units are added in order of axis-weight magnitude: high-weight-first (saturated teal), random order (grey), low-weight-first (muted teal); mean across sessions. Accuracy climbs gradually rather than saturating in the top-weighted units. (**b**) Per-unit held-out AUC versus axis weight (scaled within session, sign preserved: − = S3, + = S4), pooled across sessions with the binned trend; quantified per session as the Spearman ρ between weight and held-out AUC, tested against zero across sessions (signed-rank, median ρ = +0.80, p = 0.0039). Weights are from training trials, AUC from held-out trials (non-circular). (**c**) Each layer's share of the axis (Σa^2; layer colour) versus a label-shuffle null (grey) and the neuron-count reference (dotted); no layer exceeds its shuffle (all $p \geq 0.10$). (**d**) Per-layer cumulative decoding (layer means); the apparent deep-layer advantage tracks unit count, not laminar concentration (see c). Panels a and d show means only; the statistics are in b and c.

Although deeper layers appeared to contribute more strongly to decoding (Fig. 10c-d), this reflected differences in unit count and variance rather than laminar specialization. Each layer's share of the decoding axis (Σa^2) was indistinguishable from a label-shuffle null matched for neuron count (L5/6 observed 0.56 versus shuffle 0.55 versus count 0.57, p = 0.25; no layer differed from its shuffle, all p ≥ 0.10; Fig. 10c), indicating that prediction-content information was distributed broadly across cortical depth rather than concentrated within a specific layer.

To determine whether this distributed architecture was specific to prediction-content decoding, we applied the same decomposition to the deviant-versus-control and omission-versus-rebound decoding axes (Figs. S10, S11). In contrast to the prediction-content axis, both mismatch-related decoders saturated rapidly within the highest-weighted subset of units. For the deviant-versus-control axis, the top 20% of units alone reached AUC 0.711 versus 0.489 for the lowest-weighted units (signed-rank p = 0.0020), approaching the full-population AUC of 0.721 (p = 0.0039; Fig. S10). The omission-versus-rebound axis showed a similar pattern (top-20% high 0.575 versus low 0.499, p = 0.0020; full AUC 0.575, p = 0.0039; Fig. S11). Both also showed strong positive weight–tuning relationships (deviant median ρ = +0.86, omission median ρ = +0.79, both p = 0.0039) but, like the prediction-content axis, no laminar concentration beyond that expected from neuron count (all layer shares ≈ shuffle; the single nominal exception, omission L2/3 p = 0.012, did not survive correction across layer comparisons). Together, these findings indicate that decoded signals in V1 differ not only in content but also in architecture: mismatch-related signals are concentrated within a relatively

small high-weight subset of neurons, whereas prediction-content information emerges from broadly distributed weak contributions.

Single-unit examples dissociate the violation response from content selectivity

The population analyses (Figs. 9, 10) indicate that violation-related responses and S3-versus-S4 prediction-content selectivity are only at most partly overlapping at the individual unit level. To visualize this dissociation, we selected example single units with clear evoked responses — transient deflections above the pre-sequence baseline — and characterized each along two axes: a violation axis (deviant versus control or omission versus rebound) and a content axis (S3 versus S4; Fig. 11).

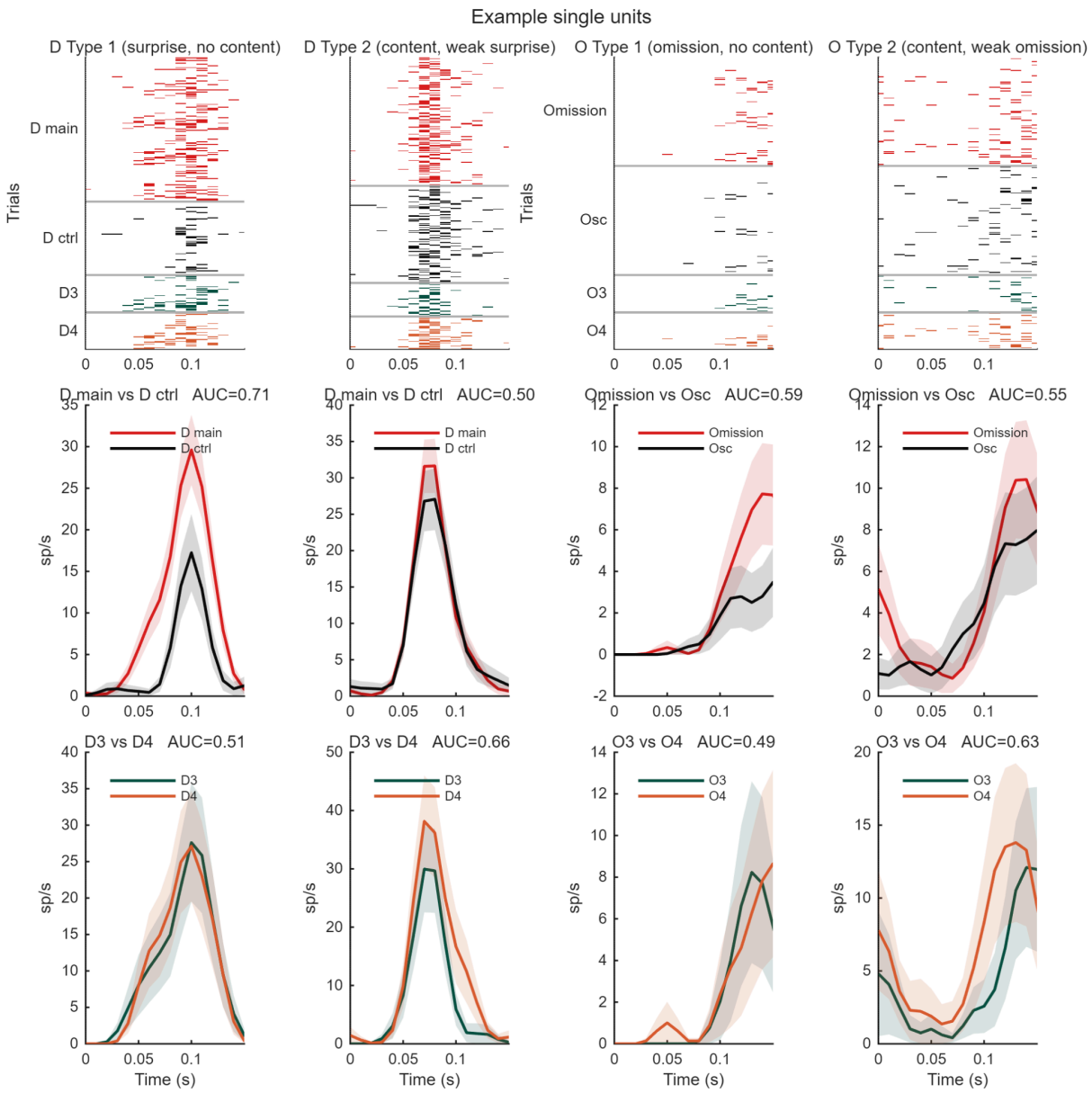


Figure 11: Single-unit examples dissociate the violation response from content selectivity.
Four example single units (columns), each selected for a genuine, sensibly shaped evoked response (a transient deflection above the pre-sequence baseline taken before the A onset; tonically active non-responsive cells excluded) and characterised on the violation axis (deviant vs control, or omission vs rebound) and the content axis (S3 vs S4). All trials shown; AUC reported on all trials (illustrative — population statistics are in Figs. 9–10). **Row 1**: rate raster (binned firing rate, sp/s; white→colour, unsmoothed) for the four conditions across trials, onset marked (dotted); display cropped to the 0–0.15 s response window. **Row 2**: violation-axis PSTH: D main vs D control (red/black) or omission vs rebound; AUC inset. **Row 3**: content-axis PSTH: S3 vs S4 (teal/coral), labelled D3/D4 or O3/O4; AUC inset. Shading, SEM across trials. **Columns 1, 3** show units with a clear deviant/omission response but no content selectivity (content AUC ≈ 0.5). **Columns 2, 4** show units with S3-vs-S4 content selectivity carried on a response that differs little (column 2) or only modestly (column 4) from control/rebound — a dissociation between the violation and content signals at the single-unit level. Single units (quality and firing-rate criteria as in Methods).

The examples illustrate a partial dissociation between the two signals. Some units showed strong violation-related responses but little content selectivity, including a

deviant-responsive unit with robust deviant-versus-control separation (AUC = 0.71) but no S3-versus-S4 difference (content AUC = 0.51; Fig. 11, column 1) and an omission-responsive unit with omission-versus-rebound separation (AUC = 0.59) but no content selectivity (AUC = 0.49; column 3). Conversely, other units exhibited clear prediction-content selectivity despite weak or absent violation responses, including a deviant-responsive unit with reliable S3-versus-S4 separation (content AUC = 0.66) but no deviant-versus-control difference (violation AUC = 0.50; column 2), and an omission-related example with content selectivity (AUC = 0.63) but only modest omission-versus-rebound separation (AUC = 0.55; column 4).

These examples are illustrative rather than statistical, with population-level conclusions established in Figs. 9 and 10. Nonetheless, they make concrete the dissociation implied by the population analyses: violation-related responses do not necessarily predict content selectivity, and prediction-content selectivity can be present even when violation responses are weak. Consistent with the distributed, weakly tuned architecture identified above, S3-versus-S4 selectivity often appeared as modest differences superimposed on largely overlapping response profiles, whereas violation-related responses were more often associated with larger separations in firing-rate responses.

Discussion

In this study, we asked whether the content of a prediction shapes sensory representations in the mouse primary visual cortex, using a sequence paradigm designed to hold the bottom-up sensory event, recent sensory history, and the overall degree of expectation violation approximately constant while varying only which stimulus was predicted. Two findings structure our interpretation. First, expectation violations produced modest and heterogeneous changes in population-averaged activity: deviant stimuli showed modest enhancement, omission responses weakly exceeded post-sequence rebound activity, and these effects differed in their dependence on sequence exposure, sequence position, and stimulus context. In

contrast, replacing an expected sequence element with an additional familiar stimulus did not produce enhancement. These findings indicate that responses scaling with expectation violation are not unitary and, by themselves, do not uniquely identify prediction-content-specific processing. Second, despite the modest changes in mean firing rate, the same sensory event — an identical deviant stimulus, omission, or familiar stimulus — was represented differently depending on which stimulus had been predicted. This prediction-content signal was modest in size but reliable, distributed across many weakly informative neurons rather than concentrated in a selective subset, and not reducible to trial time within the sequence, randomized control conditions lacking stable sequence structure, or short- and long-timescale adaptation controlled through matched sensory history and adaptation-specific control analyses. Moreover, representational analyses showed that prediction content systematically organized the geometry of population activity across sequence families, revealing a common representational structure that extended beyond pairwise decoding. Importantly, the sequence was designed such that prediction context depended on temporally extended sequence structure beyond the immediately preceding stimulus, and control analyses showed that sensory-history signals in this task decayed over shorter timescales than the prediction-context effect, consistent with predictive influences operating over longer temporal windows than local adaptation alone. Together, these findings suggest that mismatch-related response enhancement and prediction-content representations reflect at least partly dissociable processes: expectation violation can alter response magnitude through multiple interacting mechanisms, while prediction additionally modulates population activity in a prediction-specific manner.

Mismatch-related enhancement is not uniquely diagnostic of prediction-specific modulation

We showed that mismatch-related changes in response magnitude were heterogeneous across forms of expectation violation rather than reflecting a single uniform process. Unexpected omission of an expected sequence element elicited a transient enhancement

relative to post-sequence rebound activity, but this effect varied with sequence exposure and sequence position, consistent with sensitivity to broader sequence structure. Deviant stimuli likewise evoked modest enhancement relative to both expected B/C stimuli and matched randomized-control presentations, but unlike omission responses this effect remained stable across sequence exposure and within-session time and showed little dependence on sequence position. In contrast, replacing an expected sequence element with an additional familiar stimulus (extra B/C) did not produce enhanced responses despite violating the learned sequence. This dissociation suggests that deviant enhancement may depend not simply on violation of learned sequence content, but also on the relative rarity or novelty of the unexpected stimulus, because D responses increased in the standard sequence where D was infrequent whereas extra B/C responses — whose relative frequency matched that of expected B/C stimuli — did not. At the same time, rarity and expectation are closely linked in this paradigm, and the observed enhancement may reflect multiple interacting expectation-related mechanisms. Thus, different forms of expectation violation produced distinct response profiles, indicating that mismatch-related enhancement in V1 is not unitary and does not scale uniformly with the mere presence of violated expectation.

These findings complicate interpretations that equate enhanced responses to unexpected events with a unitary prediction-error signal or prediction-content-specific computation. Predictive-processing frameworks have often interpreted mismatch-related enhancement as reflecting violated expectations, particularly when unexpected stimuli evoke larger responses than expected ones (Rao & Ballard, 1999; Keller & Mrsic-Flogel, 2018; Shipp, 2016). However, several interacting processes can produce similar changes in average response magnitude, including saliency- or novelty-related enhancement of rare stimuli, behavioral relevance or attentional orienting, long-timescale sensory adaptation, and expectation-related modulation tied to broader sequence structure rather than the specific content of a violated prediction (Aitchison & Lengyel, 2017; Weber & Fairhall, 2019). Consistent

with this idea, Zhang et al. (2025) reported that different forms of expectation violation in mouse V1, including novel stimuli, omissions, and reordered or repeated familiar elements, evoke dissociable mismatch-related responses, arguing against a unitary temporally precise prediction-error signal underlying enhanced activity to unexpected events. Similarly, Homann et al. (2022) found that unlike reordered familiar elements, sparse novel visual stimuli evoke excess activity in mouse V1 that emerges rapidly and can be explained by adaptive local mechanisms, even without learned prediction content. Likewise, Hamm et al. (2021) showed that deviant stimuli recruit structured neuronal ensembles rather than producing a purely global increase in activity, while Furutachi et al. (2024) found that unexpected stimuli alter representational structure in ways more consistent with amplification of unexpected responses than purely content-specific mismatch coding. Importantly, these mechanisms are not necessarily independent and may interact nonlinearly, amplifying or masking one another depending on stimulus statistics and behavioral context. For example, Peterka et al. (2026) showed that strong local adaptation within repetitive sequences can partially mask global deviance responses, illustrating how adaptation and higher-order predictive effects can coexist while shaping sensory responses in opposing directions. This complexity provides an additional reason why isolating prediction-specific effects requires approximately matching nonspecific sources of expectation-related modulation — including rarity, broader sequence structure, and sensory history — while selectively varying prediction content. Together, these observations suggest that enhanced responses to expectation violation alone are insufficient to establish prediction-content-specific processing and should be interpreted cautiously when average response magnitude is used as the primary measure of predictive computation.

Prediction content shapes population activity beyond nonspecific expectation-related mechanisms

Although changes in average response magnitude were heterogeneous and difficult to interpret mechanistically, population representations revealed a more specific effect of prediction context. When the same sensory event was presented under different prediction contexts — while recent sensory history and nonspecific forms of expectation violation were approximately matched — neural population activity separated according to which stimulus had been predicted. In the cleanest comparisons (deviant and omission conditions), the unexpected event itself was identical across positions and occurred with matched frequency, reducing potential contributions of rarity- or saliency-related enhancement. Likewise, broader sequence-structure-related mismatch was approximately matched across contexts: omission and deviant responses showed little systematic position dependence, extra-versus-familiar responses were comparable across positions, and overall response magnitudes were closely matched. Thus, any residual differences in mismatch magnitude are unlikely to account for the observed separability. Instead, the same sensory event was represented differently depending on the prediction context. Beyond demonstrating pairwise separability, the representational analyses asked what variable best explained that separability. While an unmeasured contextual factor could in principle distinguish a single pair of conditions, it would not be expected to reproduce the broader representational geometry observed here, in which the relationships among within-family sequence positions are best explained by predicted stimulus identity. This therefore strengthens the interpretation that prediction content itself, rather than an unmeasured correlate of a particular comparison, is the primary factor underlying the representational differences observed in the decoding analyses.

Our findings complement recent work examining how temporal context and expectation shape sensory representations while addressing an important limitation of many prior paradigms. Wyrick et al. (2023) showed that neural population activity across visual cortical areas reflects temporal context and expectation, and that predicted stimuli can become represented differently across the cortical hierarchy in a manner broadly consistent with

hierarchical predictive coding. However, because prediction context covaried with stimulus identity, those analyses could not determine whether identical sensory events would be represented differently solely as a function of prediction content. By approximately matching sensory input, recent sensory history, and nonspecific forms of expectation-related modulation while selectively varying predicted identity, our design isolates this component more directly and suggests that prediction context itself can alter sensory representations in V1.

Our results also provide a complementary perspective on prior attempts to test prediction specificity in V1. Price et al. (2023) used violations of learned grating sequences to ask whether prediction errors were tuned to the spatiotemporal structure of training, finding responses broadly consistent with predictive coding but limited evidence for prediction-specific tuning of mismatch signals. One possibility is that higher-order predictions are not necessarily represented in terms of simple stimulus features such as orientation, but instead reflect broader latent or statistical regularities of sensory sequences. Under this view, prediction content may be expressed not as tuning to a specific expected feature but as a distributed shift in population state. Moreover, because the prediction-content effects observed here were subtle and distributed across many weakly informative neurons rather than concentrated in a highly selective subset, prediction-specific modulation restricted to a low-dimensional feature space such as orientation may be difficult to detect at the level of individual neurons or mean responses. By using natural image sequences and decoding population representations while approximately holding sensory events constant, our approach may have been better positioned to reveal this type of weak but distributed prediction-content signal. In addition to these experimental design differences, our use of representational analyses allowed us to move beyond explaining variance in individual responses or low-dimensional components and instead ask a more abstract question about the organization of the population representation itself. By characterizing the geometry of relationships among neural population responses, this approach

can reveal latent representational variables even when they account for only a small fraction of the variance in individual neurons.

Likewise, Furutachi et al. (2024) explicitly contrasted content-specific prediction-error accounts with additive surprise and gain-like amplification accounts, concluding that unexpected stimuli primarily amplify sensory representations rather than encoding the specific content of violated predictions. Importantly, this amplification result was derived from comparisons of expected versus unexpected stimuli and therefore may reflect mechanisms related to mismatch enhancement that are partly dissociable from prediction-content representations. Their argument against content specificity relied largely on the high similarity of neural responses to the same unexpected stimulus across different prediction contexts, which indicates that dominant sensory representations are largely preserved but does not preclude more subtle changes in the geometry of population representations. In our data, prediction-content effects were modest and broadly distributed across many weakly informative neurons, leaving dominant sensory responses largely preserved while remaining linearly decodable at the population level. Moreover, representational analyses that consider the relationships among all conditions simultaneously revealed systematic prediction-content structure that would not necessarily be apparent from pairwise comparisons alone. Thus, prediction-content information may coexist with amplification-like mismatch responses and emerge most clearly at the level of population representational geometry rather than pairwise response similarity.

A broader implication of these comparisons is that isolating prediction-content representations requires careful control of processes that can produce neural differentiation independent of prediction itself. Sensory adaptation is a major concern in temporally structured paradigms because recent sensory history can alter neural representations over extended timescales, sometimes persisting across intervening stimuli (Fritsche et al., 2022; Weber & Fairhall, 2019). Prior studies have addressed this challenge in different ways. For example, Wyrick et al. (2023) reduced local adaptation confounds by comparing across distinct sequence

contexts, whereas many oddball and contextual-deviance paradigms leave adaptation partially coupled to expectation.In the present task, we attempted to minimize these effects through several complementary approaches: brief stimulus presentations separated by gray intervals, matching immediate sensory history across prediction contexts, and explicit control analyses showing that one-back identity signals were decodable but decayed to chance by the two-back delay used in the primary comparison. Thus, although adaptation is difficult to eliminate entirely, our results suggest that the relevant timescale of sensory-history effects was substantially shorter than the prediction timescale probed here.

