Spontaneously emerging internal models of visual sequences combine abstract and event-specific information in the prefrontal cortex

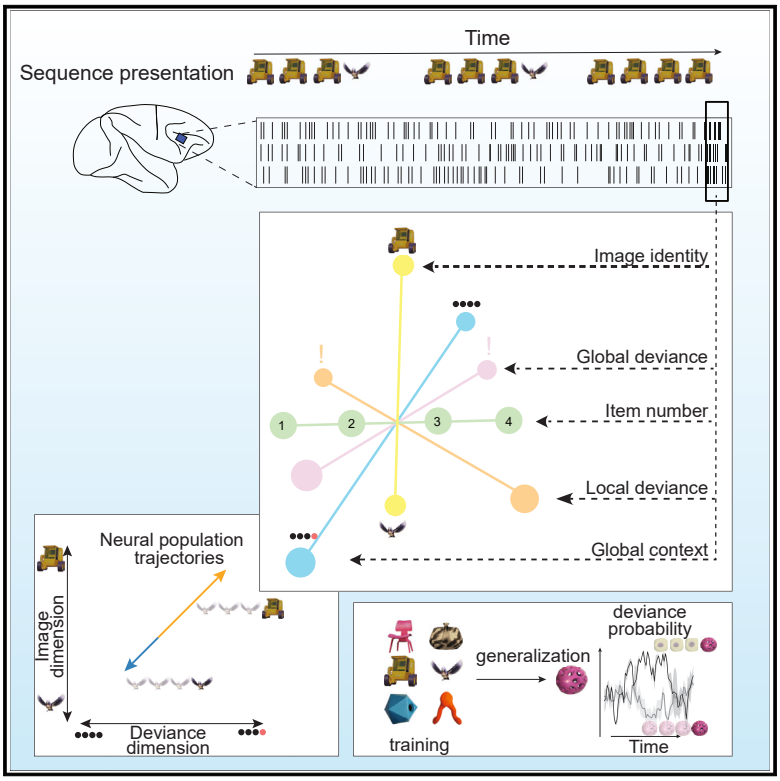
Marie E. Bellet, Marion Gay, Joachim Bellet, Bechir Jarraya, Stanislas Dehaene, Timo van Kerkoerle, Theofanis I. Panagiotaropoulos

Cell Reports

Article

Spontaneously emerging internal models of visual sequences combine abstract and event-specific information in the prefrontal cortex

Graphical abstract



Highlights

- Prefrontal populations recorded during a no-report local-global paradigm of visual sequences
- Structure, identity, and prediction errors decoded from superimposed population subspaces
- Rapid update of sequence-structure representations following exposure to a mismatch sequence
- Generalization of sequence structure and prediction error representations

Authors

Marie E. Bellet, Marion Gay, Joachim Bellet, Bechir Jarraya, Stanislas Dehaene, Timo van Kerkoerle, Theofanis I. Panagiotaropoulos

Correspondence

marieestelle.bellet@gmail.com (M.E.B.), theofanis.panagiotaropoulos@cea.fr (T.I.P.)

In brief

Bellet et al. show that neuronal populations in the prefrontal cortex spontaneously form rich internal models that represent abstract and identity-specific information of visual sequences with different complexity. These models, detected in superimposed population subspaces, rapidly update after presentation of mismatch sequences and generalize to sequences with identical structure.

Bellet et al., 2024, Cell Reports *43*, 113952
March 26, 2024 © 2024
https://doi.org/10.1016/j.celrep.2024.113952

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Spontaneously emerging internal models of visual sequences combine abstract and event-specific information in the prefrontal cortex

Marie E. Bellet,[1,*] Marion Gay,[1] Joachim Bellet,[1] Bechir Jarraya,[1,2,3] Stanislas Dehaene,[1,4,7] Timo van Kerkoerle,[1,5,6,7] and Theofanis I. Panagiotaropoulos[1,7,8,*]

[1]Cognitive Neuroimaging Unit, INSERM, CEA, Université Paris-Saclay, NeuroSpin Center, Gif-sur-Yvette, France
[2]Université Paris-Saclay, UVSQ, Versailles, France
[3]Neuromodulation Pole, Foch Hospital, Suresnes, France
[4]Collège de France, Université Paris-Sciences-Lettres (PSL), Paris, France
[5]Department of Neurophysics, Donders Center for Neuroscience, Radboud University Nijmegen, Nijmegen, the Netherlands
[6]Department of Neurobiology and Aging, Biomedical Primate Research Center, Rijswijk, the Netherlands
[7]Senior author
[8]Lead contact
*Correspondence: marieestelle.bellet@gmail.com (M.E.B.), theofanis.panagiotaropoulos@cea.fr (T.I.P.)
https://doi.org/10.1016/j.celrep.2024.113952

SUMMARY

When exposed to sensory sequences, do macaque monkeys spontaneously form abstract internal models that generalize to novel experiences? Here, we show that neuronal populations in macaque ventrolateral prefrontal cortex jointly encode visual sequences by separate codes for the specific pictures presented and for their abstract sequential structure. We recorded prefrontal neurons while macaque monkeys passively viewed visual sequences and sequence mismatches in the local-global paradigm. Even without any overt task or response requirements, prefrontal populations spontaneously form representations of sequence structure, serial order, and image identity within distinct but superimposed neuronal subspaces. Representations of sequence structure rapidly update following single exposure to a mismatch sequence, while distinct populations represent mismatches for sequences of different complexity. Finally, those representations generalize across sequences following the same repetition structure but comprising different images. These results suggest that prefrontal populations spontaneously encode rich internal models of visual sequences reflecting both content-specific and abstract information.

INTRODUCTION

How do we spontaneously encode the specific elements of experiences and at the same time learn their structure to generalize to new situations? Resolving how the brain encodes sequential patterns of sensory experience on the fly, without any explicit task demands, is key to understanding the fundamental computations underlying higher-order cognition. At the neuronal level, internal models implemented by neuronal populations could encode information about a sensory sequence at both an abstract structural level and at a concrete sensory level. Importantly, the reinstatement of an internal model, i.e., reactivating the same neuronal ensemble to encode the same structure when it is encountered in new circumstances, may provide a simple mechanism for generalizing knowledge across similar experiences.[1–3] Nevertheless, a complete model should also encode specific aspects of sequential experiences, such as the pictures being presented. Therefore, understanding how neuronal populations spontaneously encode sequences of stimuli may provide insights about the abstract scaffolding mechanism that supports generalization as well as the event-specific neural populations that encode the contents of perception.

Both abstract processing of sequential information and conscious perception of specific stimuli appear to converge in the prefrontal cortex (PFC). The PFC is critical for temporal organization of task-related behavior through the encoding, maintenance, and flexible use of abstract rules, representations, and schemas that guide behavior and facilitate cognitive control.[4,5] At the same time, apart from such contextual and abstract information that can be applied to new circumstances, prefrontal populations also encode the specific consciously perceived contents of sensory experience.[6–8] This dual role of the PFC suggests that the instantaneous prefrontal population activity should encode both properties, even when a sequence of events is encountered without any task-related demands. This hypothesis leads to predictions concerning the expected structure of the population code during a sequential sensory experience. On the one hand, the neural code should disentangle, or factorize, the experience into its main structural variables, such as the number of items or the presence of repeated elements,[9–11]

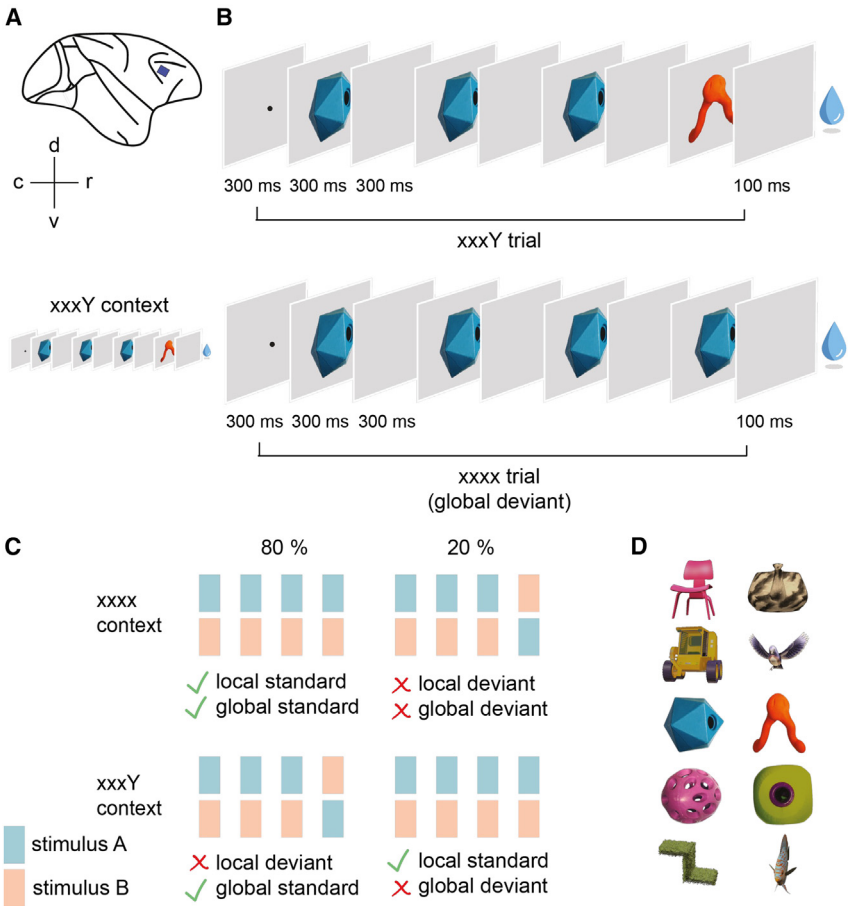


Figure 1. Recording vlPFC spiking activity during the visual local-global paradigm

(A) Location of the implanted Utah arrays in the macaque vlPFC.

(B) Example trials. A sequence of four stimuli was presented and 100 ms after offset of the last stimulus, the monkeys received a liquid reward. Reward time was constant for all conditions; therefore, a differential impact on each condition was unlikely. Examples show a single xxxY and an xxxx trial within the context of frequent xxxY sequences.

(C) Each session consisted of four blocks comprising a frequent sequence (global standard, which could have the structure xxxx or xxxY) and a rare sequence (global deviant). In each block, the x was a fixed image (A or B, taken from the pairs in D) and the Y was the other image (B or A).

(D) The five pairs of visual stimuli (rows) used in the experiments.

thus allowing for generalization when similar variables are encountered, a hallmark of abstract processing.[1,12] Simultaneously, apart from abstract information about structure, a complete model should also include concrete information about specific events or contents. Using different dimensions within the same overall population, the PFC would represent the flow of incoming information by a dynamic flow of neural activity encoding, e.g., "This is the first picture, it is a tree…now here is a second tree…a third tree…but the fourth picture is different! It is a face." According to our hypothesis, all the above labels, whether abstract (first, second, third, different…) or concrete (face, tree), should be decodable from PFC activity.

The neuronal properties in the PFC are ideal for a population code that can accommodate multiple representations that balance abstract and specific information. In the PFC, a large percentage of neurons exhibit mixed selectivity for the different variables of a task and encode multiple task-related overlapping representations.[13–16] This mixed selectivity property facilitates high-dimensional representations and, therefore, linear readouts of many different variables from population activity.[12,13] However, the computational capabilities of the PFC would be constrained if the population code allowed only for higher-dimensional representations. In particular, the role of the PFC in abstract processing suggests that prefrontal population activity should also converge into a low-dimensional representation that is optimal for extracting contextual information and generalizing across similar circumstances.[12,13]

Here, we studied how prefrontal population activity encodes visual sequences and generalizes across them spontaneously without any overt task or response demands. We used a visual version of the "local-global" paradigm that probes sequence processing at two hierarchical levels, local (sequence element transition probabilities) and global (whole sequence).[9,17,18] First, we asked whether the PFC encodes all aspects of the sequences in the local-global paradigm, therefore holding a complete internal model of the ongoing sensory stream and its occasional violations. Second, are some of these neural representations abstract enough to be independent of the specific stimulus identities within the sequence pattern, as predicted by earlier work using functional magnetic resonance imaging?[11]

We tested these hypotheses by recording from chronically implanted multielectrode arrays in macaque ventrolateral PFC (vlPFC) during the local-global paradigm. Using multivariate decoders, we show for distinct prefrontal population subspaces encoding of all aspects of the visual sequences, including image identity, serial position of stimuli, and abstract sequence structure, as well as local and global structure violations. These results reveal the fundamental computations in prefrontal ensembles mediating the spontaneous encoding of sequential information.

RESULTS

vlPFC spiking activity during the visual local-global paradigm

We recorded spiking activity from neuronal populations with multielectrode Utah arrays, chronically implanted in the vlPFC of two macaque monkeys (Figure 1A), during a visual version

Table 1. Terminology of sequence types in the local-global paradigm

Sequence (one trial)	Global context[a]	Local deviance	Global deviance	Terminology
xxxx (aaaa, bbbb)	xx block	standard	standard	xx|xx (frequent xx)
	xY block	standard	deviant	xx|xY (rare xx)
xxxY (aaaB, bbbA)	xx block	deviant	deviant	xY|xx (rare xY)
	xY block	deviant	standard	xY|xY (frequent xY)

[a]The global context corresponds to the structure of the frequent trials in a block. xx is short for an xxxx trial and xY is short for an xxxY trial.

of the local-global sequence paradigm[17] (Figures 1B–1D). To eliminate activity related to decision making and association of sequences with specific motor responses that could confound sequence specific population coding, the task did not require overt behavioral reports but mere sequence observation. Therefore, this no-report paradigm allowed studying the spontaneous emergence of population codes during passive observation of visual sequences. On each trial, we presented a sequence of four images (300 ms stimulus duration; 300 ms inter-stimulus interval) while the monkeys maintained their gaze for the whole trial duration within the image region (Figure 1B). For completion of a trial, the monkeys received a liquid reward 100 ms after offset of a sequence. A sequence consisted either of four repeats of the same stimulus (xxxx sequence, abbreviated as xx) or of three repeats and one local deviant in the last position (xxxY sequence, abbreviated as xY; Figure 1C). We presented these sequences in blocks of 200 trials where one sequence type (xx or xY) was the frequent sequence (global standard). The global standard sequence was presented in the first 50 trials of each block. Of the remaining 150 test trials, 80% were global standards and 20% were global deviants, which differed only in the last position compared to the standard. We will use a notation that indicates trials according to their local structure and global context: e.g., a rare xY trial (in an xx block) will be denoted as xY|xx. The first two letters indicate the current trial and the last two letters the global context in which it occurred (Table 1). The design enabled us to distinguish the effects of first-order (local) vs. higher-order (global) sequence regularity, which requires representing the whole sequence pattern (STAR Methods).