Beyond minimizing these potential confounds, our experimental design also enabled them to be characterized independently. Dedicated control conditions isolated the representational structure associated with recent sensory history and sequence order, allowing these factors to be evaluated directly within the representational geometry of the main sequences rather than through comparisons across different sequence contexts. This allowed us not only to test whether prediction content alone distinguished the main-sequence representations, but also to ask how much of the organization of the main-sequence geometry could be explained by independently characterized representations of sensory history and sequence order. Under this framework, prediction-related and sensory-history-related influences were well approximated by linearly additive representational components, whereas sequence order contributed more weakly and primarily along the dominant onset-to-offset dimension expressed within the sequence representation.

Importantly, these analyses do not assume that sequence order or other contextual influences are represented identically in the control and main sequences. Embedding stimuli within an extended predictive sequence may itself alter the representational geometry associated with sequence order or introduce additional behavioral or contextual influences absent from the controls. Rather than directly comparing representations across blocks, our analyses asked how much of the geometry observed in the main sequences could be explained

by representational structures independently characterized in the control conditions. Consequently, any additional sequence-dependent influences would have to reproduce the systematic prediction-content organization observed across the full representational geometry of the main sequences to account for our findings. This provides a substantially stronger test than demonstrating separability of a single comparison alone.

Behavioral relevance provides another potential source of differentiation. Expectation-related signals in the sensory cortex can depend strongly on task demands and behavioral goals (Solomon et al., 2021), reward contingencies (Garner & Keller, 2022), and proximity to reward or goal states (Furutachi et al., 2024). Under such conditions, even when expectation violation itself is approximately matched, neural responses may still diverge because prediction context covaries with additional motivational or behavioral signals. This possibility is particularly important because a matched expectation violation combined with an additional differentiating signal can still produce separable neural states. We therefore designed the paradigm to approximately match or explicitly control several major nonspecific influences simultaneously — including recent sensory history, longer-timescale adaptation, relative rarity, sequence-structure violation, trial timing, and reward timing — while varying primarily the identity of the predicted stimulus. The persistence of population separability under these conditions suggests that prediction context contributes information beyond these alternative sources of differentiation. Moreover, representational analyses showed that the broader organization of the population code is best explained by prediction content rather than by independently characterized representations sequence order, making it unlikely that residual behavior alone accounts for the observed representational structure.

Prediction content contributes a shared linear component across sensory contexts

The representational analyses extend the pairwise decoding result by showing that prediction content contributes a common geometric structure across qualitatively different

sensory conditions. In the RSA, prediction explained the organization of sequence positions after accounting for adaptation and order, and its contribution did not differ detectably between the omission and stimulus-containing families. This consistency is particularly informative because adaptation was significantly reduced in the omission family, demonstrating that the analysis could detect family-dependent changes when they occurred. Prediction content therefore contributed comparably to the representational geometry whether the sensory outcome was a deviant stimulus, an omission, or a familiar B or C stimulus.

The shared-axis analysis in Figure 6 provided an initial, restricted indication of this generality: an axis defined using the deviant and omission conditions also separated prediction contexts for familiar B and C stimuli that did not contribute to its estimation. Parallelism and CCGP tested this property systematically across all possible sensory families while preventing held-out families from contributing to either the representational subspace or the classifier. Prediction-content directions were closely aligned across families, and a linear boundary learned from some families generalized to families excluded from training. Moreover, held-out CCGP was numerically comparable to decoding obtained using family-specific axes. Although equivalence was not tested formally, this similarity suggests that generalization incurred little additional loss relative to decoding each family in its own representational frame: the modest accuracy primarily reflects the weakness of the prediction-content signal overall, rather than a substantially different coding direction in each sensory condition.

Within the geometric framework developed by Bernardi et al. (2020) and subsequently applied by Courellis et al. (2024), parallel coding directions and above-chance CCGP indicate that a variable is represented in an abstract or disentangled format: its value can be recovered across changes in the other conditions defining the population state. In this operational sense, prediction content was represented at least partly independently of the sensory event through which it was expressed. Together with the comparable RSA coefficients, these results are consistent with prediction contributing an approximately additive component that is partially

preserved across sensory states. They therefore constrain the transformation underlying the contextual effect more strongly than separability alone and are less consistent with prediction being expressed entirely through event-specific mismatch responses or strongly input-dependent modulation. At the same time, the prediction direction was not aligned with the sensory direction separating B from C, indicating that this shared component was not simply a reinstatement of the sensory representation of the expected stimulus.

Additivity itself, however, is not unique to prediction. Short-timescale sensory-history effects can retain stimulus-like structure resembling feedforward drive and can combine approximately linearly with current input, particularly for excitatory response components, although suppressive components may combine sublinearly (Deveau et al., 2026). We did not systematically test whether the adaptation component exhibited the same parallelism or cross-family generalization as prediction. The present analyses therefore do not establish that prediction is uniquely additive relative to sensory history. Rather, they show that prediction explains an additional component of representational geometry after accounting for an empirically derived adaptation geometry and that an additive decomposition provided an adequate description of their joint contributions. Consistent with this approximation, the model explained a substantial fraction of the reliable geometry, residuals showed no systematic relationship with fitted values or position pairs, the coefficients were robust across representational spaces, and an explicit adaptation-by-prediction interaction geometry explained negligible additional variance. These diagnostics support the additive model over the conditions sampled, but do not establish that the underlying circuit operations are globally linear or exclude nonlinear interactions outside this regime. The central result is therefore that prediction contributes a partially preserved, linearly accessible component across distinct sensory outcomes, rather than that additivity alone uniquely identifies predictive coding.

Prediction-content signals are weakly distributed across population activity

A striking feature of the prediction-content signal was its distributed organization. Although prediction context was reliably decodable at the population level, individual neurons carried only weak information, with no excess of strongly informative units and little evidence for highly selective single-neuron tuning. Instead, decoding accuracy accumulated gradually across many units, and prediction-context information was distributed broadly across both neurons and cortical layers rather than concentrated within a small subset of strongly weighted units or a specific laminar compartment. Apparent differences in layer-level decoding were explained by neuron count and variance rather than selective concentration of the signal. In contrast, deviant and omission responses exhibited a more concentrated architecture, with a subset of high-weight neurons carrying a disproportionate share of the signal. Thus, prediction content in V1 appears to be represented not by a specialized population of strongly tuned neurons but by modest, coordinated biases distributed across many weakly informative cells spanning cortical layers.

This distributed organization may help explain why prediction-content effects have been difficult to detect in previous studies focused on average response magnitude or single-neuron selectivity. If prediction context acts primarily by subtly shifting population state while leaving dominant sensory responses largely intact, individual neurons may show only weak or heterogeneous modulation even when the population representation remains reliably distinguishable. In this view, prediction-content information resembles a contextual modulation of sensory representations rather than a discrete prediction-specific response superimposed on sensory activity. Such an organization is broadly consistent with hierarchical predictive-processing accounts in which top-down expectations modulate ongoing sensory representations, while remaining agnostic about the precise circuit or computational implementation of those signals.

These findings place only modest constraints on predictive-processing theories but suggest that prediction-content signals in V1 may be implemented in a more distributed manner

than strong single-neuron “prediction-error” accounts might imply. Importantly, not all predictive-processing frameworks posit explicit populations of neurons encoding prediction errors. For example, predictive-coding formulations differ in whether predictions and mismatches are represented by dedicated error units or emerge through distributed recurrent inference within sensory populations (Heeger, 2017; Keller & Mrsic-Flogel, 2018). Moreover, even models that emphasize error signaling do not necessarily require prediction content to be concentrated in strongly selective neurons: information about violated expectations could instead be multiplexed with sensory signals across partially distinct population subspaces or routed through circuit-specific pathways while remaining weakly expressed at the level of individual cells.

The absence of strong single-neuron prediction-content tuning also does not preclude prediction signals from being implemented through subcellular or compartment-specific mechanisms. Several theories propose that mismatch-related signals are represented in dendritic or membrane-potential dynamics rather than spiking output alone, and recent work has demonstrated differential prediction-related effects across somatic and apical dendritic compartments (e.g., Gillon et al., 2024). More generally, recurrent interactions between cortical populations imply that prediction-related modulation may propagate broadly through sensory networks even when its local expression is modest. Under this view, the weak but distributed prediction-content signal observed here may reflect one component of a larger predictive computation rather than a dedicated prediction-specific neural code in V1.

Limitations

Several limitations should be considered when interpreting these findings. First, although our design approximately matched multiple nonspecific influences on neural responses — including recent sensory history, relative rarity, broader sequence structure, reward timing, and trial timing — these factors cannot be perfectly equated in a temporally structured sensory task.

For example, prediction contexts necessarily differed in position within the learned sequence and therefore in elapsed sequence time, even when immediate sensory history and expected identity were matched. However, randomized control conditions lacking stable sequence structure suggested that sequence position or trial order alone were insufficient to explain the observed separability, and we found little evidence for systematic position dependence in the primary prediction-content comparisons. Thus, although nonlinear interactions between sequence structure and prediction content cannot be fully excluded, they provide a less parsimonious explanation for the observed effects than a contribution of prediction context itself. Likewise, prediction-content comparisons involved natural images embedded in a high-dimensional visual space, making it difficult to precisely define or equalize the relationship between expected and observed sensory inputs. Residual differences in mismatch magnitude could therefore contribute weak nonlinear effects, for example through changes in precision or gain. However, the absence of systematic position dependence in deviant responses, closely matched omission and deviant response magnitudes across prediction contexts, matched relative frequencies, and explicit controls for short- and longer-timescale adaptation argue against these factors fully accounting for the observed population separability. Moreover, to the extent that residual differences reflect the specific mismatch between predicted and observed sensory states, they remain part of a prediction-content-dependent signal rather than an alternative explanation for it.

Several aspects of the present design also constrain interpretation while motivating future work. The experiments were conducted in passively viewing mice receiving predictable rewards, minimizing explicit task demands. Although this limits conclusions about how prediction-content signals are behaviorally used, it also demonstrates that such representations can emerge in the sensory cortex without active task engagement or explicit behavioral relevance (Gavornik & Bear, 2014; Hamm et al., 2021; Zhang et al., 2025). At the same time, prior work suggests that predictive signals are shaped by task demands, reward contingencies,

and behavioral goals (Solomon et al., 2021; Garner & Keller, 2022; Furutachi et al., 2024), raising the possibility that prediction-content representations may be amplified, reshaped, or selectively routed during active behavior. Recordings were restricted to V1, leaving unresolved whether the observed signals emerge locally or reflect feedback from higher cortical areas. Likewise, our analyses focused primarily on temporally averaged response windows; although time-resolved decoding was underpowered in the present dataset, prediction-content signals may evolve dynamically across time in ways not captured here. Finally, the limited number of random- and sequence-exposure sessions constrained statistical power to assess learning-related differences. More extensive longitudinal and multi-area recordings, particularly during behaviorally engaged tasks, will be important for clarifying how distributed prediction-content signals are learned, generated, and routed across cortical hierarchies.

Conclusion

Together, these findings show that expectation-related activity in mouse V1 cannot be reduced to average mismatch responses, but instead includes a subtle, distributed, prediction-content-dependent population signal that coexists with broader response modulations driven by sequence structure, stimulus context, and exposure history.

Author Contributions

S.M. conceived the study, designed the research, performed the experiments, analyzed the data, and wrote the original draft of the manuscript. N.A.S. contributed to conceptualization and methodology, supervised the project, and provided resources and manuscript revisions. R.P.N.R. contributed to conceptualization and study design and provided manuscript revisions.

Data Availability:

The data that support the findings of this study are openly available in figshare at https://figshare.com/s/3351ea07edea95874d5b , reference number …

Acknowledgments:

We thank Koosha Khalvati, Preston Jiang, Ares Fiser, and members of the Steinmetz laboratory for helpful discussions and feedback on the project. We thank Ljuvica Kolich for assistance with animal training and surgical procedures. We are especially grateful to Fred Rieke, Adrienne Fairhall, and Elizabeth A. Buffalo for insightful feedback, thoughtful discussions, and guidance that substantially improved this work. Large language model-based artificial intelligence tools were used to aid in the preparation of this manuscript. The authors take responsibility for the content of the work.

References

Aitchison, L., & Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current opinion in neurobiology*, *46*, 219-227. https://doi.org/10.1016/j.conb.2017.08.010

Audette, N. J., Zhou, W., La Chioma, A., & Schneider, D. M. (2022). Precise movement-based predictions in the mouse auditory cortex. Current biology : CB, 32(22), 4925–4940.e6. https://doi.org/10.1016/j.cub.2022.09.064

Audette, N. J., & Schneider, D. M. (2023). Stimulus-specific prediction error neurons in mouse auditory cortex. bioRxiv : the preprint server for biology, 2023.01.06.523032. https://doi.org/10.1101/2023.01.06.523032

Bastos, G., Holmes, J. T., Ross, J. M., Rader, A. M., Gallimore, C. G., Wargo, J. A., … Hamm, J. P. (2023). Top-down input modulates visual context processing through an interneuron-specific circuit. *Cell Reports, 42*(9). https://doi.org/10.1016/j.celrep.2023.113090

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. Cell, 183(4), 954–967.e21. https://doi.org/10.1016/j.cell.2020.09.031

Biderman, D., Whiteway, M. R., Hurwitz, C., Greenspan, N., Lee, R. S., Vishnubhotla, A., Warren, R., Pedraja, F., Noone, D., Schartner, M. M., Huntenburg, J. M., Khanal, A., Meijer, G. T., Noel, J. P., Pan-Vazquez, A., Socha, K. Z., Urai, A. E., International Brain Laboratory, Cunningham, J. P., Sawtell, N. B., … Paninski, L. (2024). Lightning Pose: improved animal pose estimation via semi-supervised learning, Bayesian ensembling and cloud-native open-source tools. Nature methods, 21(7), 1316–1328. https://doi.org/10.1038/s41592-024-02319-1

Burgess, C. P., Lak, A., Steinmetz, N. A., Zatka-Haas, P., Reddy, C. B., Jacobs, E. A., … Carandini, M. (2017). High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice. *Cell Reports, 20*(10), 2513–2524. https://doi.org/10.1016/j.celrep.2017.08.021

The International Brain Laboratory, Banga K, Boussard J, Chapuis G, Faulkner M, Harris K, Huntenburg J, Hurwitz C, Lee HD, Paninski L, Rossant C, Roth N, Steinmetz N, Windolf C, Winter O. Spike sorting pipeline for the International Brain Laboratory. figshare. 2022. https://doi.org/10.6084/m9.figshare.19705522

Clark, A. (2016). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Courellis, H. S., Minxha, J., Cardenas, A. R., Kimmel, D. L., Reed, C. M., Valiante, T. A., Salzman, C. D., Mamelak, A. N., Fusi, S., & Rutishauser, U. (2024). Abstract representations emerge in human hippocampal neurons during inference. *Nature*, 632(8026), 841–849. https://doi.org/10.1038/s41586-024-07799-x

den Ouden, H. E., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in psychology*, *3*, 548. https://doi.org/10.3389/fpsyg.2012.00548

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS computational biology, 13(4), e1005508. https://doi.org/10.1371/journal.pcbi.1005508

Finnie, P. S., Komorowski, R. W., & Bear, M. F. (2021). The spatiotemporal organization of experience dictates hippocampal involvement in primary visual cortical plasticity. *Current Biology, 31*(18), 3996–4008.e3. https://doi.org/10.1016/j.cub.2021.07.056

Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., & Keller, G. B. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nature Neuroscience, 19*(12), 1658–1664. https://doi.org/10.1038/nn.4370

Friston K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London*. Series B, Biological sciences, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Furutachi, S., Franklin, A. D., Aldea, A. M., Mrsic-Flogel, T. D., & Hofer, S. B. (2024). Cooperative thalamocortical circuit mechanism for sensory prediction errors. *Nature, 633*(8029), 398–406. https://doi.org/10.1038/s41586-024-04422-y

Garner, A. R., & Keller, G. B. (2022). A cortical circuit for audio-visual predictions. *Nature Neuroscience, 25*(1), 98–105. https://doi.org/10.1038/s41593-021-00966-y

Garrett, M., Manavi, S., Roll, K., Ollerenshaw, D. R., Groblewski, P. A., Ponvert, N. D., … Olsen, S. R. (2020). Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. *eLife, 9*, e50340. https://doi.org/10.7554/eLife.50340

Garrett, M., Groblewski, P., Piet, A., Ollerenshaw, D., Najafi, F., Yavorska, I., ... & Olsen, S. R. (2023). Stimulus novelty uncovers coding diversity in visual cortical circuits. *BioRxiv*, 2023-02. https://doi.org/10.1101/2023.02.14.528085

Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience, 17*(5), 732–737. https://doi.org/10.1038/nn.3683

Gillon, C. J., Pina, J. E., Lecoq, J. A., Ahmed, R., Billeh, Y. N., Caldejon, S., … Zylberberg, J. (2024). Responses to pattern-violating visual stimuli evolve differently over days in somata and distal apical dendrites. *Journal of Neuroscience, 44*(5). https://doi.org/10.1523/JNEUROSCI.1234-23.2024

Hamm, J. P., Shymkiv, Y., Han, S., Yang, W., & Yuste, R. (2021). Cortical ensembles are selective for context. *Proceedings of the National Academy of Sciences, 118*(14), e2026179118. https://doi.org/10.1073/pnas.2026179118

Heeger D.J (2017). Theory of cortical function, *Proc. Natl. Acad. Sci. U.S.A.* 114 (8) 1773-1782, https://doi.org/10.1073/pnas.1619788114

Homann, J., & Berry, M. J. (2022). Novel stimuli evoke excess activity in the mouse primary visual cortex. *Proceedings of the National Academy of Sciences, 119*(5), e2108882119. https://doi.org/10.1073/pnas.2108882119

Huang, Y., Shamsnia, A., Chen, M., Wu, S., Stamm, T., Medico, S., & Najafi, F. (2025). Intrinsic timing, not temporal prediction, underlies ramping dynamics in visual and parietal cortex during passive behavior (Advance online publication). *bioRxiv*.