In each recording session, a specific pair of images (A and B) was chosen out of five possible pairs (Figures 1C and 1D), and four blocks were run with this picture pair in random order (two xx blocks, aa and bb; two xY blocks, aB and bA). This allowed us to test whether the population code of sequence structure generalized within sessions, where the order of the images changed, and across sessions, where stimulus identities were different. For the latter, it was important that the signal was stable across recording days. We found that the recorded multi-unit activity (MUA), i.e., the sum of recorded spikes from each electrode, was robust over several days in both animals, with overall more active sites in monkey A. Figure S2 shows that the channels that carried most information about the sequences (see later decoding analysis) contained mostly MUA and that isolated single units were not stable across all recording days. We therefore decided to perform most analyses using MUA.

To analyze the selectivity properties of the recorded PFC neurons, we used isolated neurons from monkey A, for which we were able to obtain 11 neurons on average per recording. To assess selectivity of single neurons in our experiments, we performed an ANOVA on the responses of individual neurons[19,20] with the independent variables stimulus, local and global deviance, and first-order interactions to reveal non-linear mixed selectivity (NMS) (STAR Methods). The total degree of freedom (df) was 279 (monkey A) and 267 (monkey H) on average per neuron, and the df for factors and interaction was 1. p values were corrected using false discovery rate (FDR)[21] across all six recording sessions, and values at $p < 0.01$ were regarded as significant (see Table S1 for statistics). We found that on average 28% of neurons had NMS (27% for stimulus and local, 8% for local and global, 3% for stimulus and global—note that these percentages do not sum to 28% because one neuron could have several interaction effects). Four percent of neurons had linear mixed selectivity to local deviance and stimulus identity, 33% had classical selectivity to only one variable (13% stimulus, 19% local, 2% global), and 35% were not selective (Figure 2D; for statistics see Table S1). Local deviants and stimulus identity were most prominently represented by single-neuron responses, whereas global deviants yielded responses of only a few neurons.

vlPFC ensembles form a rich representation of visual sequences

Given the diversity of response types and signs of mixed selectivity, we hypothesized that visual sequences could be better represented by neuronal population vectors rather than by individually specialized units, as previously shown for the PFC.[12–15,22–24] We used regression analyses to test whether the vlPFC represented all the variables that defined the visual sequences in the local-global paradigm, namely, (1) stimulus identity (one of two possible images in each session), (2) serial position of the image within each sequence (from 1 to 4), (3) global context (xx or xY block), (4) local deviance, and (5) global deviance. For this analysis, we used only the data from the test trials that followed the first 50 habituation trials in each block to ensure that the current global context was learned (see also Figures 4G and 4H).

We applied multiple linear regression for all variables but serial position, for which we used multinomial logistic regression (STAR Methods). This approach allowed us to determine which population vectors, if any, carried maximum information about each sequence variable. We then used these vectors to reduce the dimensionality of the MUA and obtain trajectories of the neural population within each neural subspace (Figure 3). Using the subspace trajectories for classification allowed us to perform a decoding analysis for each of the sequence variables. We quantified such decoding, relative to chance level, using the

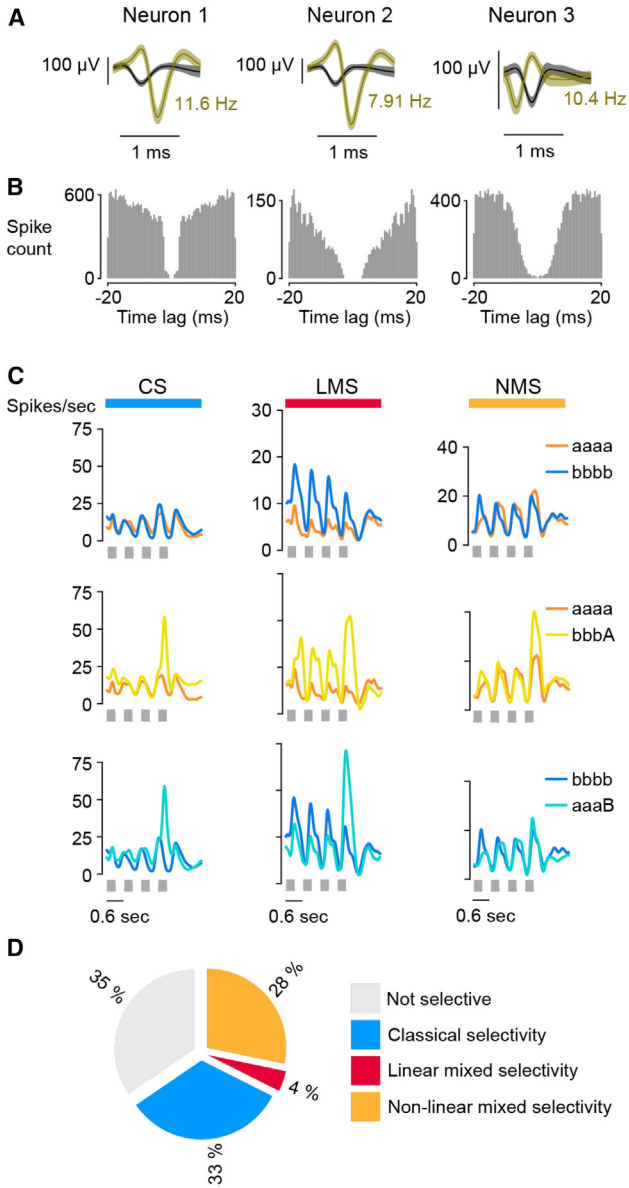


Figure 2. Neural selectivity during the local-global paradigm

(A) Spike waveforms from three example neurons in monkey A, selected on the basis of their selectivity (see C). Isolated neurons are shown in green.

(B) Auto-correlograms of isolated neurons in (A).

(C) Peri-stimulus time histograms of three isolated neurons for different sequences: aaaa in orange, bbbb in blue, bbbA in yellow, and aaaB in cyan. The neuron on the left shows classical selectivity (CS), in this case selectivity for local deviants. The neuron in the middle exhibits linear mixed selectivity (LMS), i.e., joint selectivity for stimulus b and for local deviants. The neuron on the right shows non-linear mixed selectivity (NMS), i.e., selectivity for a combination of stimulus A and local deviance. Statistics are listed in Table S1. In this table, neuron 1 corresponds to ID 14, neuron 2 to ID 1, and neuron 3 to ID 10.