Jiang, L.P., & Rao R.P.N.(2022). Predictive Coding Theories of Cortical Function , in S Murray Sherman (ed.), *Oxford Research Encyclopedia of Neuroscience.* https://doi.org/10.1093/acrefore/9780190264086.013.328

Keller, G. B., Bonhoeffer, T., & Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron*, *74*(5), 809–815. https://doi.org/10.1016/j.neuron.2012.03.040

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Frontiers in systems neuroscience, 2, 4. https://doi.org/10.3389/neuro.06.004.2008

Millidge, B., Seth, A.K., & Buckley, C.L. (2021). Predictive Coding: a Theoretical and Experimental Review. ArXiv, abs/2107.12979. https://doi.org/10.48550/arXiv.2107.12979

Murray, J., Bernacchia, A., Freedman, D. et al. A hierarchy of intrinsic timescales across primate cortex. Nat Neurosci 17, 1661–1663 (2014). https://doi.org/10.1038/nn.3862

Nestvogel, D. B., & McCormick, D. A. (2022). Visual thalamocortical mechanisms of waking state-dependent activity and alpha oscillations. Neuron, 110(1), 120–138.e4. https://doi.org/10.1016/j.neuron.2021.10.005

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. PLoS computational biology, 10(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Latimer, K. W., Barbera, D., Sokoletsky, M., Awwad, B., Katz, Y., Nelken, I., Lampl, I., Fairhall, A. L., & Priebe, N. J. (2019). Multiple Timescales Account for Adaptive Responses across Sensory Cortices. The Journal of neuroscience : the official journal of the Society for Neuroscience, 39(50), 10019–10033. https://doi.org/10.1523/JNEUROSCI.1642-19.2019

Peterka, D. S., Imai, F., Ross, J. M., Bastos, G., Hornick, M., Rachmany, L., Gallimore, C. G., Hockley, A., & Hamm, J. P. (2026). Global context rapidly shapes sensory responses in V1. *bioRxiv : the preprint server for biology*, 2026.01.07.698143. https://doi.org/10.64898/2026.01.07.698143

Price, B. H., Jensen, C. M., Khoudary, A., & Gavornik, J. P. (2023). Expectation violations produce error signals in mouse V1. *Cerebral Cortex, 33*(13), 8803–8819. https://doi.org/10.1093/cercor/bhad163

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87. https://doi.org/10.1038/4580

Senzai, Y., Fernandez-Ruiz, A., & Buzsáki, G. (2019). Layer-Specific Physiological Features and Interlaminar Interactions in the Primary Visual Cortex of the Mouse. Neuron, 101(3), 500–513.e5. https://doi.org/10.1016/j.neuron.2018.12.009

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. Neuron, 61(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Solomon, S. S., Tang, H., Sussman, E., & Kohn, A. (2021). Limited evidence for sensory prediction error responses in visual cortex of macaques and humans. *Cerebral Cortex, 31*(6), 3136–3152. https://doi.org/10.1093/cercor/bhaa390

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 576(7786), 266-273.

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., … Harris, T. D. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science, 372*(6539), eabf4588. https://doi.org/10.1126/science.abf4588

Stuart, G. J., & Spruston, N. (2015). Dendritic integration: 60 years of progress. *Nature neuroscience*, 18(12), 1713–1721. https://doi.org/10.1038/nn.4157

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature neuroscience*, 26(5), 858–866. https://doi.org/10.1038/s41593-023-01304-9

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, 1464(1), 242–268. https://doi.org/10.1111/nyas.14321

Weber, A. I., & Fairhall, A. L. (2019). The role of adaptation in neural coding. Current opinion in neurobiology, 58, 135–140. https://doi.org/10.1016/j.conb.2019.09.013

Wyrick, D. G., Cain, N., Larsen, R. S., Lecoq, J., Valley, M., Ahmed, R., … Mazzucato, L. (2023). Differential encoding of temporal context and expectation under representational drift across hierarchically connected areas (Advance online publication). *bioRxiv*. https://doi.org/10.1101/2023.06.02.543483

Ye, Z., Bull, M. S., Li, A., Birman, D., Daigle, T. L., Tasic, B., … Steinmetz, N. A. (2023). Brain-wide topographic coordination of traveling spiral waves (Advance online publication). *bioRxiv*. https://doi.org/10.1101/2023.12.07.570517

Zhang, S., Xu, M., Chang, W. C., Ma, C., Hoang Do, J. P., Jeong, D., Lei, T., Fan, J. L., & Dan, Y. (2016). Organization of long-range inputs and outputs of frontal cortex for top-down control. Nature neuroscience, 19(12), 1733–1742. https://doi.org/10.1038/nn.4417

Zhang, Q., Grødem, S., Gracias, A., Lensjø, K. K., Fyhn, M., Stringer, C., & Pachitariu, M. (2025). Spatial predictive coding in visual cortical neurons. *bioRxiv*, 2025-09.

Methods

Animals

The experiments were conducted in adherence to the US National Institutes of Health guidelines for animal research and were approved by the Institutional Animal Care and Use Committee at the University of Washington. Six C57BL/6J mice (1 female, 5 males), aged two to three months, were utilized for the experiments. After undergoing head plate implant surgery, the animals were housed individually and maintained on a 12-hour dark/light cycle, with all experiments conducted during the light cycle. Mice were put on water restriction five days after head-plate surgery, and their weights were carefully monitored to ensure they remained above 85% of their baseline body weight.

Surgery

Head Plate Implant: The procedure for implanting the head plate, as described in Ye et al. (2023), involved inducing anesthesia with 5% isoflurane, which was then maintained at 2-3%. Briefly, Carprofen (5mg/kg) and Lidocaine (2mg/kg) were administered for analgesia. The scalp skin and underlying connective tissue were removed to implant a titanium head plate and a 3D-printed chamber using Metabond (Parkell). Carprofen was administered at 0.05 mg/ml in the water post-surgery for three days.

Craniotomy: One or two days before recordings, a 2 x 2 mm craniotomy was performed over the primary visual cortex using a dental drill and stereotaxic techniques. The craniotomy after random exposure was opened either at the right side or left side and another craniotomy was opened after sequence exposure either on the opposite side. The order of craniotomies (right vs left) was counterbalanced between animals. The area was covered with silicone gel. This process used similar anesthesia as the head plate implant, and Carprofen (5mg/kg) was given for analgesia. A cap was placed on the chamber to prevent dirt from entering the craniotomy area.

Experiments

Stimulus: Visual stimuli were presented on three 60 Hz screens (LG LP097QX1) surrounding the mice and covering a 270 ° × 70° visual angle. Each trial featured a natural image from a set of 20, sourced from the Allen Institute Brain Observatory, and repeated across all screens. Images A, B, C, and the surprising image D were selected from pilot recordings (N=2) for having the most distinct neural representations while eliciting similar overall response levels. For doing so the sum of euclidean distance between neural representations of each pair was maximized while the difference between average population PSTHs were minimized. The identities of images were fixed across all animals and sessions, so the representation space would not vary. These images were equiluminant and underwent histogram equalization to normalize contrast. Furthermore, the images were warped to appear rectilinear from the animals' viewpoint. Each image was displayed for 50 ms, followed by 100 ms of grayscreen, and 5 images were presented within a trial unless an image was omitted or only image A,B,C, or D was shown. A reward of 5 µL of 10% sucrose water was administered at 1100 ms after the first image to maintain engagement. The inter-trial interval followed an exponential distribution with a minimum of 2 and a mean of 2.6 seconds.

Experimental Flow

Habituation: After two days of water restriction, animals were habituated to head fixation for two days. Subsequently, they went through two more days of habituation to recording rigs after the first craniotomy, during which they received up to 30 random 5 µL sucrose-water (10%) rewards.

Training: After first habituation, animals were exposed to a series of 5-image sequences randomly chosen from among a set of 20 natural images across 1000 trials, followed by a 1.5 µL sucrose-water reward after 1100 ms of first image for 5 days (Random Exposure, RE). Moreover, after the first 2 days of recording, they were exposed to 1000 trials of a fixed sequence, ABBCC, for 5 days over a 5-7-day period (Sequence exposure, SE).

Recording day experiments. Following either Random Exposure (RE) or Sequence Exposure (SE), animals underwent recording sessions in the primary visual cortex. Following RE, animals additionally underwent craniotomy and habituation to the recording rig prior to recordings; by SE recordings, animals were already habituated to the apparatus. On each recording day, visual receptive fields were first mapped for 3 min using an asynchronous sparse-noise stimulus consisting of black and white checkerboard elements presented on a 10 × 36 spatial grid. Grid locations were updated independently at 60 Hz using sparse probabilistic sampling, and individual checkerboard elements persisted for 10 frames (~167 ms) before returning to baseline.

Following receptive-field mapping, animals were presented with visual image sequences organized into interleaved control and main blocks (control–main–control–main–control; Fig. 1b), separated by at least 30 s of rest. Images were displayed for 50 ms followed by a 100 ms gray-screen interval. All trials ended with delivery of 10% sucrose water reward 500 ms after onset of the final image, and inter-trial intervals were drawn from an exponential distribution with a minimum of 2 s.

The full set of experimental conditions and trial counts is summarized in Table S1. Each control block contained 200 trials designed to preserve stimulus identities and temporal positions while disrupting stable sequence structure (Fig. 1c). Forty trials consisted of isolated presentations of natural images A, B, C, or D (10 trials per image), with reward timing matched to the sequence task. The remaining 160 trials embedded A, B, C, or D at positions 2–5 of otherwise random five-image sequences (10 repetitions per image-position combination). In these sequences, all non-target images were random natural images (X) drawn from a pool of 20 stimuli excluding A, B, C, and D (e.g., XAXXX, XXBXX, XBXXX). These control conditions preserved lower-order sensory history and sequence timing while eliminating stable higher-order sequence predictions.

Each main block contained 460 trials consisting primarily of the standard ABBCC sequence (200 trials per block), interleaved with probe conditions designed to perturb learned sequence expectations (Fig. 1c). Probe conditions included omission trials, in which one expected sequence element was omitted (AOBCC, ABOCC, ABBOC, ABBCO), an omission-control condition in which the initial A stimulus was replaced by a random image X while omitting the expected second element (XOBCC), deviant substitution trials, in which an unexpected image D replaced an expected sequence element (ADBCC, ABDCC, ABBDC, ABBCD), and familiar-stimulus substitutions, in which an additional B or C image appeared at an unexpected position (ACBCC, ABCCC, ABBBC, ABBCB). Each probe condition occurred 20 times per main block (40 times per recording session across the two main blocks), whereas standard ABBCC trials occurred 400 times per recording session.

Electrophysiology

Mice were head-fixed during recordings while seated in a plastic support with their forepaws resting on a wheel (Burgess et al., 2017). This setup permitted movement of body parts other than the head but prevented locomotion. Behavioral state was monitored using two Basler acA2440-75um cameras (560 × 560 resolution): one focused on the eye to monitor pupil diameter and the other on the face to detect licking and ensure the animal was not in distress. Animals remained awake throughout all experiments.

Recordings were obtained from all layers of primary visual cortex using Neuropixels 2.0 probes (Steinmetz et al., 2021) sampled at 30 kHz. Probe insertion sites were targeted using the Pinpoint system (Birman et al., 2023) based on stereotaxic coordinates in the Allen Common Coordinate Framework (CCF). The recording reference was placed at the probe tip. Following each exposure condition, two recording sessions were attempted. If a session failed because of technical issues, the remainder of the experimental protocol was completed as planned.

In total, 15 recording sessions were attempted across 6 animals. Two sessions were excluded prior to analysis: one because neural responses did not reliably discriminate visual stimuli, preventing assessment of sequence-related representations, and one due to technical failure of stimulus timing measurements (photodiode failure), which precluded reliable alignment of neural activity to stimulus onset. The remaining dataset comprised 13 recording sessions from 6 animals and was used for all primary analyses unless otherwise stated. Of these sessions, 5 followed random exposure and 8 followed sequence exposure. Two animals did not contribute a random-exposure recording session, whereas all animals contributed at least one sequence-exposure session.

Spike sorting was performed separately for each of the four probe shanks using Kilosort 2.5 (Steinmetz et al., 2021; RRID: SCR_016422). Quality-control metrics were computed following criteria described by the International Brain Laboratory et al. (2022) and Roth et al. (2026). Units were included if median spike amplitude exceeded 50 μV and the lower-quantile mean amplitude remained within 5 standard deviations of the overall mean amplitude. Also, units with less than average firing rate of 0.1 sp/s during trials (main and control blocks) were excluded. Units with estimated contamination below 10%, assessed using the sliding-window method, were classified as single units; remaining included units were classified as multi-units. Session-by-session and layer-by-layer counts are reported in Table S2.

Session trimming and trial exclusion

To minimize the influence of recording drift and stimulus-alignment issues occurring near the end of some recording sessions, a subset of sessions was trimmed prior to analysis. Trimming criteria were determined based on data quality (e.g., probe drift, loss of stable units, or photodiode alignment problems) and applied before statistical analyses.

For main-block analyses, 4 of 13 sessions required trimming, with retained trial counts reduced to 460 trials in three sessions and 400 trials in one session. For control-block analyses,

5 of 13 sessions required trimming, including the same four sessions and one additional session with reduced control-block quality; depending on the session, between 200 and 400 control-block trials were retained.

Analyses requiring stable matched trial counts across conditions excluded sessions with insufficient post-trimming data. In particular, prediction-content decoding analyses (Figs. 6 and 8), which required matched control and main-block comparisons, were restricted to sessions with sufficient stable trials following trimming. Early-versus-late analyses of omission and deviant responses likewise excluded sessions with insufficient retained trials after quality control.

Analysis

Pupil size extraction

Pupil size was extracted using the Lightning Pose package (Biderman et al., 2024). Eye videos from two recording sessions were unavailable due to technical issues during acquisition and were excluded from pupil analyses. To train the model, we manually labeled 100 frames sampled from 5 of the 12 available videos (20 frames per video), marking the top, bottom, left, and right edges of the pupil. The selected videos represented different animals to capture between-subject variability in eye appearance, and frames were chosen to span a range of pupil sizes and gaze directions. Because the task involved relatively constrained eye movements and pupil geometry, a supervised pose-estimation approach was used. Model performance was assessed through manual inspection of labeled videos.

After inference, pupil size was estimated for frames in which all four tracked points were detected with confidence >80% (approximately 99% of frames). The short and long pupil axes were computed from the tracked points, and pupil area was estimated as the product of these axes, yielding a measure proportional to the area of an ellipse.

For trial-aligned analyses, pupil traces were normalized relative to the pre-trial baseline period (−0.2 to 0 s) by z-scoring each trial using the mean and standard deviation of its baseline window. This normalization facilitated comparison across sessions and conditions while reducing the influence of known nonlinear relationships between baseline pupil size and task-evoked pupil responses, which can complicate interpretation of absolute pupil changes.

Lick Extraction

For each video, an area around the lick spout was selected by visual inspection. Because the tongue is very bright, a peak-detection algorithm was used to detect brightness changes in the selected area. The minimum peak height was set to the mean brightness, the minimum distance to 85 ms, and the minimum prominence to 25% of the difference between the maximum and average brightness. Moreover, another area anterior to the nose was chosen as animals occasionally started grooming, and as their hands crossed the area around the spout, false licks were detected. If the brightness in the area anterior to the nose peaked, indicating grooming at the proximity of the time ($\pm$85 ms) of licks that were detected, those licks were excluded.

Layer Identification

Cortical layers were identified using the amplitude view of raw data across different shanks, inspecting spike waveforms in raw data files with SpikeGLX 4.1.3, drift maps, and analyzing changes in amplitude along the depth of the probes. Layer 5/6 was identified by its prominent large amplitude spikes (Senzai et al., 2019).

Population-level averages

Population-level responses were computed using trial-averaged peri-stimulus time histograms (PSTHs). Spike counts were binned (10 ms bins) and converted to firing rates (spikes/s). For each unit, PSTHs were first averaged across trials.

For population analyses, PSTHs were then averaged across units within each session.All units were included irrespective of isolation quality to avoid biasing population estimates toward high-firing or well-isolated neurons.

Two types of population responses were computed: (1) pooled population responses, obtained by averaging across all units (multi+single) within a session regardless of layer, and (2) layer-specific population responses, obtained by averaging across units assigned to layers 2/3, 4, or 5/6.

Population traces shown in figures represent the mean across sessions, with shaded regions indicating the standard error of the mean (SEM).

Estimation of Trial-by-trial Amplitude of Response

To compute the amplitude of omission and oscillation responses, let the trial-averaged PSTH for response type (r) in session (s) over the corresponding 150 ms window be $\mathbf{p_{r,s}} \in \mathbb{R}^{\mathrm{T}}$, where T is the number of time bins. The PSTH was normalized to obtain a unit vector:

$$\widehat{\mathbf{p}_{r,s}} = \frac{\mathbf{p_{r,s}}}{||\mathbf{p_{r,s}}||}$$

For each trial i, let the PSTH response vector within the same time window be $\mathbf{x_{i,s}}$. The response amplitude was computed by projecting the trial response onto the unit vector:

$$a_{i,r,s} = \mathbf{x_{i,s}^{\mathrm{T}}}\widehat{\mathbf{p}_{r,s}}.$$

To express amplitudes in units of spikes/s, the projection was rescaled by the dynamic range of the template PSTH

$$A_{i,r,s} = a_{i,r,s}\left[max\left(\widehat{\mathbf{p}_{r,s}}\right) - min\left(\widehat{\mathbf{p}_{r,s}}\right)\right].$$

To compute the amplitude of omission and oscillation responses, we formed unit vectors from the trial-averaged Peristimulus Time Histogram (PSTH) for each response type over the corresponding 150 ms window in each session, and projected each trial's response onto those unit vectors. Furthermore, to translate those amplitudes back to changes in units of spikes/s, we multiplied the result by the maximum minus the minimum of the unit vector.

To compute the relative amplitudes, we followed the same strategy and subsequently divided the amplitude of responses (to omissions) in each trial by the average response amplitude to other elements (B/C) in the same trial. In this way, we accounted for systematic changes in excitability during the session due to state changes or signal loss (damage to neural tissue/drift) in order to reveal relative changes in response to unexpected elements.

Cluster-based permutation test

Time-resolved differences between conditions were assessed using a **cluster-based permutation test**:

- Test statistic (paired t across sessions) was computed at each time bin
- Contiguous time bins exceeding a predefined threshold ($P<0.05$) formed clusters
- Cluster mass was defined as the sum of test statistics within a cluster
- A null distribution was generated by randomly permuting condition labels across sessions in a paired manner
- The observed cluster mass was compared against this null distribution in a one sided manner testing if surprising condition has a higher mean than non-surprising one

Reported values include:

- **P cluster**: cluster-level significance
- **d**: effect size (Cohen’s d across sessions)

Significant clusters are indicated by gray bars below the traces.

Population decoding along a shared prediction-context axis

For all population decoding analyses, neural activity was averaged within the 150-ms window corresponding to the element of interest (e.g., the unexpected stimulus in deviant trials or the omitted element in omission trials), and a trial × unit response matrix was constructed for each session. Both single- and multi-units were included. Analyses used regularized linear discriminant analysis (LDA) with fixed covariance regularization ($\Gamma = 0.1$) and a fixed number of retained principal components (50); the previously used Bayesian hyperparameter optimization was not employed, so that the same fixed pipeline applied identically to every condition and session.

The central analysis quantified whether the population representation of a sensory event depended on its prediction context, using a single shared decoding axis. For each session, the discriminant axis separating the two matched sequence elements (element 3 vs element 4) was trained on the deviant and omission conditions — the cases in which the physical sensory event is identical across the two contexts — so that the axis reflected prediction context rather than the sensory stimulus. Discriminability was quantified as the area under the ROC curve (AUC; 0.5 = chance), computed two-sided from the projections of held-out trials onto the axis (Mann–Whitney U statistic). Within each cross-validation fold, conditions were balanced to the smallest available trial count, the balanced pool was z-scored using label-free means and standard deviations, principal components were computed on the training trials, and the LDA axis was fit in that space and applied to held-out trials. Decoding used 5-fold cross-validation repeated 50 times, and the per-session AUC was the mean across folds and repetitions. Decoder orientation was standardized so that positive projections corresponded to the second

of the two elements. A per-session null was obtained by permuting element labels (200 permutations) and refitting the axis, yielding a shuffled AUC expectation centered at chance.

The familiar B and C stimuli at the two elements were not used to train the axis; their trials were projected onto the same deviant/omission-defined axis out of sample, so that their separation reflected alignment with the prediction-context axis rather than a separately optimized boundary. The pooled "All" measure concatenated the deviant, omission, B, and C trials before computing a single AUC on the shared axis.

To test whether separation along this axis could be explained by factors other than prediction, the same projection was applied to two sets of control conditions. Order/trial-time controls were the matched stimulus identities presented at the same elements of interleaved control sequences that lacked the learned ABBCC structure; longer-timescale (two-back) adaptation conditions tested whether the stimulus two elements earlier (A vs B) remained decodable, using control sequences in which A or B was followed by intervening elements and reading the response at the matched two-back delay. Both control sets were projected onto the prediction-context axis in the same way as B and C.

To confirm that low AUC for the controls and adaptation conditions on the prediction-context axis reflected an absence of decodable structure rather than an unfavourable axis, each of these conditions was additionally decoded on its own cross-validated axis using the identical pipeline (balance, label-free z-scoring, PCA, LDA with $\Gamma = 0.1$, 5-fold × 50 repetitions, label-permutation null). Each condition was decoded on its own axis without pooling trials across sub-conditions, because pooling distinct sub-conditions onto a single axis allows the boundary to exploit sub-condition identity and inflates AUC; where a condition combined two readout windows, each window was z-scored separately before being combined so that window identity could not contribute to decoding. The same own-axis analysis was applied to the two-back adaptation conditions at both the one-back and two-back readout delays to characterize the time course of the adaptation signal.

Session-level AUCs were the unit of replication. Effects were assessed with two-sided signed-rank tests of the per-session AUC against the shuffled null or of paired differences (e.g., a condition minus its matched control) against zero; confidence intervals were obtained by bootstrapping the across-session median (5,000 resamples). Differences between exposure groups (Random vs Sequence Exposure) were assessed with a two-sided rank-sum test and reported as exploratory given the small number of sessions per group. A primary set of nine sessions passing drift and trial-count quality control is reported as the main result, with the full thirteen-session set reported as a robustness check.

Representational similarity analysis of within-family geometry

To determine whether prediction content organizes population representations beyond the local comparison described in Figure 6, we asked whether it systematically shapes the representational geometry across sequence positions 2–5, where prediction, recent sensory adaptation, and sequence order can all contribute to the population response. Specifically, we asked whether positions sharing the same predicted stimulus are represented more similarly than positions predicting different stimuli while explicitly accounting for competing contributions from recent sensory adaptation and sequence order. We addressed this question using representational similarity analysis (RSA; Kriegeskorte et al., 2008; Kriegeskorte & Kievit, 2013). RSA characterizes a neural representation by the pattern of pairwise dissimilarities among conditions, summarized as a representational dissimilarity matrix (RDM), which describes the geometry of the representation independently of its orientation in neural state space.

For each recording session, we constructed an observed RDM describing the geometry of sequence positions 2–5 within each sequence family (deviant, omission, extra B, and extra C). We then asked whether this observed geometry could be explained as a weighted linear combination of three candidate geometries representing recent sensory adaptation, prediction content, and sequence order using multiple linear regression (Fig. S12 a-c). The adaptation and

sequence-order geometries were estimated empirically from interleaved control conditions designed to isolate these factors in the absence of stable sequence prediction. The adaptation geometry was computed from the single-stimulus control conditions (AOOOO, BOOOO, COOOO, and DOOOO) using responses to the stimulus presented at the second sequence position, thereby isolating stimulus-identity effects independent of sequence prediction. The sequence-order geometry was computed from the position-control conditions, in which the same stimulus was presented at sequence positions 2–5 within otherwise random image sequences, thereby isolating ordinal-position effects independent of stable sequence prediction. The prediction-content geometry was specified directly from the experimental design as a model RDM. The observed, adaptation, and sequence-order RDMs were constructed from cross-validated Mahalanobis (crossnobis) distances as described below, whereas the prediction-content RDM encoded whether each pair of conditions shared the same predicted stimulus (0 = shared prediction, 1 = different prediction).

Because the goal of the analysis was to explain the representational geometry observed during the familiar sequence, the empirical adaptation and sequence-order geometries were evaluated within the dimensions organizing that geometry. Specifically, we first identified a three-dimensional linear discriminant analysis (LDA) subspace using the familiar-sequence responses from positions 2–5. The discriminant identifies the directions that best separate these sequence positions without imposing any grouping corresponding to the prediction hypothesis tested by the regression model.

The observed, adaptation, and sequence-order geometries were all measured within this common representational subspace. Multiple regression therefore asked how much of the familiar-sequence geometry could be explained by the empirical adaptation and sequence-order components expressed along these dimensions, and whether prediction content accounted for additional structure beyond them. By expressing the empirical adaptation and sequence-order

geometries within the same discriminant subspace before regression, the analysis allows these competing factors to explain as much of the modeled geometry as possible before attributing any additional structure to prediction content, providing a conservative test of prediction-specific organization. To verify that the results did not depend on the discriminant projection, all principal analyses were repeated in the unsupervised 50-dimensional principal-component space, yielding nearly identical observed geometries. Component and observed geometries were visualized using classical multidimensional scaling of the square-root grand-average RDMs, and their shapes were compared across spaces by Procrustes alignment(Procrustes distance = 0.01; Fig. S13).

Cross-validated representational distances. Neural responses were quantified within the corresponding 150-ms analysis window for each condition. Population responses were projected onto the first 50 principal components computed from per-neuron z-scored responses pooled across all conditions entering the analysis. Pairwise representational dissimilarities were then estimated using the cross-validated squared Mahalanobis (crossnobis) distance (Nili et al., 2014; Walther et al., 2016; Diedrichsen & Kriegeskorte, 2017),

$$d_{ij}^{\ 2} = \ < (m_i^{\ A} - m_j^{\ A})\Sigma^{-1}(m_i^{\ B} - m_j^{\ B})^T >_{A \neq B},$$

where $m_i^{\ A} and\ m_i^{\ B}$ denote the mean population responses for condition i estimated from independent trial partitions, and Σ is the pooled noise covariance matrix estimated using a Ledoit-Wolf shrinkage estimator. Because the crossnobis estimator is unbiased, conditions with identical underlying representations have an expected distance of zero. Each empirical RDM was obtained by averaging crossnobis estimates across 20 random cross-validation partitions before subsequent analyses.

Additive decomposition. Each RDM was symmetrized and reduced to its six unique off-diagonal entries (Fig. S12b). Within each sequence family, the observed dissimilarities were mean-centered, and both the observed and model vectors were standardized to unit variance.

The standardized vectors from all four sequence families were concatenated and fit simultaneously by multiple linear regression (Fig. S12c),

$$D^2_{obs} = \beta_A D^2_A + \beta_P D^2_P + \beta_O D^2_O + \varepsilon,$$

where D^2_A, D^2_P, and D^2_O denote the adaptation, prediction-content, and sequence-order squared-distance geometries, respectively.

The additive model was formulated in squared-distance space. Cross-validated Mahalanobis (crossnobis) distances provide an unbiased estimate of squared representational distance, such that conditions with identical underlying representations have an expected distance of zero, yielding a meaningful model intercept. More fundamentally, squared representational distances are the natural quantity for additive decomposition: when multiple underlying representational components contribute linearly to a neural representation, the squared distance between two representations equals the sum of the squared component displacements plus interaction terms. Thus, if the underlying components are orthogonal, squared distances combine additively exactly; when interactions are weak, the additive model provides a close approximation. An explicit adaptation-by-prediction interaction geometry explained negligible additional variance and was therefore omitted from the final model. Predictor collinearity was low (variance inflation factors < 1.6).

Statistical inference and robustness. The regression was fit independently for each recording session, and inference was performed across the resulting nine session-level coefficients (two-sided Wilcoxon signed-rank test; bootstrap confidence intervals across sessions). The consistency of each component across sequence families was summarized by the range (maximum minus minimum) of the four per-family coefficients (the geometric parallelism of the prediction axis is quantified separately, below). Model fit was quantified by adjusting relative to a split-half noise ceiling, computed by splitting trials within each condition, correlating the resulting within-family half-RDMs, and applying the Spearman-Brown correction (Nili et al., 2014).

To assess robustness, the entire analysis was repeated in the unsupervised 50-dimensional principal-component space. The regression was also repeated after additionally standardizing each family's observed geometry to unit variance before fitting, thereby equalizing the contribution of each family (Fig. S14). Residual standard deviations were computed for each position pair in both the discriminant and unsupervised spaces to compare model specification (Fig. S14).

Parallelism and cross-family generalization of the prediction axis

To test whether predicted context enters the representation additively (a fixed, sensory-independent shift) or multiplicatively (a sensory-dependent change), we measured the parallelism of the prediction-coding direction across the four sequence families and its ability to generalize to held-out families. Population responses were projected onto the first 50 principal components (fit on all conditions) and then onto a three-dimensional linear discriminant subspace trained to separate the four sequence positions (refit independently for each session). Within each family, the prediction-content direction was defined as the difference between the centroids of elements 4 and 3, the pure prediction edge, where the two elements share the same preceding stimulus and ordinal position, differing only in predicted content. This centroid difference requires no within-class covariance estimate and is therefore stable at low trial counts. Within the discriminant subspace it was essentially collinear with the covariance-weighted (Fisher) discriminant between the two elements (mean cosine = 0.99), so the simpler centroid-difference estimator was used throughout.

Parallelism score. The parallelism score was defined as the mean pairwise cosine similarity among the four family-specific prediction directions. A rotational null distribution was generated by drawing, for each family, a random direction with matched norm and recomputing the mean pairwise cosine (1000 draws). Observed parallelism was compared with this null across sessions using a two-sided Wilcoxon signed-rank test.

Cross-family generalization (CCGP). Cross-condition generalization performance (CCGP) was used to test whether a prediction-coding axis learned from one subset of sequence families generalized to unseen families. A linear decision boundary separating elements 3 and 4 was trained on a subset of families and evaluated on the held-out families using the area under the receiver operating characteristic curve (AUC). Generalization was assessed by leaving out one family (train 3, test 1) and two families (train 2, test 2).

To prevent information leakage, the position-discriminant subspace was refitted using only the training families, and the held-out families were projected into this subspace before classification. Because the principal-component space is unsupervised, it was fit on all conditions without introducing predicted-context label information. Chance performance was estimated using a label-permutation null in which the labels of elements 3 and 4 were randomly permuted before refitting the decision boundary. Training the classifier directly in the unsupervised principal-component space, without refitting the discriminant subspace, produced equivalent results. For each evaluation, conditions were randomly subsampled to 40 trials, metrics were averaged across 20 independent subsamples, and inference was performed across the nine recording sessions using a two-sided Wilcoxon signed-rank test.

Alignment with sensory representations

To test whether the prediction axis coincided with the sensory representation, we computed the absolute cosine between the prediction axis and two control sensory directions: the direction separating the two predicted stimuli (B vs. C) and a control direction separating two other stimuli (A vs. D), each obtained from the single-stimulus control conditions in the same space. The prediction axis was estimated from trial-balanced conditions (each condition subsampled to a common trial count, averaged over draws); unbalanced pooling, in which high-count conditions dominate the discriminant, produces a spurious alignment with the sensory direction that balancing removes. Each alignment was referenced to a rotational null

(random directions of matched norm), and the two were compared directly by a paired signed-rank test across sessions. The result was unchanged across analysis dimensionality (20 and 50 principal components) and, in a layer-resolved analysis, did not differ across cortical depth.

Unit-by-unit ROC analysis

Changes in neuronal activity were quantified by the area under the ROC curve (AUC) for each unit (single- and multi-unit; multiunit AUC reflects a mixture and results are reported as "per-unit"). For each unit a null distribution was generated by permuting condition labels (1,000 iterations); two-sided p-values were computed against this null, and $|AUC-0.5|$ was taken as the deviation from chance, with the same statistic computed on the permuted data to give a null expectation. For visualization, AUC distributions were computed per session, stratified by cortical layer (L2/3, L4, L5/6), and plotted as the fraction of units per bin. Population-level effects were assessed per session and aggregated across sessions: (i) the mean $|AUC-0.5|$ across units was compared to its permutation null by a one-sided signed-rank test (observed > null); (ii) the fraction of units individually significant ($p < 0.05$) was compared to the 5% chance level by a one-sided sign test; (iii) the tail fraction $P(|AUC-0.5| \geq \tau)$ was compared to the pooled null, with a sign test on the per-session tail fraction; and (iv) mean $|AUC-0.5|$ was compared between exposure groups (RE vs SE) by a rank-sum test. All tests treated the session as the unit of replication.

Response-difference contrasts

The same per-unit AUC framework was applied to contrasts between two response windows rather than between trial types within a window: omission vs rebound (omission window vs the post-sequence window on the same trials), deviant vs matched control, and familiar vs extra (computed separately for each standard stimulus, B and C, using only the window that excludes same-stimulus repetition so the contrast is free of the directional response

reduction that repetition suppression would impose). For each unit, AUC and its 1,000-iteration label-permutation null were computed as above, and the population-level mean |AUC−0.5|, significant-unit fraction, and tail fraction were tested across sessions exactly as for the per-unit analysis (one-sided signed-rank, one-sided sign test, and tail-fraction sign test, respectively).

Because |AUC−0.5| is unsigned, it cannot distinguish a directional shift (units modulated consistently in one direction) from a bidirectional one (units split between increased and decreased discriminability, cancelling at the mean). To separate these, we additionally computed the signed per-session mean of (AUC−0.5) and tested it against the signed permutation null with a two-sided signed-rank test, and report the ratio |signed mean| / mean |AUC−0.5| (≈1 indicating a directional effect, ≈0 a balanced bidirectional one). The sign structure of contrasts that pool across positions (deviant, omission) is not interpreted, since position-dependent adaptation alone produces a near-balanced split; the signed test is interpretable for the familiar-vs-extra contrasts, where each side is a single unadapted window.

Decoder-based neuron ordering

To examine how individual neurons contribute to discriminating between the element-3 and element-4 conditions, we used a linear decoding approach based on linear discriminant analysis (LDA). A multiclass LDA classifier was trained to discriminate among all sequence-element conditions on odd trials. From this model, the decision boundary separating the element-3 and element-4 conditions was extracted, and the corresponding weight vector was used to quantify each neuron's contribution to this specific discrimination.

For each neuron, trial-averaged activity was computed separately for the element-3 and element-4 conditions on even trials. The difference between conditions (element 3 − element 4) was then calculated over time and z-scored to normalize response magnitude across neurons.

Neurons were sorted according to their weights along the element-3-vs-4 decision boundary, so that neurons with similar contributions to the classifier were grouped together.

Heatmaps were constructed by plotting the time-resolved element-3-minus-element-4 difference for each neuron in sorted order.

To assess the effect of dimensionality reduction, decoding was performed both with and without prior principal component analysis (PCA). For PCA-based analyses, neural activity was projected onto a lower-dimensional space prior to fitting the LDA model, and the resulting weight vectors were back-projected into the original neuron space to enable interpretation at the level of individual units.

Because decoder weights reflect both signal and noise structure, they are not directly interpretable as encoding weights. To obtain interpretable activation patterns, we applied the Haufe transform to the LDA weights. This transformation maps discriminative weights back into the data space by accounting for the covariance structure of neural activity, yielding patterns that more directly reflect how neural responses contribute to the encoded variable (i.e., the element-3-minus-element-4 difference) rather than how they are used by the classifier for discrimination; the shrinkage regularization applied to the LDA covariance estimate was accounted for in this transform.

Supplementary material

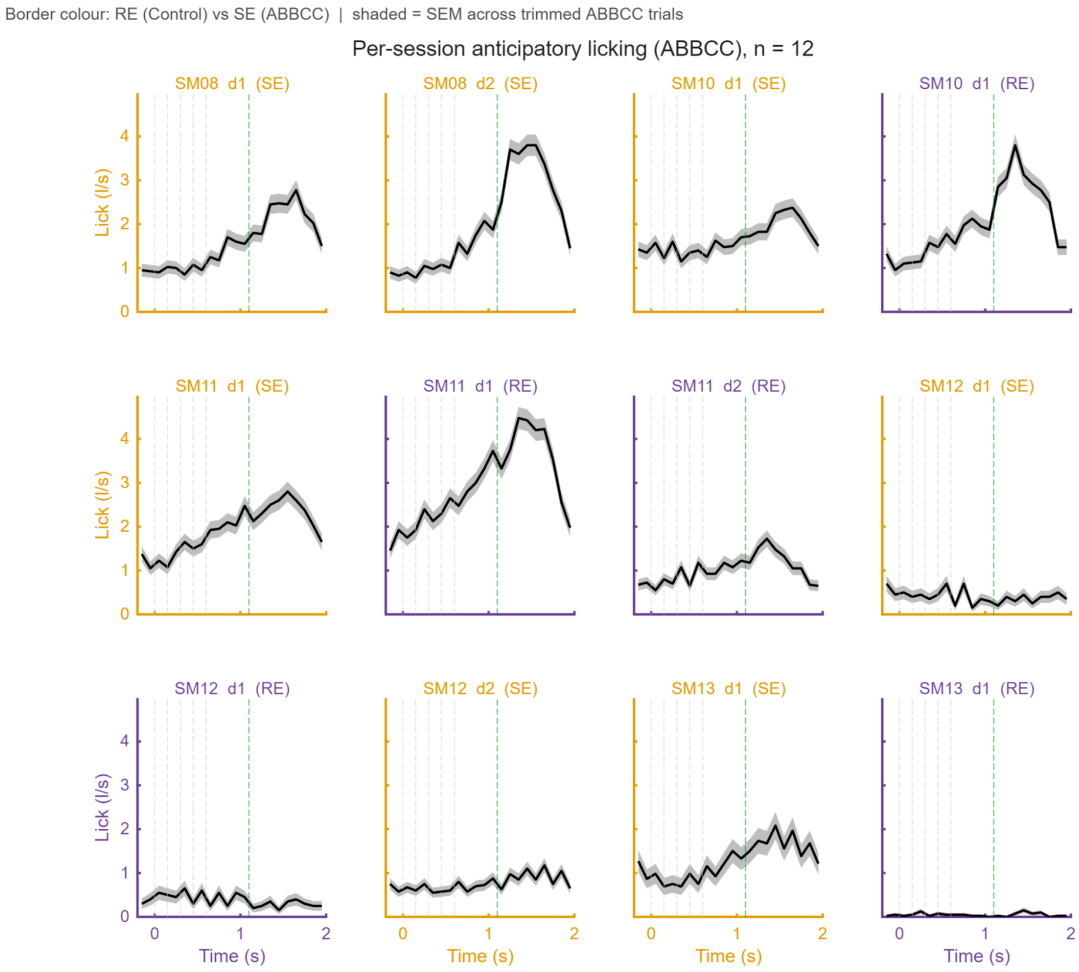


Figure S1: Session-by-session anticipatory licking and recording stability.
ABBCC lick PSTHs for individual recording sessions (mean ± SEM across trimmed trials), ordered by animal and recording day. Border and title colors indicate exposure condition (Random Exposure, RE, purple; Sequence Exposure, SE, gold). Most sessions show progressive anticipatory licking toward reward delivery, although several sessions exhibit weaker or minimal ramping. *n* = 12 sessions (SM_0009 excluded due to missing behavioral video).