(D) Ratio of selectivity types, averaged across six recording sessions in monkey A, with 11 neurons per session on average.

area under the receiver-operating characteristic curve (AUROC; see Figure S1 for methods). A random permutation test with cluster-based correction for multiple comparisons was used to obtain p values for each tested time bin (STAR Methods). Bins were regarded as significant at $p < 0.05$. For the number of trials per recording and condition, see Table S2.

The results showed that all sequence variables could be decoded at above-chance levels (Figure 3). First, the decoding performance for stimulus identity (image A vs. image B within each recording session) was close to 1 for every item in a sequence, including the last sequence item when it changed on xY trials (Figures 3A–3F).

Second, using a separate decoder for serial position within a sequence, we could predict serial position, particularly for the first and last items, reflecting known primacy and recency effects,[25–27] but also at above-chance levels for positions 2 and 3 (Figures 3B–3F, predictions are indicated by horizontal bars). Since the sequences used a fixed timing, this decoding could reflect numerical codes, temporal codes, or both.[10,28] However, elapsed time alone could not explain all of the findings, such as the fact that the code for "1st item" was partially reactivated for the last image of xY trials (the first image with this identity) or that the code for "4th item" was reactivated at ordinal positions 1, 2, or 3 on trials when the monkey broke fixation and the visual sequence was aborted, suggesting that it actually responded to "last item" (Figure S3). These findings indicate that those population codes were partially locked to the phase of the task and not solely to timing.

Third, to test whether vlPFC neurons contained a model of the upcoming sequence structure, we decoded the global context (xx or xY block) from the neuronal population activity prior to the last stimulus of a sequence. We indeed identified a population subspace whose activity allowed us to infer the global context even before the sequence presentation started (Figures 3C–3H). Could this effect be carried by stronger firing-rate adaptation in xx blocks compared to xY blocks (because of more repetitions of the stimulus x)? To address this alternative explanation, we tested context sensitivity in individual channels using independent t tests on the average response during the first three stimuli. p values were corrected for multiple comparisons across channels that were responsive to the sequences (STAR Methods). For channels with a p value of <0.01, we assessed the direction of the effect, separating channels with decreased firing rate in xx blocks from those with increased firing rate in xx blocks, compared to xY blocks, using an independent t test. On average across sessions, 64 channels were responsive in monkey A (22 in monkey H), out of which 12 channels were context sensitive (6 in monkey H). Statistics are given in Tables S3 and S4. Eight out of the 12 channels showed decreased firing rate during the first three stimuli of xx blocks compared to xY blocks, but four channels had an increased firing rate (3 and 3 in monkey H). The presence of both increased and decreased firing indicates that context encoding cannot be explained solely by firing-rate adaptation. In the following section, we will look in more detail at the properties of the neural subspace-encoding context and how it builds up during the habituation period.

Fourth, we assessed responses to violations of either local or global sequence regularity. The population that was sensitive to local deviance showed a response to both predicted (xY|xY) and unpredicted (xY|xx) local deviants (Figures 3D and 3I, orange and

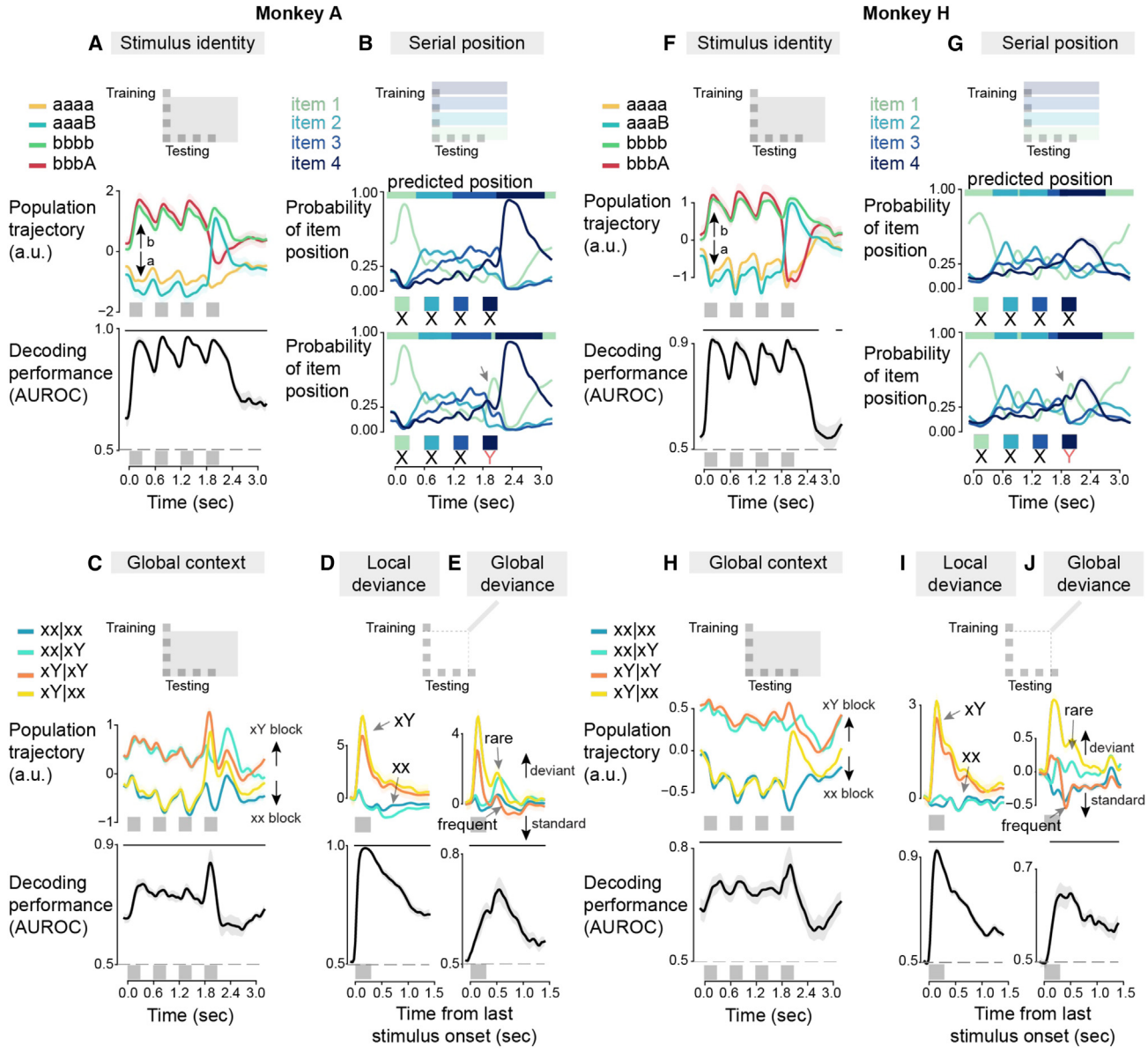


Figure 3. Decoding neural population codes for different aspects of the sequences

Upper plots in all panels (and lower plot in B and G) show the population trajectories resulting from the multiple linear regression (with the time windows used for training and testing indicated in the insets above the plots). Black traces in the bottom plots show decoder performance in terms of AUROC, relative to chance level 0.5. Horizontal lines on top of each graph indicate the time points for which the decoding performance was significantly above chance ($p < 0.05$). Data from monkey A (left, A–E) and monkey H (right, F–J).

(A and F) Decoding of image identity (picture A vs. picture B).

(B and G) Decoding of the four ordinal positions in the sequence. Colored curves show the predictive probability of decoders (derived from multinomial logistic regression) trained on xx trials and generalized to xY trials.

(C and H) Decoding of global context, i.e., xY blocks vs. xx blocks. Note how decoding is significant even prior to sequence presentation, indicating an anticipation of the forthcoming sequence.

(D and I) Decoding of local deviance, i.e., xx vs. xY sequences.

(E and J) Decoding of global deviance, i.e., frequent vs. rare sequences.

In (D), (I), (E), and (J) only, decoding was time-locked to the last sequence item (−100 ms to 1,400 ms relative to last stimulus onset). Training/testing scheme shows that decoder was trained and tested on each time bin after last stimulus onset. A positive activation indicates a local or global deviance signal, respectively.

yellow). Decoding local deviance was almost perfect on a single-trial basis, with an early peak (~200 ms) indicating a very robust and fast response to local novelty. Nevertheless, the activation was stronger upon unpredicted than predicted local deviants, in agreement with the predictive-coding framework (Figures S4G and S4H, left). We also decoded global deviance

(rare vs. frequent sequences in a given block), with the effect showing a later peak ~500 ms after last item onset. In contrast to the unimodal phasic response observed for local deviants, the trajectories encoding global deviance showed a biphasic response. First, until around 300 ms (200 ms in monkey H) following the last stimulus onset, only trials with a local deviant showed a positive activation (Figures 3E–3J, orange and yellow), again with a higher amplitude for unpredicted local deviants. Somewhat later, both types of rare trials (xY|xx and xx|xY) evoked a response in the same direction (yellow and cyan) and opposite to the frequent trials (orange and blue). We separately measured the global deviance decoding performance by computing the AUROC for rare vs. frequent xY (local and global deviants, see Figures S4C and S4D, dashed lines) and rare vs. frequent xx trials (pure global deviants, see Figures S4C and S4D, solid lines). The effect was significant for both and slightly earlier for xY than for xx trials, suggesting that deviant detection for xx trials requires a longer integration time.

We next investigated whether global deviance detection was abstract and invariant for sequence pattern by evaluating a "cross-condition" decoder trained on xx trials and tested on xY trials (Figures S3E and S3F). We did indeed find that decoding significantly generalized across the two sequence structures, occurring late after the last item onset (~300–500 ms), indicating that by that time, the neural code for global surprise was shared by the two sequences with a different local structure. This finding also suggests that the population subspace coding for global deviance might be different from the respective subspace coding earlier for local deviance. We further assessed the independence of these two subspaces by decoding global deviance from the subspace that represented local deviance and found that the late global mismatch response was indeed not encoded in the subspace that showed a late local mismatch response (Figures S4G and S4H).

Importantly, the generation of eye movements, found to be a behavioral readout of local novelty detection, could not explain neural deviance responses, as the effects remained after the removal of trials in which the monkeys made an eye movement (Figures S5A and S5B). The different timescales of the local and global effect are consistent with previous results showing that processing of global sequence violations requires longer times that reflect conscious integration compared to the detection of local deviants that can happen non-consciously.[17,29] We assessed how distributed the subspaces are by computing a "participation ratio" (PR), which indicates the fraction of sites that contributes significantly to the subspace (STAR Methods). The PR for the stimulus subspace was 73% in monkey A and 74% in monkey H, for context 72% and 73%, for local deviance 72% and 69%, and for global deviance 74% and 69%. For item position, the PR was 74% and 70% for item 1, 72% and 70% for item 2, 74% and 67% for item 3, and 72% and 66% for item 4. Only sites that were responsive to the sequences were considered, which were 64 for monkey A and 22 for monkey H. These ratios indicate that the different subspaces were distributed across the populations and not restricted to only a few responsive sites.

Decoding performance of the main effects after the last stimulus in the sequence was not altered when interaction effects were included (Figures S5C and S5D). Furthermore, all pairwise interactions could be decoded from the population activity, potentially due to the block structure of the paradigm.

Altogether, these findings indicate that the vlPFC population comprised multiple, overlapping, and distributed representations of all the features of the visual sequences used in the local-global paradigm.

The representation of global context is learned spontaneously during habituation and is updated upon errors

Predictive-coding models suggest that following a global context violation, the internal model is destabilized or updated, at least transiently. We examined whether this could be detected in the population activity within the subspace that represented global context (xx or xY sequence) (Figures 4A and 4B). In both monkeys, there was sustained activity persisting throughout all trials and distinguishing xx from xY context, regardless of the previous trial (Figures 4C and 4D). A trial in an xY block, for example, led to an activation into the "xY direction" of this subspace during the first three stimuli in a sequence, whether following a global standard xY|xY (Figures 4A and 4B, orange solid lines) or a global deviant xx|xY trial (Figures 4A and 4B, orange dashed lines). This observation is important, as it indicates that the activity was not simply due to a lingering memory of the previous trial but was sustained in the long term and resistant to an occasional global deviant, as needed to encode the global context of an entire session. Nevertheless, the encoding strength of global context was reduced after the occurrence of a global deviant (Figures 4A and 4B, dashed vs. solid lines). This effect was transient, and activity was quickly restored within the following 1–2 standard trials (Figures 4E and 4F).

We also examined how fast this activity built up during the first 50 habituation trials of a given block (Figures 4G and 4H). The divergence between xx and xY blocks increased continuously through the habituation period of 50 trials (linear regression of the average difference between xY and xx blocks onto the trial number for monkeys A and H, respectively: slope mean, 0.0069 and 0.0031; confidence interval, ±0.0015 and ±0.0007; Z, 8.87 and 8.65; $p = 7.40 \times 10^{-19}$ and 5.26×10^{-18}).

Together, these findings show that the population activity reflected the global sequence context that was inferred at a long timescale while at the same time being updated on a shorter timescale whenever a deviant sequence occurred. The former finding indicates the spontaneous emergence of a neural representation of global sequence regularity in the vlPFC, while the latter fits with a transient destabilization or update of this model after a mismatch in the learned global structure.