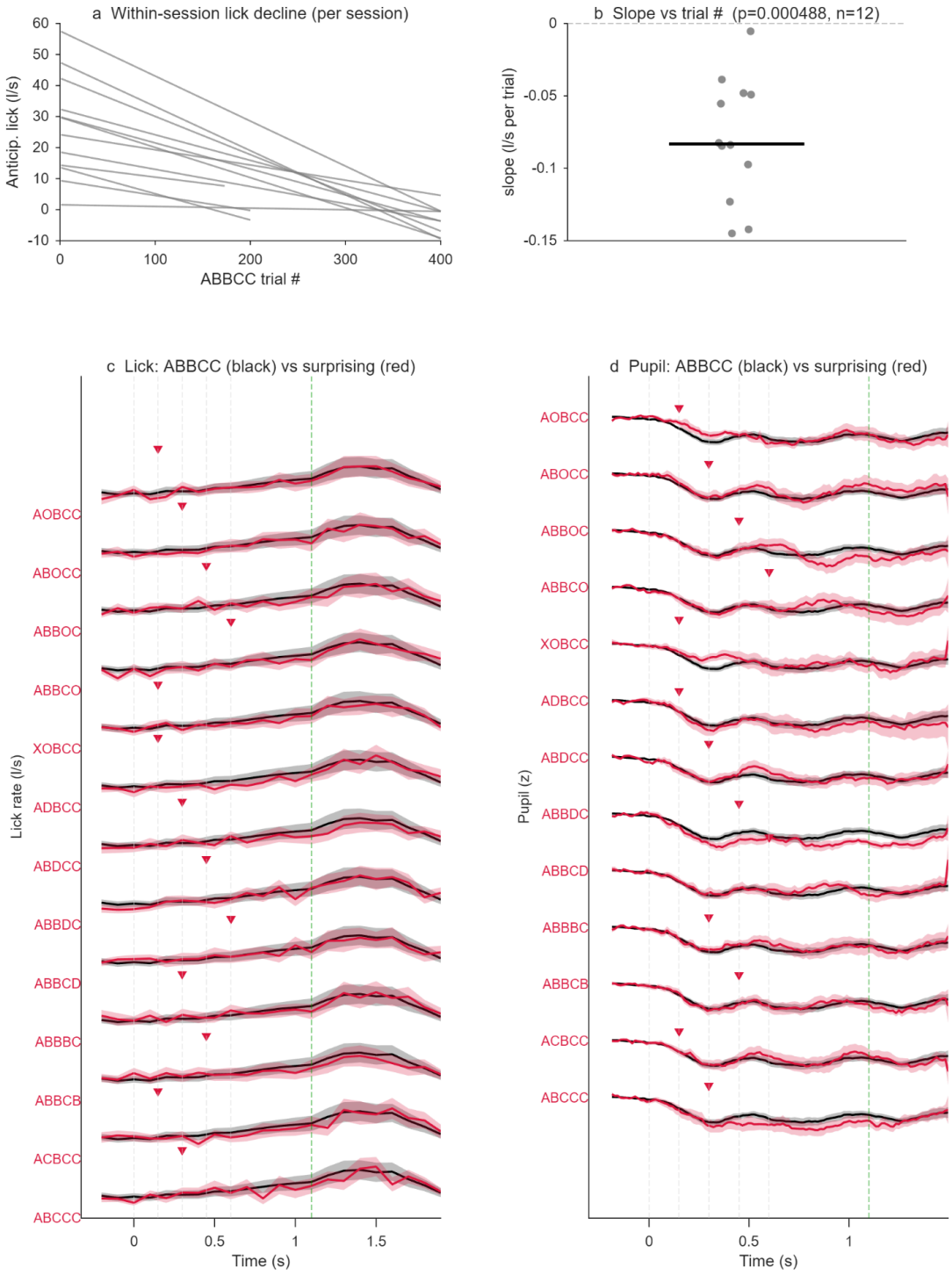


Figure S2: Within-session behavioral changes and condition-specific licking and pupil responses.
(a) Per-session regressions of anticipatory lick rate against ABBCC trial number, showing a gradual decline in anticipatory licking over the course of the session. Each line represents one recording session. **(b)** Session-wise slopes from (a), tested against zero (signed-rank test, $p = 5 \times 10^{-4}$, $n = 12$ sessions), consistent with progressive satiation during the session. Horizontal line indicates the median across sessions. **(c)** Lick responses for individual surprising trial types (red) overlaid on familiar ABBCC trials (black). Traces are vertically offset for visualization. Red triangles indicate the sequence position of the surprising event (omission or substitution) for each trial type. Shaded regions indicate SEM across trimmed trials. Lick dynamics were broadly similar across trial types. **(d)** Pupil responses for individual surprising trial types (red) overlaid on familiar ABBCC trials (black), plotted as in (c). Red triangles indicate the sequence position of the surprising event. Pupil dynamics were broadly similar across trial types and showed a small common constriction following sequence onset. $n = 12$ behavioral sessions for licking; pupil analyses include sessions with usable eye tracking ($n = 11$).

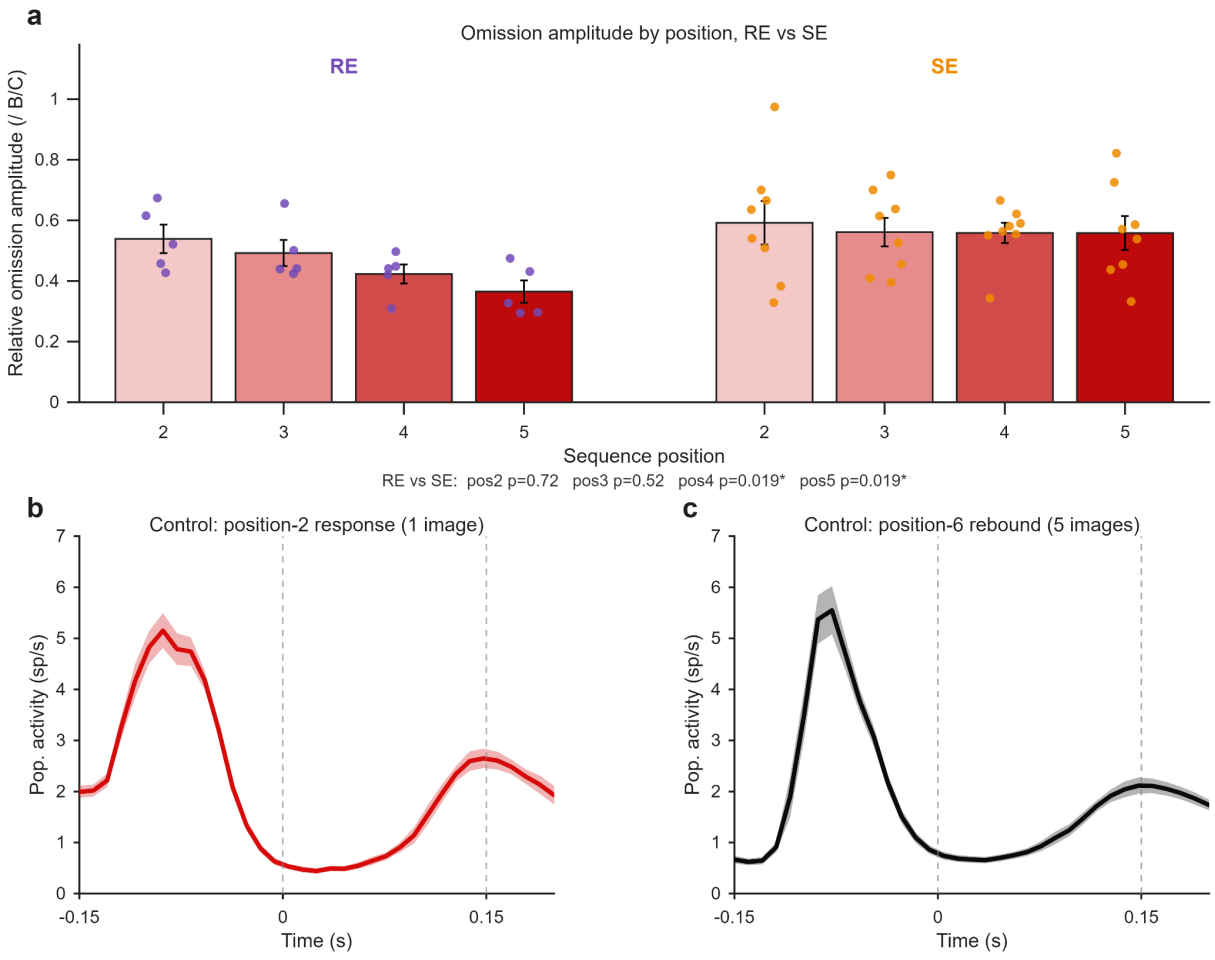


Figure S3: Omission amplitude by sequence position and exposure, and control-block omission-like and rebound responses.
(a) Relative omission amplitude (normalized to within-trial B/C responses) at each omission position of the ABBCC sequence (positions 2–5; AOBCC, ABOCC, ABBOC, ABBCO), shown separately for Random Exposure (RE, purple) and Sequence Exposure (SE, gold). Dots, individual sessions; bars and error bars, mean ± s.e.m. across sessions.

Under RE relative omission amplitude declined monotonically but nonesignificantly across positions (per-session slope, signed-rank p = 0.125, n = 5), whereas under SE it remained approximately constant; a session-level slope contrast did not support a formal Position × Exposure interaction (ranksum p = 0.22). **(b)** Population-averaged response in the control block following a single isolated stimulus, aligned to stimulus onset (0 s); the omission-like response occurs at the subsequent expected stimulus time (second dashed line, ~0.15 s). Shaded region, s.e.m. across sessions. **(c)** Population-averaged rebound following the final image of a five-image control sequence, aligned to the onset of the final image (0 s); the rebound occurs at the subsequent expected stimulus time. Shaded region, s.e.m. across sessions.

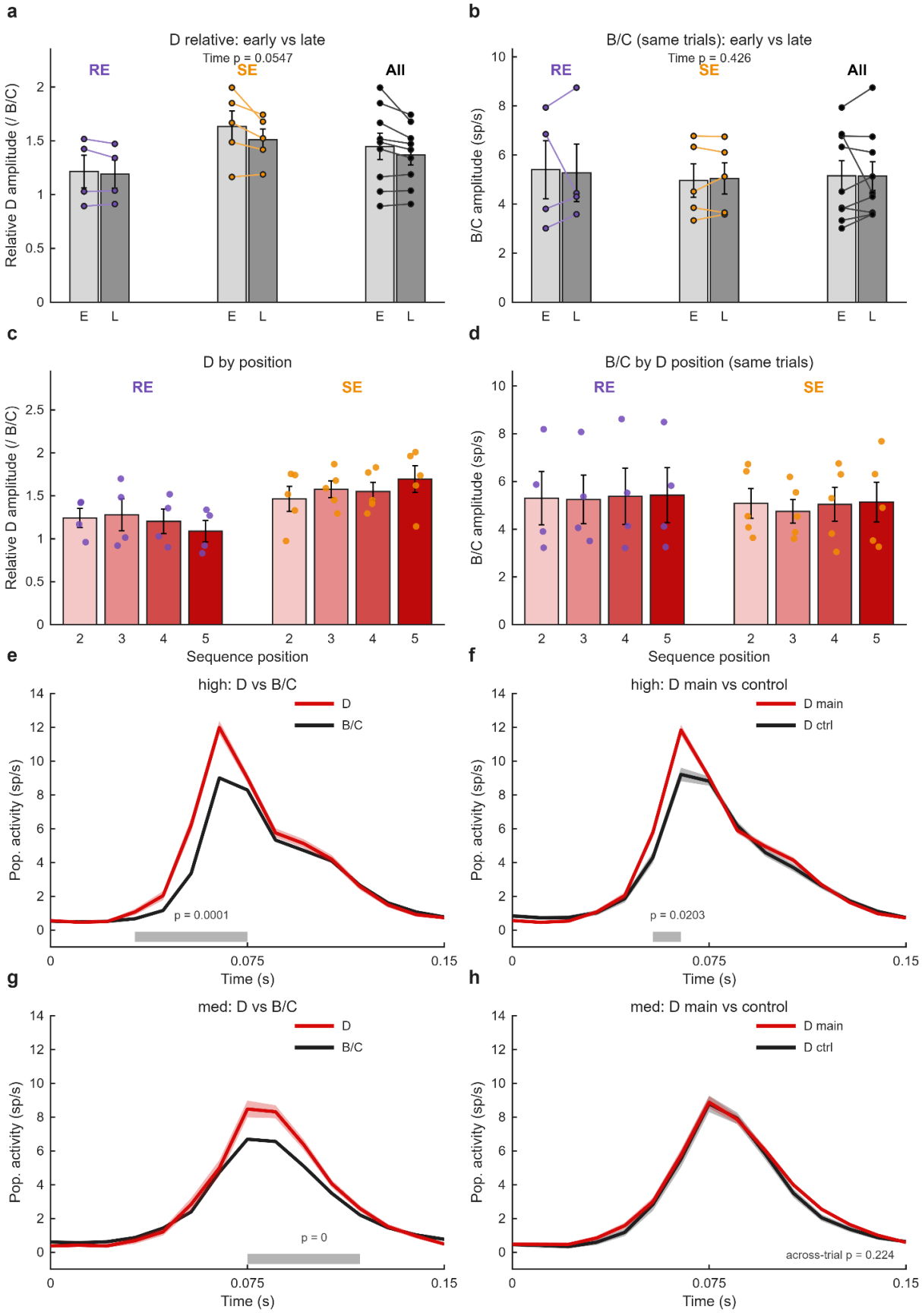

a
D relative: early vs late
Time p = 0.0547
RE
SE
All
Relative D amplitude (/ B/C)
E L E L E L
b
B/C (same trials): early vs late
Time p = 0.426
RE
SE
All
B/C amplitude (sp/s)
E L E L E L
c
D by position
RE
SE
Relative D amplitude (/ B/C)
Sequence position
d
B/C by D position (same trials)
RE
SE
B/C amplitude (sp/s)
Sequence position
e
high: D vs B/C
D
B/C
Pop. activity (sp/s)
p = 0.0001
Time (s)
f
high: D main vs control
D main
D ctrl
Pop. activity (sp/s)
p = 0.0203
Time (s)
g
med: D vs B/C
D
B/C
Pop. activity (sp/s)
p = 0
Time (s)
h
med: D main vs control
D main
D ctrl
Pop. activity (sp/s)
across-trial p = 0.224
Time (s)

Figure S4: Within-session stability, positional distribution, and example sessions for the deviant enhancement.
(a) Relative D amplitude (normalized to within-trial B/C) for early versus late trial halves, shown for Random Exposure (RE), Sequence Exposure (SE), and all sessions combined (sessions with intact trial counts, n = 9; see Methods). Lines connect halves within a session; bars and error bars, mean ± s.e.m. Relative D amplitude was stable across the session (Time $p = 0.055$). **(b)** Within-trial B/C amplitude (sp/s) on the same trials as (a), early versus late, formatted as in (a). The B/C normalizer was stable across the session (Time $p = 0.43$). **(c)** Relative D amplitude (normalized to within-trial B/C) at each deviant position (sequence positions 2–5), shown for RE and SE. Dots, individual sessions; bars and error bars, mean ± s.e.m. Relative D amplitude did not differ between exposure conditions at any position (rank-sum, all $p \geq 0.19$). **(d)** Within-trial B/C amplitude (sp/s) at the same deviant positions, on the same trials, formatted as in (c). The B/C normalizer was positionally invariant (rank-sum, all $p \geq 0.9$). **(e–h)** Two example sessions spanning the range of effect sizes (combined D-vs-B/C and D-main-vs-control score). **(e)** Example session with strong enhancement: D (red) versus B/C (black), aligned to stimulus onset. **(f)** Same session, D in main (red) versus control (black) block. **(g,h)** As in (e,f) for an example session with median enhancement. Shaded regions, s.e.m. across trials. Gray bars mark significant within-session, across-trial clusters (cluster-based permutation test on trials; descriptive, distinct from the session-level test in Fig. 4); the across-trial cluster p is shown.

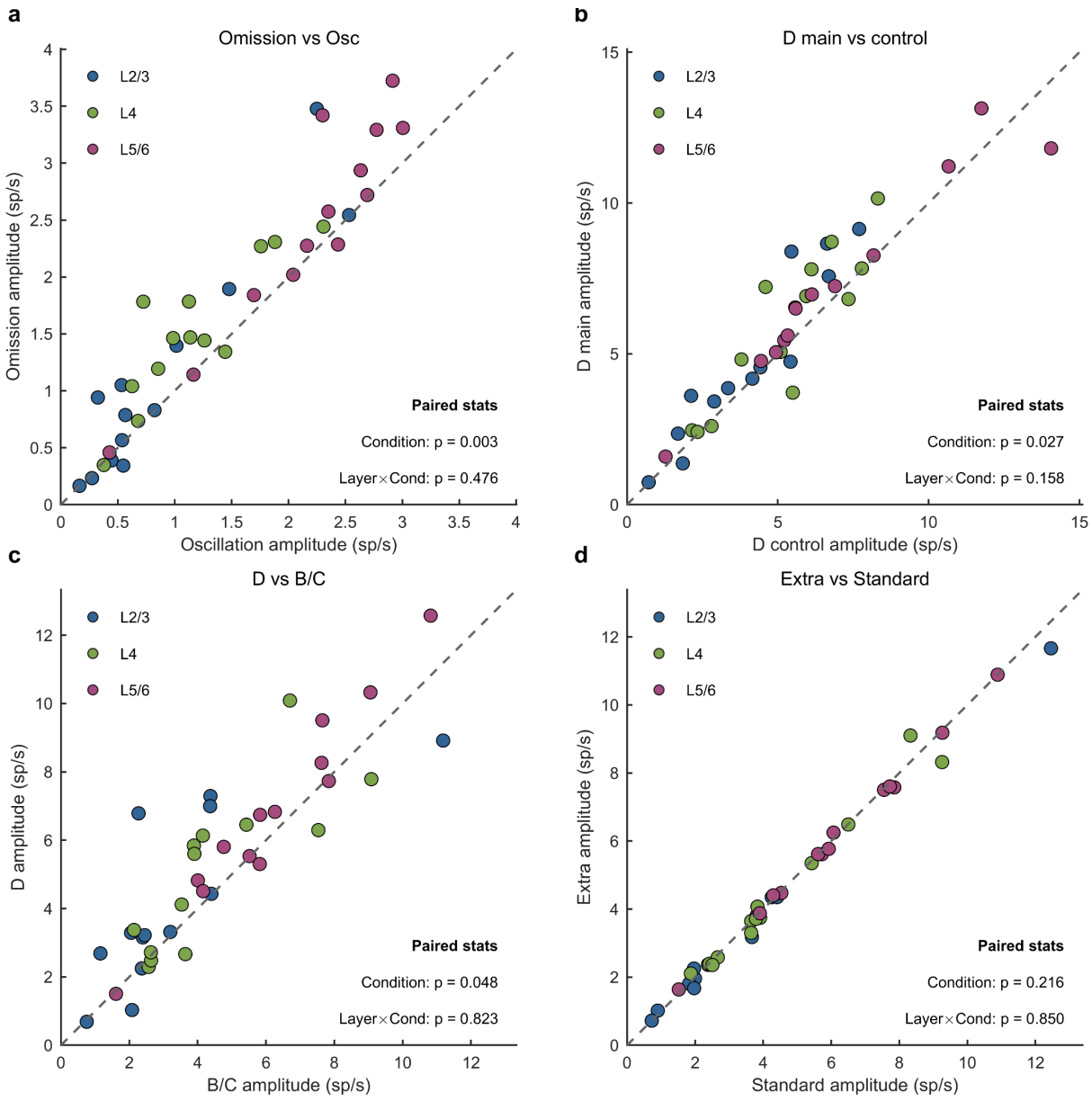


Figure S5: Mismatch-related response enhancement is broadly distributed across cortical layers.
Units were grouped into superficial (L2/3), middle (L4), and deep (L5/6) layers. Each point is one recording session within a layer (colors denote layer); dashed line, unity. Amplitudes were quantified by projecting the population response onto a pooled (both-condition) unit template, as in Fig. 4. Condition effects, signed-rank on session-averaged paired differences; layer dependence, linear mixed-effects model of the paired difference (Diff ~ Layer + (1|Session)). **(a)** Omission versus oscillation (rebound) amplitude. Omission exceeded rebound across layers (condition p = 0.0034) with no layer dependence (layer × condition p = 0.48). **(b)** D main versus D control. D was larger in the main block across layers (condition p = 0.027; layer × condition p = 0.16). **(c)** D versus B/C (main). D exceeded B/C across layers (condition p = 0.048; layer × condition p = 0.82). **(d)** Extra versus Standard. Extra and Standard did not differ (condition p = 0.22; layer × condition p = 0.85). In every comparison the modulation showed no significant layer dependence.