A shared population code for global deviance and context

Current models of predictive coding hypothesize that, at each level of the cortical hierarchy, separate neural populations code for predictions and prediction errors[30] According to this hypothesis, expectation and mismatch or error signals should be detected in distinct, segregated ensembles. Alternatively, the representation of context might happen within the same neural population that also emits the mismatch responses. Such integration might be especially relevant for higher-order

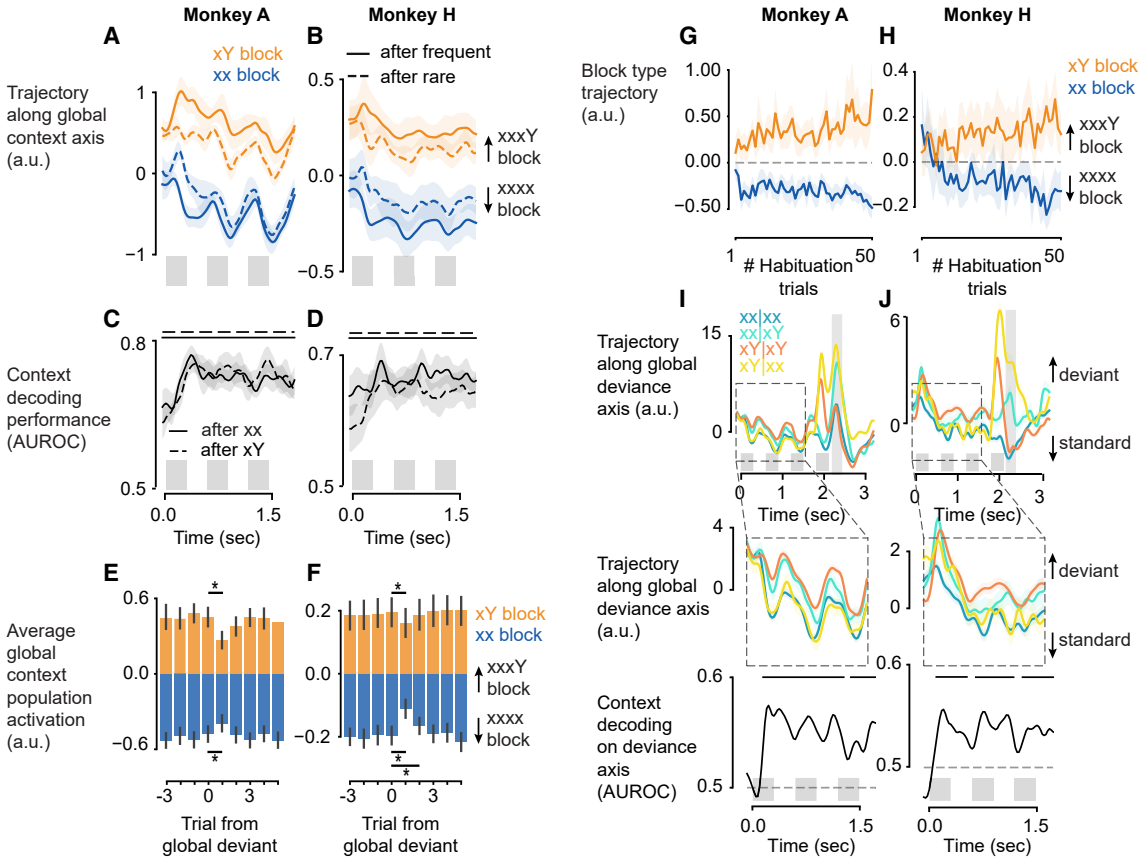


Figure 4. Neural signals reflecting the updating of global sequence knowledge

(A and B) Population activity projected onto the axis coding for global sequence knowledge, i.e., xY blocks (orange) vs. xx blocks (blue). Dashed curves indicate a reduction in the neural separation of xY and xx blocks after a rare global deviant.

(C and D) In both blocks, the global context can still be decoded after a global deviant trial (after xx trials in xY blocks and after xY trials in xx blocks), i.e., when the previous trial is suggestive of the opposite block.

(E and F) Activation of the global context axis averaged over the first three stimuli in each trial, aligned to rare trials in xY blocks (orange) or xx blocks (blue). Height of bars indicates average across trials from pooled sessions, and error bars are the 95% confidence interval. Asterisks denote a significant change in global context signal between the trial before the global deviance occurred (0) and the following trial ($p < 0.05$, paired t test).

(G and H) Buildup of sequence knowledge during habituation. The population activity during the 50 habituation trials of each block was projected onto the population axis that encoded global context and averaged over the first three stimuli in each trial. Lines show average of blocks from all sessions (12 in monkey A, 20 in monkey H), and the shaded area is the SE. The divergence between block types over trials was significant for both animals (linear regression of the difference between xY and xx blocks, $p < 0.05$, see main text).

(I and J) Trajectories of the neural population that lead to maximal overall global deviance decoding performance (Figures 3E–3J), indicated by the gray shaded area. Middle panel: zoom into the time prior to the onset of the last stimulus. Bottom panel: quantification of the context-decoding performance based on the activation of the global deviance population.

Horizontal bars on top of the AUROC plots indicate periods of significant context decoding. Horizontal dashed line shows the chance level. Only trials following a global standard sequence were used. Curves indicate average across all trials from all sessions, and shaded areas are ±SEM.

areas such as the PFC, where mismatch signals interact with model representations to promote model update. To address this issue, we studied the overlap of global deviant and context representations in the PFC. We projected the unfolding population MUA onto the vector corresponding to maximal global deviance decoding performance (±100 ms around the maximum time bin) (Figures 4I and 4J). We found that the resulting population trajectories also segregated as a function of the global context, i.e., xx vs. xY blocks (Figures 4I and 4J). During the first three items, there was a slightly larger population trajectory deviation into the direction of global deviance on xY blocks than on xx blocks. This was reflected by a significant AUROC for context decoding from these trajectories (Figures 4I and 4J [bottom], obtained via random permutation test and cluster-based correction for multiple comparisons; STAR Methods). It is important to note that, conversely, this effect of context cannot entirely explain the deviance response to rare xx trials because the latter elicited an additional trajectory deviation after the last stimulus (Figures 4I and 4J [top], cyan). We therefore conclude that at least part of the population code for global deviance detection was shared with the representation of global context.