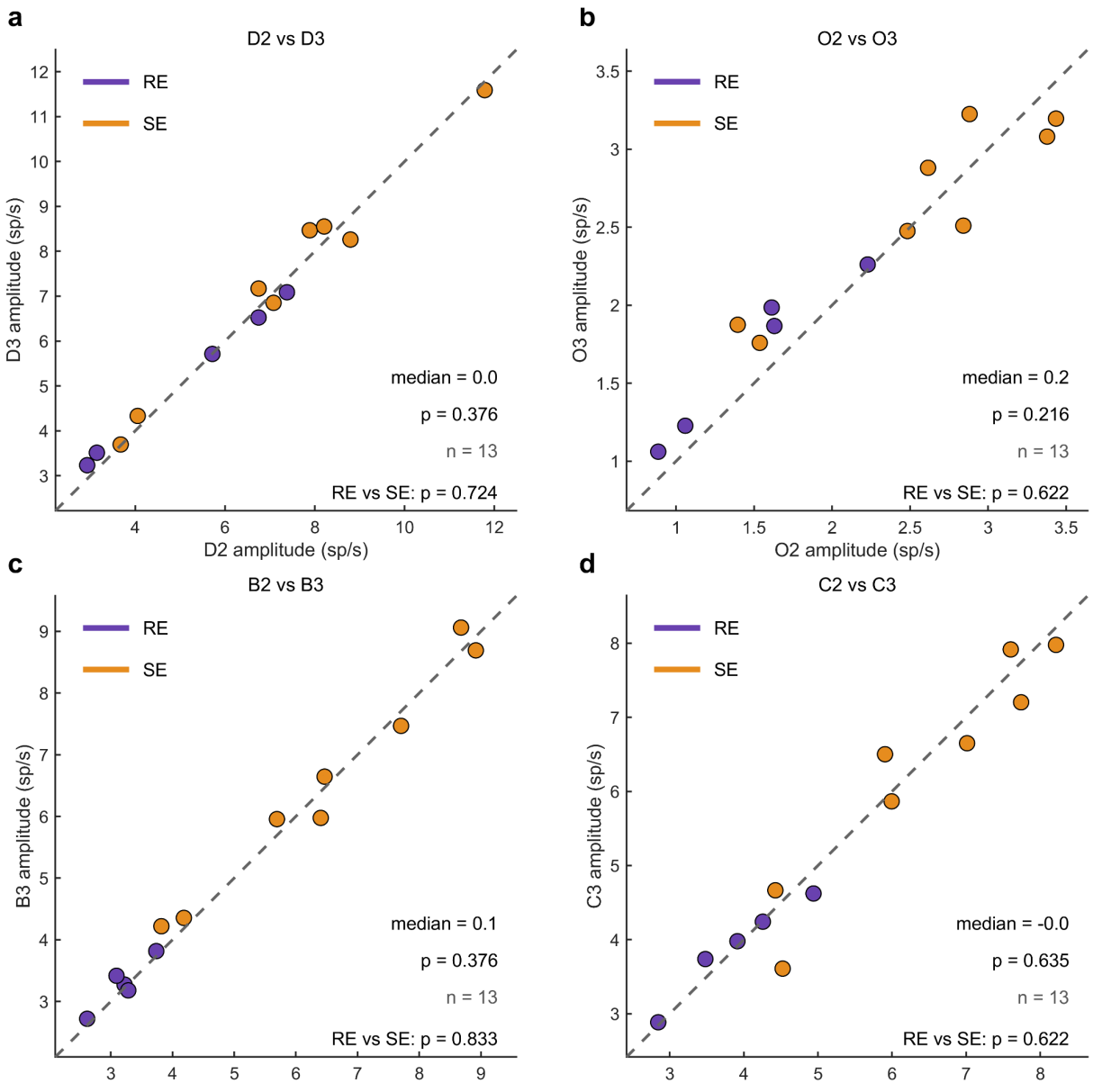

a
D2 vs D3
RE
SE
D3 amplitude (sp/s)
D2 amplitude (sp/s)
median = 0.0
p = 0.376
n = 13
RE vs SE: p = 0.724
b
O2 vs O3
RE
SE
O3 amplitude (sp/s)
O2 amplitude (sp/s)
median = 0.2
p = 0.216
n = 13
RE vs SE: p = 0.622
c
B2 vs B3
RE
SE
B3 amplitude (sp/s)
median = 0.1
p = 0.376
n = 13
RE vs SE: p = 0.833
d
C2 vs C3
RE
SE
C3 amplitude (sp/s)
median = -0.0
p = 0.635
n = 13
RE vs SE: p = 0.622

Figure S6:Response amplitudes are comparable across position-2 and position-3 prediction contexts.

Population response amplitudes measured within the analysis window for events occurring at positions 2 and 3 of the familiar sequence. Panels show comparisons for **(a)** deviant stimuli (D2 vs D3), **(b)** omissions (O2 vs O3), **(c)** B stimuli occurring at position 2 or 3 (B2 vs B3), and **(d)** C stimuli occurring at position 2 or 3 (C2 vs C3). For B and C comparisons, one condition corresponded to an expected familiar stimulus whereas the other corresponded to an unexpected extra familiar stimulus. Colored points indicate session means for Random Exposure (RE) and Sequence Exposure (SE) sessions; faint points show trial-level responses. Dashed lines indicate equality between position-2 and position-3 conditions. Across comparisons, response amplitudes did not systematically differ between positions 2 and 3 (all signed-rank $p > 0.05$), suggesting that overall mismatch magnitude and saliency were broadly comparable across prediction contexts. Because immediate sensory history was matched (both positions followed B in the familiar sequence), these comparisons provide a controlled test of whether prediction context alters neural population representations independent of gross differences in firing-rate magnitude.

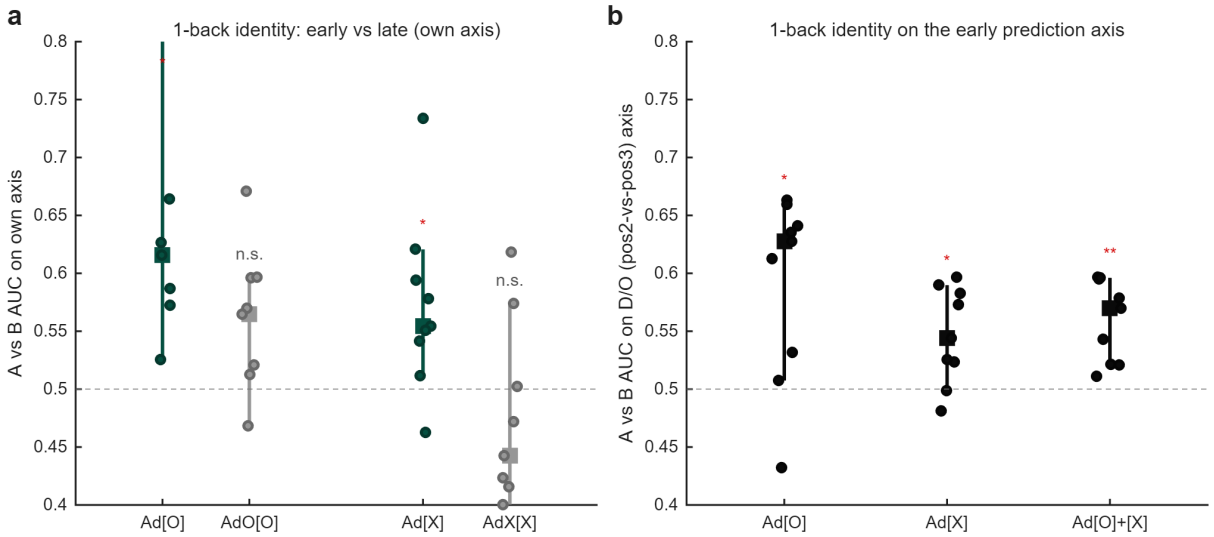


Figure S7:One-back stimulus identity is decodable in V1 and decays by the two-back delay.

Controls establishing that the two-back adaptation signal tested in Figure 6 has decayed by the readout delay rather than being undetectable for methodological reasons. Primary session set (n = 9). Throughout, bracket notation denotes the readout delay relative to the A/B stimulus: a single bracket (Ad[O], Ad[X]) is the one-back read, a doubled bracket (AdO[O], AdX[X]) the two-back read; [O] indicates the O-anchored sequences (AOOOO vs BOOOO) and [X] the X-embedded sequences (XAXXX + XXAXX vs XBXXX + XXBXX). **(a)** Stimulus identity (A vs B) decoded on its own cross-validated axis at the one-back (Ad[O], Ad[X]) and two-back (AdO[O], AdX[X]) delays. Identity is read at the element one or two positions after the A/B stimulus; the two windows of the X-embedded condition are z-scored separately before combining, so window identity cannot contribute to decoding. Points, individual sessions; squares, across-session median; bars, bootstrap 95% confidence interval of the median; dashed line, chance (AUC = 0.5). Identity is decodable at the one-back delay (Ad[O], median AUC = 0.616, $p = 0.039$; Ad[X], 0.555, $p = 0.012$) and falls to chance at the two-back delay (AdO[O], 0.565, $p = 0.20$; AdX[X], 0.443, $p = 0.30$; both n.s.). **(b)** The same one-back identity contrast projected onto the prediction axis decoded one window earlier (deviant/omission, element 2 vs element 3), for the O-anchored (Ad[O]), X-embedded (Ad[X]), and pooled (Ad[O]+[X]) conditions; conventions as in (a). One-back identity is decodable on this early prediction axis as well (Ad[O] = 0.628, $p = 0.020$; Ad[X] = 0.544, $p = 0.020$; Ad[O]+[X] = 0.570, $p = 0.004$). Together these show that the decoder reliably recovers stimulus identity when it is present in the population, so the chance-level two-back result in Figure 6i,l reflects genuine decay of the adaptation signal rather than insufficient sensitivity. Asterisks, $p < 0.05$; statistics are two-sided AUC with signed-rank tests against the within-session shuffled null.

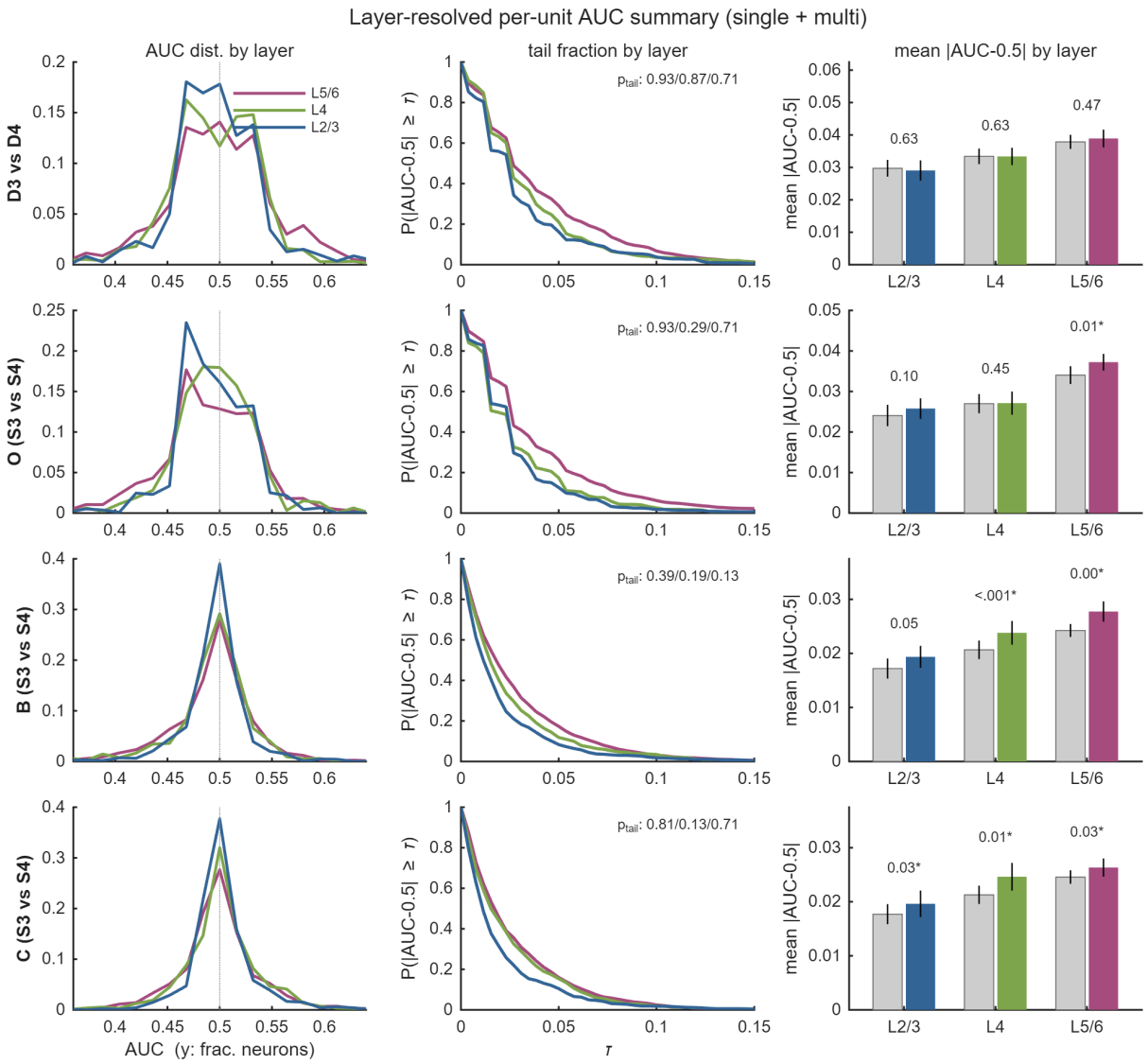


Figure S8: No laminar concentration of stimulus-position information.

The Fig. 7 contrasts resolved by cortical layer (L2/3 blue, L4 green, L5/6 magenta; displayed superficial → deep). Column 1, per-layer observed AUC distribution. Column 2, per-layer tail fraction; corner text, per-layer tail-fraction p (L2/3 / L4 / L5/6). Column 3, mean |AUC−0.5| per layer, null (grey) versus observed (layer colour), with per-layer one-sided signed-rank p. No layer shows a significant-unit excess or a heavy tail, and the superficial-vs-deep gap contrast is non-significant in every comparison. Scattered single-layer hits (e.g., C, L4) do not survive correction across the 12 per-layer tests. All units (single + multi).

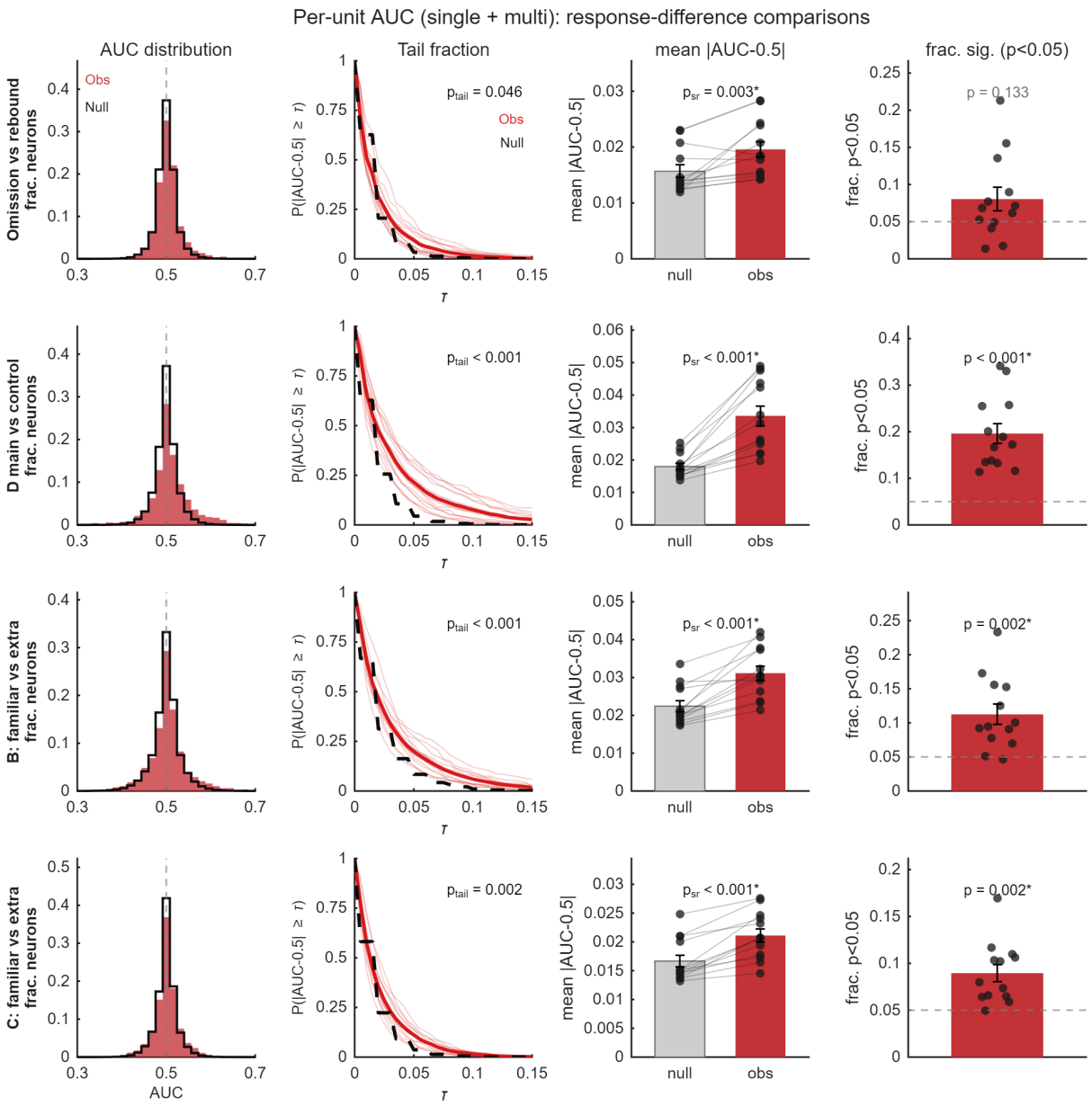


Figure S9: Response-difference contrasts show bidirectional, unit-level structure.

Per-unit AUC analysis as in Fig. 7, applied to contrasts where a genuine per-unit signal is expected (rows: omission vs rebound; deviant vs matched control; B familiar vs extra; C familiar vs extra). The familiar-vs-extra rows use only the window that excludes same-stimulus repetition (familiar = first B / first C), isolating predictive context from repetition suppression. Column 1, observed AUC distribution (red, session-averaged) with the pooled permutation null (black outline), dashed line = chance; this column is unfolded, so directional shifts are visible. Column 2, tail fraction P(|AUC−0.5| ≥ τ); p, tail-fraction sign-test. Column 3, mean |AUC−0.5|, null versus observed, per-session dots and connectors; p, one-sided signed-rank. Column 4, fraction of significant units (dashed line, 5% chance); p, one-sided sign-test. Deviant-vs-control is concentrated (heavy-tailed, large significant-unit excess) with a modest net enhancement; omission-vs-rebound is significant in aggregate without a significant-unit excess; the familiar-vs-extra contrasts are strongly discriminable yet have a signed per-unit mean near zero with an elevated folded mean, indicating balanced bidirectional, identity-specific modulation. The sign structure of the pooled deviant and omission contrasts is not interpreted (positions pooled). All units (single + multi); the single-unit-only version reproduces each effect.

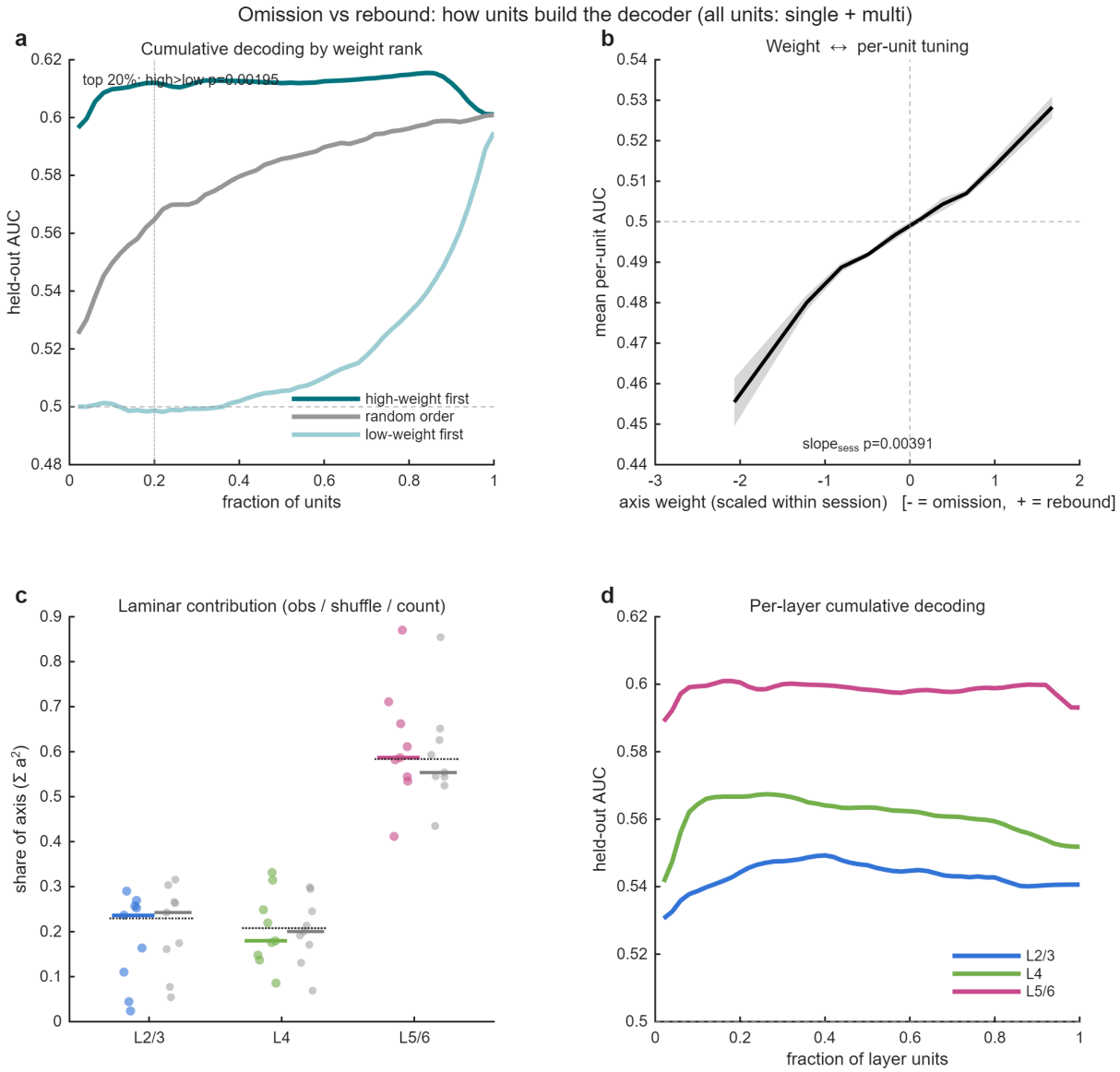


Figure S10: The deviance axis is concentrated in a tuned subset.
Decomposition of the deviant-vs-control decoding axis (all units, single + multi; 9 sessions), as in Figure 8. **(a)** Cumulative held-out decoding by axis-weight rank (high-weight-first saturated teal, random grey, low-weight-first muted teal; session means). Accuracy saturates within the top-weighted units: the top 20% high-weight-first reach AUC 0.711 versus 0.489 for the lowest-weighted (signed-rank $p = 0.0020$), against a full-population AUC of 0.721 (vs 0.5, $p = 0.0039$). **(b)** Per-unit held-out AUC versus axis weight (scaled within session; − = D control, + = D main); per-session Spearman ρ tested against zero across sessions (signed-rank, median ρ = +0.86, $p = 0.0039$). Weights from training trials, AUC from held-out trials (non-circular). **(c)** Per-layer axis share (Σa^2; layer colour) versus label-shuffle null (grey) and count reference (dotted); no layer exceeds its shuffle (all $p \geq 0.73$). **(d)** Per-layer cumulative decoding (layer means); deep-layer advantage tracks unit count, not laminar concentration (see c). Panels a and d show means only; statistics are in b and c. Contrast with the distributed content axis of Figure 8.

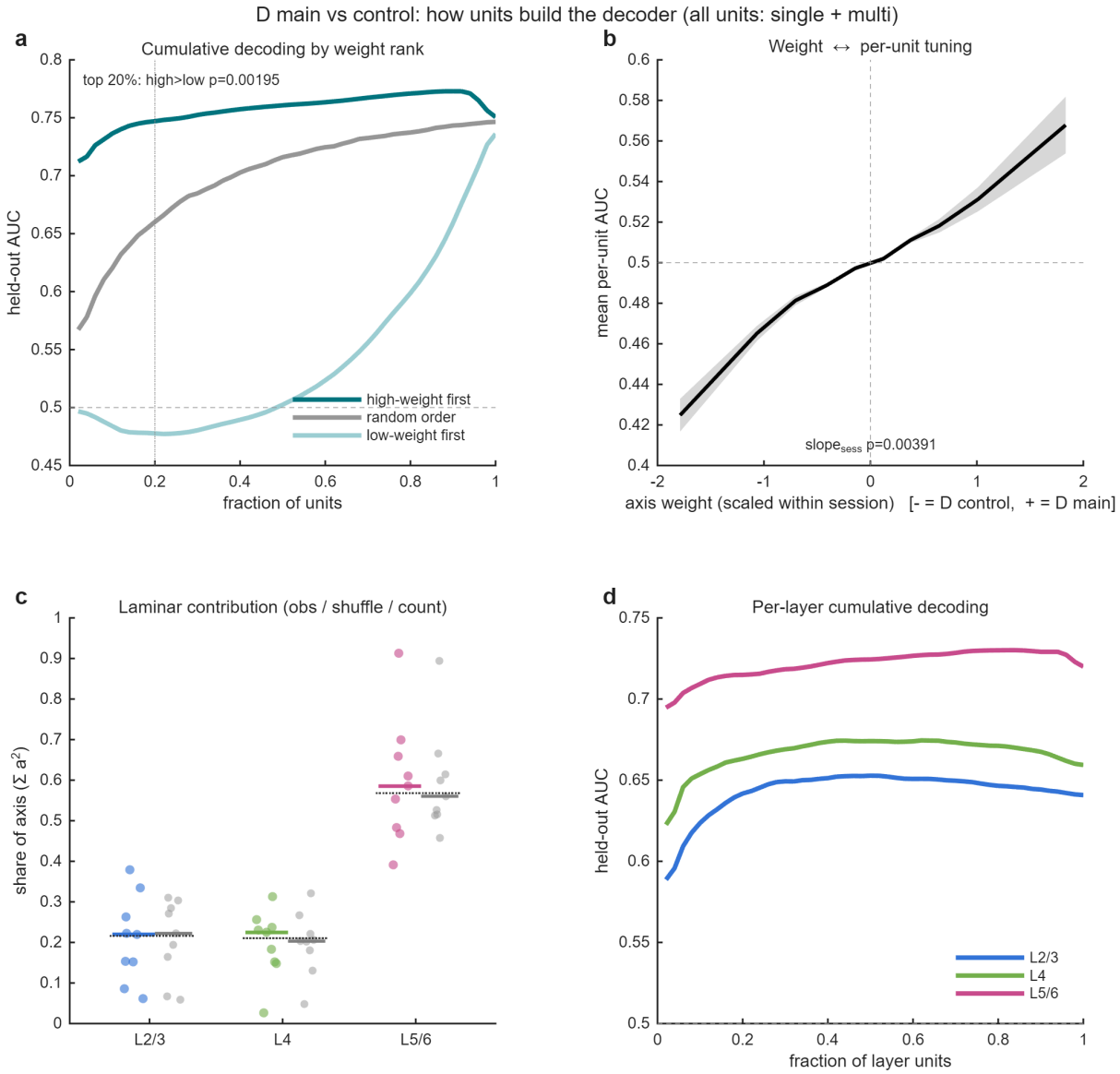


Figure S11: The omission axis is concentrated in a tuned subset.
Decomposition of the omission-vs-rebound decoding axis (all units, single + multi; 9 sessions), as in Figure 8. **(a)** Cumulative held-out decoding by axis-weight rank (conventions as in S10). Accuracy saturates early: the top 20% high-weight-first reach AUC 0.575 versus 0.499 for the lowest-weighted (signed-rank p = 0.0020), against a

full-population AUC of 0.575 (vs 0.5, p = 0.0039). **(b)** Per-unit held-out AUC versus axis weight (scaled within session; − = omission, + = rebound); per-session Spearman ρ tested against zero across sessions (signed-rank, median ρ = +0.79, p = 0.0039). Weights from training trials, AUC from held-out trials (non-circular). **(c)** Per-layer axis share (Σa²; layer colour) versus label-shuffle null (grey) and count reference (dotted). No layer shows meaningful laminar concentration; the one nominal exception (L2/3 p = 0.012) is negligible in magnitude (observed 0.24 vs shuffle 0.24) and does not survive correction across the nine layer tests, while L5/6 matches its shuffle (0.59 vs 0.55, p = 0.25). **(d)** Per-layer cumulative decoding (layer means); deep-layer advantage tracks unit count, not laminar concentration (see c). Panels a and d show means only; statistics are in b and c. Contrast with the distributed content axis of Figure 8.

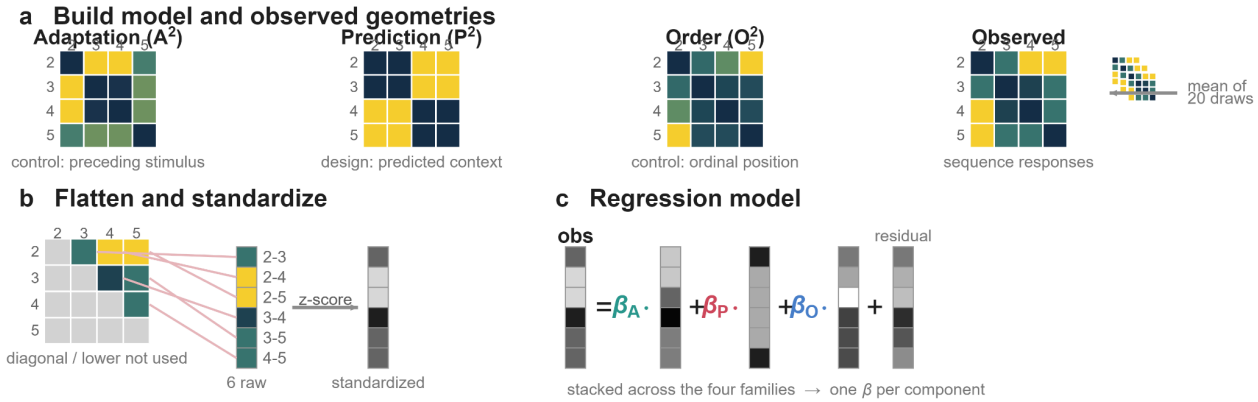


Figure S12: Decomposing the within-family representational geometry.
(**a**) The four component geometries entering the model, shown as 4 × 4 representational dissimilarity matrices (RDMs) over the within-family sequence positions (elements 2–5). Adaptation (A²) and order (O²) are empirical, computed as cross-validated (crossnobis) distances among independent control conditions and indexed onto the four positions by preceding-stimulus identity and by ordinal position, respectively; prediction (P²) is a categorical model assigning zero distance to positions sharing a predicted stimulus (2–3 and 4–5) and unit distance otherwise. The observed geometry is the crossnobis distance matrix among the four positions of each sequence family. Each RDM is averaged over 20 independent trial subsample-and-fold partitions before fitting. Blue-to-yellow indicates raw dissimilarity, scaled to each matrix's off-diagonal range. (**b**) Each RDM is symmetrized and reduced to its six unique off-diagonal entries (the diagonal and lower triangle, greyed, are not used), giving one six-element vector per geometry per session, which is then mean-centered within family and standardized to unit variance. Greyscale indicates standardized values. **(c)** The standardized observed vector is regressed onto the three standardized component vectors, with a single coefficient per component estimated jointly across the four sequence families (D, O, B, C). Palette change from blue-to-yellow to greyscale marks the transition from raw distances to the standardized quantities entering the regression.

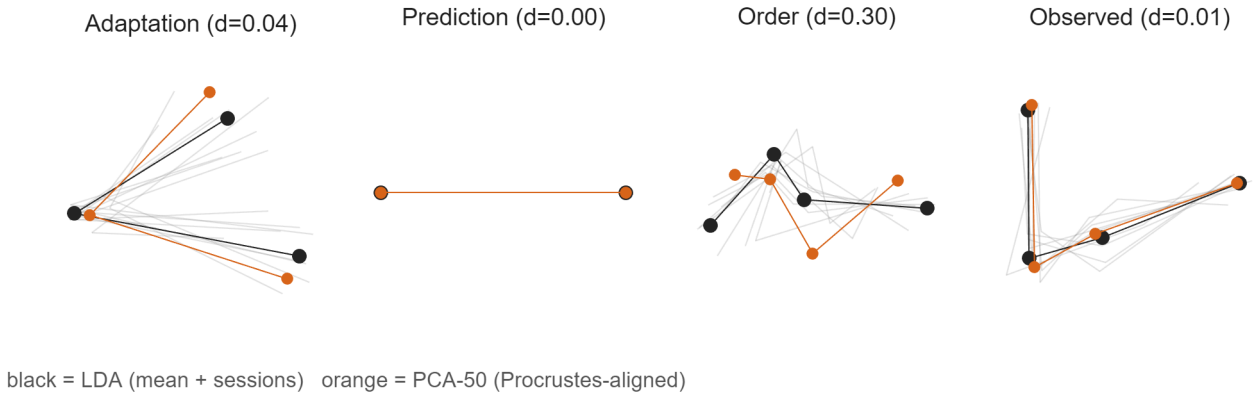


Figure S13: Representational shape is preserved across projection spaces.

Each panel shows the configuration of the four within-family sequence positions (elements 2–5) for one geometry, obtained by classical multidimensional scaling of the square root of the grand-average representational dissimilarity matrix. Black, configuration in the three-dimensional discriminant subspace (filled markers and heavy line, grand average across the nine sessions; faint lines, individual sessions Procrustes-aligned to the grand average). Orange, configuration in the unsupervised 50-dimensional principal-component space, Procrustes-aligned to the discriminant configuration. Procrustes alignment removes translation, rotation, reflection, and uniform scale, so the residual distance *d* reported in each title reflects a difference in shape alone (*d* = 0, identical shape; *d* = 1, unrelated). Adaptation (*d* = 0.04), prediction (*d* = 0.00), and the observed geometry (*d* = 0.01) have essentially identical shapes in the two spaces, and the tight clustering of the per-session configurations indicates that this shape is stable across sessions. Only the order geometry differs (*d* = 0.30): its multidimensional control geometry is expressed along a single dominant dimension within the discriminant subspace, consistent with only that component being represented in the familiar sequence (Methods). Note that the prediction model is rank-1 by construction, so its configuration is degenerate and its near-zero *d* is expected.

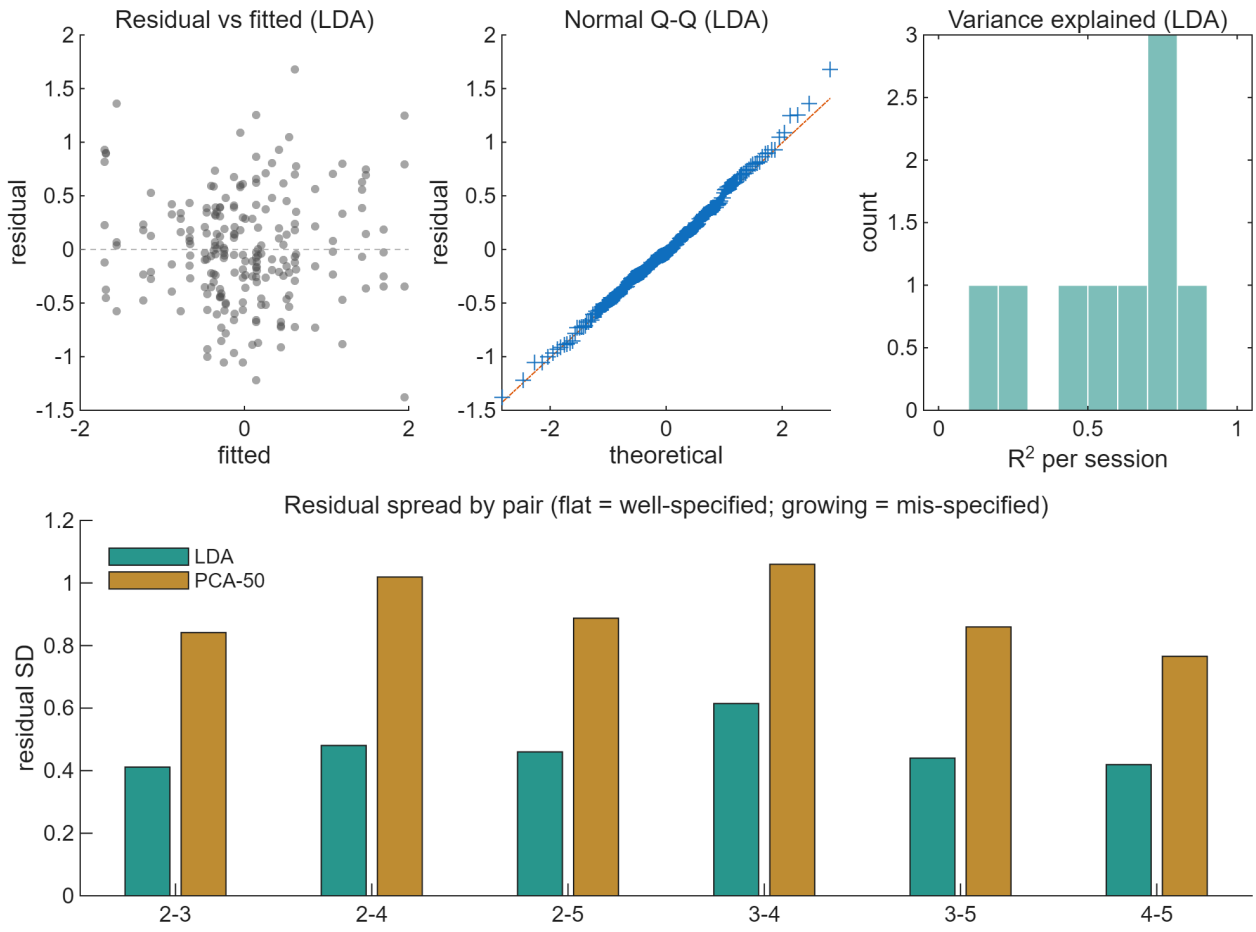


Figure S14: Regression diagnostics and model specification across spaces.