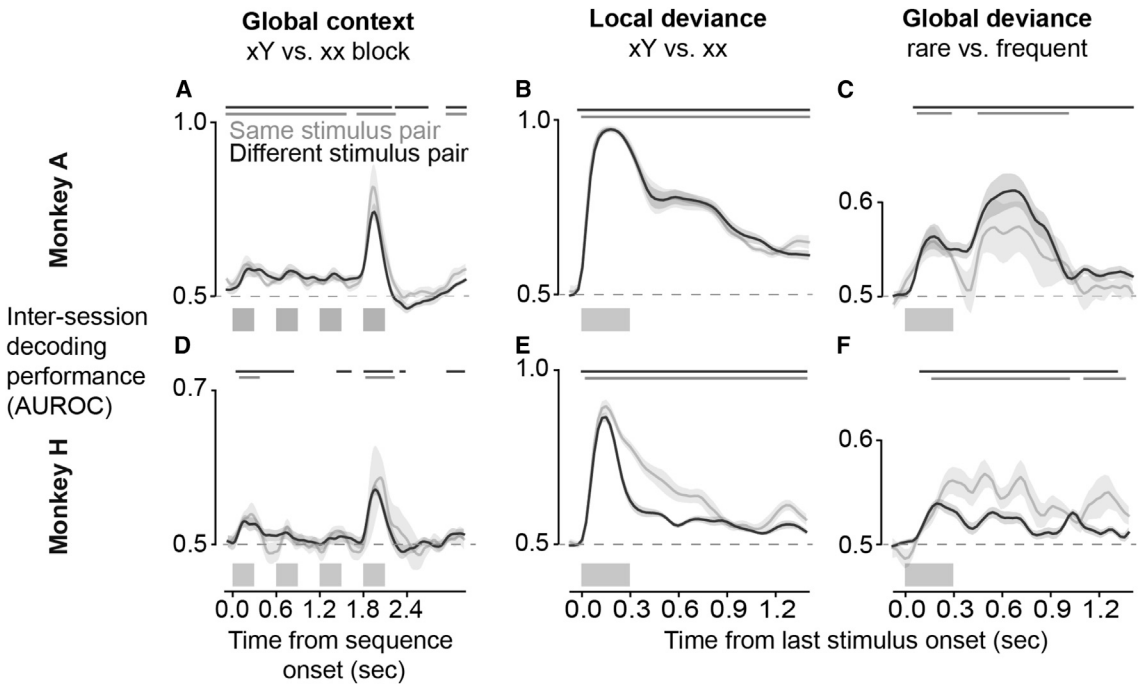


Figure 5. Population responses to global context, local deviance, and global deviance generalize to new sessions and stimuli

Plots show the generalization of decoding to new sessions with either the same visual stimuli (gray curve) or difference stimuli (black curve). Generalization performance (AUROC) is shown separately for decoders trained on global context (A and D), local deviance (B and E), and global deviance (C and F). Horizontal lines on top of each graph show time points where the decoding performance was significantly above chance ($p < 0.05$).

The population code for sequence structure generalizes across stimulus identities

The above results establish that vlPFC population activity spontaneously forms partially overlapping representations of image identity, serial order, global context, local deviance, and global deviance. Furthermore, the above analyses were applied across the two pictures that were presented in each session (A and B), hinting at neural codes that generalize across stimulus identities. To further evaluate the abstractness of these representations, we exploited the chronic nature of recordings and tested the generalization of the population codes across sessions with different picture pairs, always presented on different days. We found that the same neuronal population vectors allowed us to decode global sequence context, local deviants, and global deviants, respectively, even for stimulus pairs that differed completely from those presented in the training session (Figure 5). This finding indicates that vlPFC sequence representations were stable for multiple days and, most importantly, were reinstated across distinct picture identities, thus possibly reflecting an abstract neural code. Although the decoding performance was significantly above chance levels, it was not perfect. This may indicate that part of the representation for sequence structure is sequence specific, or it may reflect neural drift or instabilities in recording from the same neuronal populations across days. Furthermore, differences in the generalization performance of the neuronal populations between the two animals (Figure 5B vs. Figures 5E and 5C vs. Figure 5F) might relate to individual differences in attention, learning the structure, or both. It is also likely that the reduced number of recorded units in the second animal could have contributed to these differences in generalization performance.

Rather than a difference between predicted and seen pictures, the representation of local deviance could reflect the indirect effect of stimulus-specific adaptation (SSA) occurring at an earlier stage such as inferotemporal cortex (IT). Neural responses would be smaller on BBBB trials than on AAAB trials because the response to picture B would have been adapted.[31,32] Note that SSA in IT is picture specific,[33] but if the signal from multiple picture-specific neurons in IT was integrated in the vlPFC, it might explain the observed generalization across different pictures. To test whether the vlPFC population response to local deviants reflected genuine deviance detection or merely SSA, we performed an additional experiment in monkey A, presenting new random sequences with different numbers of image repetitions and changes (Figure 6A). Contrasting the activation evoked by the last stimulus in XXXY sequences with the last stimulus in WXYZ sequences allowed us to disentangle deviance and SSA, as done with the many-standards control in mismatch negativity (MMN) studies.[34] In the many-standards control paradigm, deviance detection predicts a novelty response to XXXY (where a prediction develops about X) but none to WXYZ (in which image identity changes on every trial and is unpredictable); SSA, however, predicts no difference, as the last picture is equally novel and non-adapted in both cases. The results supported deviance detection (Figures 6A and 6B): local deviants always led to a larger response of this neural subpopulation (Figures 6C and

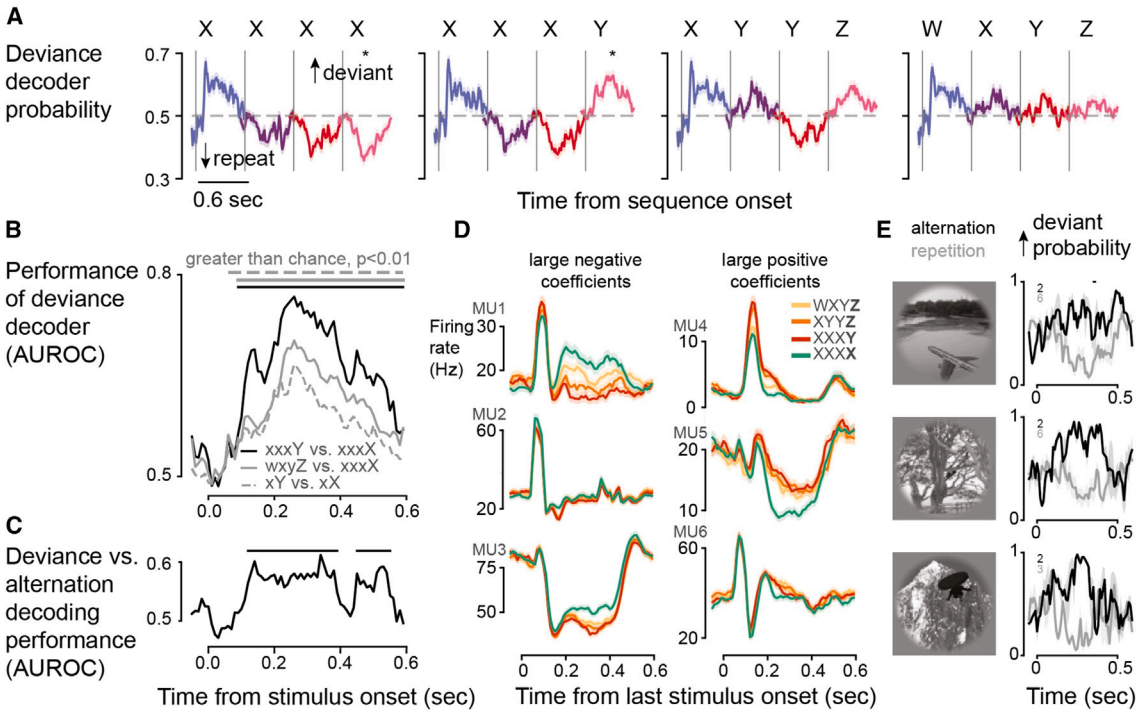


Figure 6. Abstract change and deviance detection by neural populations in monkey A

In a control experiment, four possible sequence chunks (see titles in A) were presented in a uniform random manner. Letters W–Z indicate any of 948 grayscale images from the Brainscore database,[35] changing randomly in each trial. A decoder for local deviance was trained on XXXY vs. XXXX trials (asterisk in A), using leave-one-stimulus-out cross-validation.