(**Top**) Diagnostics for the three-component model fit in the discriminant (LDA) space. Left, residuals versus fitted values, pooled across sessions and families, showing no systematic trend (dashed line, zero). Middle, normal quantile–quantile plot of the residuals, falling close to the reference line (dashed), indicating approximately normally distributed residuals. Right, distribution of the model's coefficient of determination (R^2) across the nine sessions. (**Bottom**) Standard deviation of the regression residuals computed separately for each of the six position pairs, in the discriminant space (teal) and the unsupervised 50-dimensional principal-component space (gold). A well-specified model leaves residuals of comparable magnitude across pairs; a mis-specified component geometry leaves larger, pair-structured residuals. Residual spread is uniformly smaller and flatter across pairs in the discriminant space, whereas in the unsupervised space it is larger and grows at the high-distance pairs—including the interior 3–4 pair that most directly reflects prediction—indicating that the additive model is better specified when the component geometries are expressed in the common discriminant frame.

Block	Condition category	Trial type	Sequence	Count per block	Count per session
Main	Standard sequence	Standard	ABBCC	200	400
Main	Omission	O1	AOBCC	20	40
Main	Omission	O2	ABOCC	20	40
Main	Omission	O3	ABBOC	20	40
Main	Omission	O4	ABBCO	20	40
Main	Omission control	XO	XOBCC	20	40
Main	Deviant	D2	ADBCC	20	40
Main	Deviant	D3	ABDCC	20	40
Main	Deviant	D4	ABBDC	20	40
Main	Deviant	D5	ABBCD	20	40
Main	Extra familiar	ExC2	ACBCC	20	40
Main	Extra familiar	ExC3	ABCCC	20	40
Main	Extra familiar	ExB4	ABBBC	20	40
Main	Extra familiar	ExB5	ABBCB	20	40
Control	Stimulus alone	A	A	10	30
Control	Stimulus alone	B	B	10	30
Control	Stimulus alone	C	C	10	30

Control	Stimulus alone	D	D	10	30
Control	Position control	A (positions 2–5)	XAXXX / XXAXX / XXXAX / XXXXA	10 each	30 each
Control	Position control	B (positions 2–5)	XBXXX / XXBXX / XXXBX / XXXXB	10 each	30 each
Control	Position control	C (positions 2–5)	XCXXX / XXCXX / XXXCX / XXXXC	10 each	30 each
Control	Position control	D (positions 2–5)	XDXXX / XXDXX / XXXDX / XXXXD	10 each	30 each

Table S1: Experimental trial conditions and counts per recording session
Control block total = 200 trials (3 blocks/session = 600). Main block total = 460 trials (2 blocks/session = 920).Total per recording session = 1520 trials.

Session	Group	L2/3 single/multi/all	L4 single/multi/all	L5/6 single/multi/all	All layers single/multi/all
1	SE	58 / 33 / 91	32 / 34 / 66	46 / 136 / 182	136 / 203 / 339
2	SE	25 / 36 / 61	24 / 24 / 48	21 / 95 / 116	70 / 155 / 225
3	SE	38 / 71 / 109	18 / 15 / 33	28 / 19 / 47	84 / 105 / 189
4	SE	37 / 20 / 57	41 / 25 / 66	60 / 128 / 188	138 / 173 / 311
5	RE	65 / 32 / 97	70 / 54 / 124	64 / 125 / 189	199 / 211 / 410
6	SE	14 / 7 / 21	12 / 6 / 18	108 / 178 / 286	134 / 191 / 325
7	RE	43 / 21 / 64	61 / 33 / 94	93 / 164 / 257	197 / 218 / 415
8	RE	11 / 14 / 25	54 / 64 / 118	93 / 163 / 256	158 / 241 / 399
9	SE	45 / 35 / 80	17 / 28 / 45	73 / 143 / 216	135 / 206 / 341
10	SE	107 / 35 / 142	22 / 37 / 59	88 / 180 / 268	217 / 252 / 469
11	RE	23 / 16 / 39	8 / 12 / 20	28 / 105 / 133	59 / 133 / 192
12	SE	36 / 16 / 52	28 / 22 / 50	69 / 129 / 198	133 / 167 / 300
13	RE	55 / 45 / 100	18 / 43 / 61	46 / 142 / 188	119 / 230 / 349
total	RE	197 / 128 / 325	211 / 206 / 417	324 / 699 / 1023	732 / 1033 / 1765
total	SE	360 / 253 / 613	194 / 191 / 385	493 / 1008 / 1501	1047 / 1452 / 2499
All		557 / 381 / 938	405 / 397 / 802	817 / 1707 / 2524	1779 / 2485 / 4264

Table S2: Number of recorded units.
Number of all recorded units, multi units that passed qc, and single units per cortical layers and overall for each session.

Chapter 5: General Discussion

What these studies establish

The preceding chapters distinguished among progressively stronger claims at the computational, algorithmic, and implementational levels. Sensory inference may be distributed across a hierarchy, the algorithm approximating that inference may combine constraints in a particular form, and the resulting computations may be realized through identifiable circuits and neural signals. The two studies presented here bear on all three levels, although with different degrees of certainty.

Both studies establish that V1 responses depend on information not specified by the current input alone. In the first study, responses to a fixed input pattern depended on the patterns that preceded it, and this dependence extended to neurons not directly driven by the preceding pattern. In the second study, the population representation of an identical stimulus differed according to the state implied by the preceding sequence context, even when the current stimulus and immediate sensory history were held constant. Within the framework developed here, both findings can be understood as instances of a latent-state estimate being constrained by information carried by statistically related variables.

The studies differ in the source of that constraint and are therefore complementary rather than redundant. The first isolated the influence of immediately preceding population states through recurrent interactions within V1. Responses were enhanced for sequences following temporal transitions characteristic of natural vision relative to sequences that did not, consistent with recurrent interactions that reflect the temporal statistics of visual experience. The second isolated a constraint derived from sequence structure extending beyond the interval over which V1 retained stimulus-specific information. This influence was content-specific: the population representation depended on which state was predicted, not only on whether the observed state was expected.

The two studies also reveal a parallel in the algebraic form of these influences. In the first, the response of a non-target neuron excited by an earlier pattern summed approximately linearly with its response to a later input, such that the response within the sequence could be predicted from the sum of its components. In the second, the current stimulus, immediate history, temporal order, and contextual prediction contributed separable components to population geometry that combined approximately additively. These findings concern different levels of description, namely the summation of single-neuron responses over tens of milliseconds in one case and the separability of representational components within a population space in the other. They therefore do not establish a shared mechanism. Both are nevertheless consistent with a linear-Gaussian approximation in which distinct constraints on the same latent variable contribute additively to its update. This consistency is not uniquely diagnostic of predictive coding, and the first study also identifies a boundary of the approximation: neurons suppressed by an earlier pattern responded sublinearly to the subsequent input, indicating that additivity does not hold uniformly across the population and may be altered by nonlinear neuronal input-output transformations.

The implementational results provide evidence against canonical proposals in which prediction errors are explicitly represented by a specialized population of spiking neurons. When identical sensory events are evaluated against different predicted states, their prediction errors differ. If a specialized population explicitly represented this difference, its members should contribute disproportionately to the prediction-content axis, producing a subset of high-weight neurons and a heavy-tailed distribution of single-neuron information. Neither pattern was observed. Prediction-content information was weak and heterogeneous at the single-neuron level, decoding accumulated gradually across many neurons, and the signal was neither aligned with sensory selectivity nor concentrated within a particular cortical layer. By contrast, deviant and omission signals were concentrated within high-weight subsets, demonstrating that the analyses could detect a concentrated architecture when one was present. These findings

therefore argue against a specialized firing-rate population explicitly representing the context-dependent prediction error within the recorded V1 population. They do not exclude error computations expressed through distributed population activity, subthreshold or dendritic signals, temporal codes, or populations outside the recorded area. The results consequently constrain a specific family of implementational hypotheses rather than the broader algorithmic claim that estimates are revised according to discrepancies between predicted and observed states.

Where the constraints come from, and what they carry

The most direct limitation of the present work is that it establishes that broader contextual information reaches V1 without determining where that information originates or precisely what the predicted content represents. Simultaneous recording from V1 and higher visual areas would begin to address the first question, although correlated activity across areas would not establish the direction of influence because both areas may receive input from common sources.

The use of arbitrary image identities also limits what can be inferred about the learned internal model. In the present design, learning transitions between particular images and learning an underlying rule capable of generating those transitions would produce the same predictions. A parameterized stochastic stimulus could distinguish these possibilities. Gillon et al. (2024), for example, used Gabor sequences in which the exact local image structure varied across presentations while the sequence elements shared a known mean orientation, with controlled violations produced by changing that sequence-level parameter. Because the precise images were not repeated, sensitivity to the shared pattern could not be explained simply by memorizing a fixed sequence of sensory observations.

Naturalistic textures synthesized from specified image statistics provide an analogous tool while also introducing a representational dimension that differs across successive cortical

stages. Areas beyond V1 discriminate among textures differing in higher-order statistics to which V1 is comparatively insensitive (Freeman et al., 2013). Texture parameters can be varied systematically to generate many sensory observations with known and graded relationships rather than two arbitrary categorical alternatives. The predicted state would then occupy a location in a parameter space, allowing the geometry induced by prediction in V1 to be compared with the geometry of the same stimuli in a simultaneously recorded higher visual area and with the distances defined by the generating parameters themselves. Determining whether prediction-induced distances in V1 track the higher-area representation, the parameter space, or neither would convert the question of prediction content into a quantitative comparison among candidate models.

Such a design would also allow representational level and temporal span to be varied more independently. The sequence context could predict a texture dimension that is weakly expressed in the feedforward-evoked V1 representation but represented more strongly in a higher visual area. Finding that this dimension nevertheless structures V1 activity during prediction would provide stronger evidence that the contextual signal was supplied by another representational level, although it would not by itself establish feedback as its anatomical source. If the generating parameter were additionally made to evolve according to a controlled transition process, it would become possible to ask whether neural predictions track an inferred latent trajectory and generalize to novel observations, or merely reproduce the particular transitions encountered during exposure.

Establishing the source of a contextual signal remains separate from establishing its mechanism. Causal perturbation of feedback pathways, local recurrent connections, or specific interneuron populations would be required to determine which circuits are necessary. The first study demonstrates the leverage available when a known input can be delivered under controlled preceding states rather than inferred solely from correlations in sensory responses. Combining this control with simultaneous multi-area recording and pathway-specific perturbation

would help distinguish locally generated temporal constraints from contextual information conveyed across the cortical hierarchy.

Dynamics and the form of the update

The analyses presented here characterize activity averaged over response windows and therefore primarily describe the outcome of inference rather than its temporal course. The predictive-coding algorithm developed in the introduction is explicitly iterative: estimates are repeatedly revised as constraints propagate, and a mutually consistent state is approached over time rather than computed in a single step. Time-resolved population analyses could determine whether contextual constraints act early or late relative to the sensory response, whether their influence increases as the estimate settles, and whether population trajectories differ across predictive contexts even when their endpoints are similar.

This bears directly on a distinction the present analyses cannot resolve. Contextual input may shift an estimate independently of its current value, or it may alter the dynamics through which that estimate evolves. The approximately additive relationship observed here is consistent with the first possibility, but it was measured over averaged response windows and a limited set of conditions. A contextual modulation of population dynamics operating over a restricted range of states could appear approximately additive when only the resulting average representation is examined. Characterizing the trajectory itself would therefore provide a stronger test of the update proposed by predictive-coding algorithms and could reveal nonlinear interactions that are obscured at the endpoint.

Precision and reliability

The introduction argued that precision enters the inferential objective directly, causing different constraints to be weighted according to their estimated reliability. Neither study independently manipulated predictive reliability. In the second study, reliability was held

approximately constant across prediction contents so that the effect of content could be isolated. Manipulating the consistency with which a context predicts a subsequent state would test whether its influence on the population representation scales with its estimated precision.

Such a manipulation would also help distinguish prediction content from changes in gain. Predictability can alter both the expected state and the confidence assigned to that expectation. A stronger or more reliable prediction may therefore produce a larger representational influence without changing what is predicted. Varying reliability while holding predicted content constant, and varying content while holding reliability constant, would allow these contributions to be estimated separately. The resulting population geometry could reveal whether precision scales the prediction-related component, alters the weighting of sensory evidence, or changes the population dynamics through which the two are combined.

Behavioral goals and active inference

Reliability is not the only quantity that determines how sensory information contributes to behavior. In Bayesian decision theory, the posterior describes uncertainty about the latent state, whereas the estimate or action selected from that posterior depends on the associated loss function. Asymmetric costs, rewards, or consequences can therefore change the optimal response without changing the posterior itself. In a distributed neural system optimized end to end, however, this conceptual distinction need not correspond to a strict anatomical division. Behavioral objectives may shape earlier representations through learning, feedback, attention, or neuromodulation, potentially changing which sensory features are preserved and how prior and current information are expressed in V1. Consistent with this possibility, predictive signals can depend on task demands, reward contingencies, and proximity to behavioral goals (Furutachi et al., 2024; Garner & Keller, 2022; Solomon et al., 2021).

The second study deliberately avoided making prediction content relevant to a behavioral choice, and reward did not depend on correctly identifying or predicting the stimulus.

This reduced the possibility that differences in choice, movement, reward expectation, or task strategy would account for differences between predictive contexts. The results therefore establish how broader contextual predictions alter V1 representations under similar behavioral demands. Having established this baseline, a subsequent experiment could vary the behavioral relevance or consequences of the predicted state while preserving the sensory sequences and their statistical reliability. Behavioral responses and motor requirements would need to be measured and matched so that changes in population structure could be attributed to task relevance rather than to the actions themselves. Such a design could determine whether goals merely alter the downstream use of an otherwise stable V1 representation or reshape the representation itself.

Active inference introduces a further role for behavior. Uncertainty can be reduced by updating an internal estimate from the observations already available, but it can also be reduced by acting to obtain more informative observations. Epistemic actions seek information that distinguishes among uncertain latent states, whereas pragmatic actions alter the sensory environment in ways that help attain preferred outcomes. In active-inference formulations, perceptual updating and action selection therefore form a coupled process: internal estimates guide the selection of observations, and the observations generated by those actions further update the estimates (Friston et al., 2018; Pezzulo et al., 2022).

The present experiments largely isolated perceptual updating from action-mediated information gathering because the animals could neither control the sensory input nor select observations that would resolve uncertainty about the predicted state. A closed-loop task could instead allow an animal to perform an action that reveals otherwise unavailable information, making it possible to compare uncertainty reduction through internal updating with uncertainty reduction through active sampling. Neural activity before and after the informative action could then reveal whether V1 represents the uncertainty motivating the action, the expected information it will provide, or the revised estimate obtained afterward. This would extend the

present results from prediction under externally controlled observations to the coupled perception-action loop within which sensory inference normally operates.

What the code carries

The analyses used here assume that relevant information is carried by population firing rates averaged over defined response windows. They therefore address only what such a rate-based population code can represent. Prediction-related information expressed through precise spike timing, correlation structure, coordinated assembly activity, or the order in which neurons become active would be attenuated or invisible in these measurements. Temporal and sequential codes are particularly relevant to the present paradigms because a contextual prediction may alter the trajectory or activation order of the population without producing a large change in its mean firing-rate vector.

A separate question concerns what the neural activity represents about the posterior. Cortical activity could encode a point estimate, generate samples from a distribution over time, or parameterize a distribution across the population. These possibilities assign different meanings to trial-to-trial variability. Under a point-estimate account, such variability may primarily constitute noise around the inferred value. Under a sampling account, the variability may be part of the representation itself. Under a probabilistic population code, the distribution of activity across neurons may jointly encode both an estimate and its uncertainty. The present analyses establish that prediction content is recoverable by a linear discriminant applied to population firing rates, but they do not distinguish among these representational formats.

These are not peripheral questions. The theoretical framework specifies what should be estimated and how multiple constraints may combine, but the empirical content of those claims depends on which feature of neural activity carries the estimate and what that feature encodes. Extending the analysis to temporal structure, coordinated variability, and trial-resolved population dynamics will be necessary to determine whether uncertainty and prediction are

represented only in average firing rates or also in the organization of neural activity across neurons, trials, and time.

Summary

Together, these studies show that temporal statistics operating at two timescales reshape representations in V1, the earliest cortical stage of visual processing. At the shorter timescale, preceding population states influenced subsequent responses through recurrent interactions within V1, with enhanced responses to transitions characteristic of natural vision. At the longer timescale, broader sequence context altered population representations according to the content of the predicted state, even when the current stimulus and immediate history were held constant. In the second study, the current stimulus, immediate history, temporal order, and contextual prediction contributed separable components that combined approximately additively.

The approximately additive organization observed across the two studies is consistent with linear-Gaussian formulations in which distinct constraints contribute jointly to the estimate of a latent state, although it does not uniquely identify predictive coding or establish globally linear circuit operations. The distributed organization of prediction-content information further argues against a specialized firing-rate population explicitly representing the corresponding prediction error within V1. What remains unresolved is where broader contextual predictions originate, how they update sensory representations over time, what regulates their influence, and in what format the resulting uncertainty is encoded. The present results provide a foundation for addressing these questions by establishing how recent population states and broader contextual predictions contribute to V1 representations when current sensory input and behavioral demands are controlled.

References

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature neuroscience*, *16*(7), 974-981.

Furutachi, S., Franklin, A. D., Aldea, A. M., Mrsic-Flogel, T. D., & Hofer, S. B. (2024). Cooperative thalamocortical circuit mechanism for sensory prediction errors. Nature, 633, 398-406.

Garner, A. R., & Keller, G. B. (2022). A cortical circuit for audio-visual predictions. *Nature Neuroscience, 25*(1), 98–105.

Gillon, C. J., Pina, J. E., Lecoq, J. A., Ahmed, R., Billeh, Y. N., Caldejon, S., … Zylberberg, J. (2024). Responses to pattern-violating visual stimuli evolve differently over days in somata and distal apical dendrites. *Journal of Neuroscience, 44*(5).

Solomon, S. S., Tang, H., Sussman, E., & Kohn, A. (2021). Limited evidence for sensory prediction error responses in visual cortex of macaques and humans. *Cerebral Cortex, 31*(6), 3136–3152.