(A) Predictive probability of the decoder for all sequence types and stimulus positions in a sequence (indicated by colors).

(B) Decoding performance in terms of AUROC for local deviants (XXXY vs. XXXX, black), novel stimulus, as well as vs. pure repeats (WXYZ vs. XXXX, gray) or any transition vs. single repeats (XY vs. XX, dashed gray). All conditions could be decoded above chance level with $p < 0.01$ (random permutation test). Horizontal bars indicate significant time bins.

(C) Rare stimuli violating a local pattern of repetitions (XXXY) yielded a significantly higher response of this population than rare stimuli without preceding regularity (WXYZ), indicated by horizontal bars on top.

(D) Examples of multiple units contributing to the population axis coding for local deviance. A large negative (left) or positive (right) coefficient means a lower or higher firing rate for deviant stimuli, respectively.

(E) Population deviance response to repeats (light gray) or alternations (black) of three example images. Lines show mean across N number of trials (indicated by small numbers in each plot), and shaded areas show ±SEM.

S6), indicating that the vlPFC indeed encodes deviance from a local context, regardless of picture identity. Indeed, we found that MUA responses underlying this population subspace were far more diverse than what would be expected based on the adaptation hypothesis (Figure 6D), namely a simple decrease in response amplitude upon repetition. This provides additional evidence that PFC populations spontaneously encode sequence deviance in an abstract way, as previously inferred indirectly through brain-imaging signals.[11]

DISCUSSION

Our results suggest that prefrontal population activity encodes concrete, event-specific features of sequential information such as image identity, but also all other abstract aspects of visual sequences such as the serial position of each picture and/or task phase, the global sequence pattern, and local and global sequence structure violations. The population code for global sequence patterns generalized across similar pattern sequences, suggesting that it encoded abstract structural knowledge. Importantly, these representations emerged spontaneously during passive viewing in the absence of any active task that could have resulted in the association of specific sequences with behavioral responses or other contingencies.

Abstract sequence processing in the PFC

Single neurons in the PFC encode abstract information about stimulus category,[36] number,[37] and rules.[5] More recently, the development of multichannel recording techniques and machine-learning analytical tools has allowed for the sampling and analysis of the activity of large neuronal populations. This resulted in the re-emergence of the Hebbian concept of neural assembly,[38] in which the joint neuronal population activity constitutes the basic computation unit in the brain.[16,23,39] Abstract coding in the PFC has been recently probed at this level of ensemble activity. Population activity in two frontal areas, the orbitofrontal and anterior cingulate cortex, was shown to encode abstract information in a task where stimulus identification and

knowledge of context were both necessary to predict reward.[40] Population activity in the PFC (but also in the hippocampus) was shown to have a low-dimensional geometrical structure that allowed the linear classification of abstract features defined by the association of a stimulus with its operant and reinforcement contingencies.[1] At the same time, the code was high-dimensional enough to allow the classification of diverse combinations of specific inputs. The same geometrical properties exist in the human medial frontal cortex for cognitive control, where the population code was both task-general for performance-monitoring signals and task-specific in two paradigms, the Stroop task and a multisource interference task.[41]

Our results demonstrate the ability of the prefrontal population code to represent simultaneously such abstract and content-specific variables spontaneously, through mere exposure, in a no-report sequence paradigm. We suggest that this mechanism could be the basic computational unit mediating the perception of sequences. Learning of abstract features can therefore occur spontaneously through observation without specific instructions or associations of stimuli with rewards or other contingencies.

Implications for predictive processing

We can conceptualize these prefrontal representations of sequences as internal mental models that may also operate to generate predictions onto the external world.[42–46] In previous work, the implementation of such mental models and predictive-coding computations was inferred indirectly from macroscopic brain signals related to sequence violations.[11,18,29,47,48] These studies were useful in guiding the location of our intracortical recordings, which in turn revealed in much higher detail the neural code of these computations. Our results show that even under passive sensory stimulation conditions, assemblies of PFC neurons form abstract mental models of statistical regularities that are frequently encountered in the sensory input. The function of these models could be to project predictions that bias perception, guide decisions, and facilitate sensory processing.[9,49,50]

In such predictive-coding models, the incongruence of sensory information with an already established mental model results in a surprise signal that serves to update the model.[49,51] Based on this main hypothesis of predictive coding, deviations from learned stimulus sequences of variable complexity have been used to probe the existence of multiple levels of predictive representations in the brain, showing that these models unfold at multiple hierarchical scales, reflecting different levels of abstraction and engaging different cortical areas.[9] In general, sequence violations result in mesoscale error signals with faster latencies for low-order violations than for higher-order violations.[17,18] These studies inferred that the long-latency, higher-order error signals arose from the PFC, in agreement with recent modeling work in which predictive deep neural networks develop latent image representations in higher areas of a multilayer architecture.[52] Indeed, our findings showing longer-latency error signals in spiking activity for higher-order sequence violation agree with these predictions.

The most commonly studied brain signal to sequence deviants is mismatch negativity (MMN).[53] In the classical auditory oddball paradigm, a local regularity is established by the repetition of one tone, which leads to a reduction in brain responses; at this point, the presentation of a rare (deviant) tone yields the MMN signal, a stronger negative deflection of the scalp electroencephalography compared to repeated stimuli.[54] Although feedforward adaptation may contribute to the MMN,[31,32] several results indicate that it, and similar visual violation responses,[55–58] primarily reflect predictive processing.[31,46,59,60]

The source of auditory MMN has been localized in the secondary auditory cortex, superior temporal gyrus, and PFC.[61–63]

Apart from local regularities, humans and non-human primates can infer sequence regularities at a more temporally extended and abstract level, for instance the extraction of the same structure or pattern (e.g., xxY) from sequences comprising different stimuli (e.g., AAB and CCD).[11] Processing such higher-order sequential structures necessarily abstracts away from the specific stimulus identities and may therefore engage higher-order association cortical areas, in particular the PFC. We have previously studied these two levels of sequence processing using a hierarchical "local-global" sequence paradigm where auditory stimuli are presented in short sequences while measuring macroscopic brain signals. During an entire block of trials, all sequences follow the same pattern, either four repeats of the same stimulus (e.g., AAAA) or three repeats followed by a deviant (e.g., AAAB). Following the repeated presentation of one of these global sequence patterns, the ability to detect global violations is probed by presenting sequences that deviate from the pattern (AAAB or AAAA, respectively; global deviants). In contrast to the MMN, mismatch responses to such global deviants are delayed, have been assumed to require consciousness, and predominantly arise from higher-order cortical areas of humans and macaques,[17,18,29,64] including the vlPFC. Intracranial recordings of global field potentials indicate that the PFC sends feedback signals to upstream cortical areas after the violation of global sequence expectations.[18] From the observation of such mismatch signals, it was indirectly inferred that the PFC and associated high-level areas constitute an internal mental model of the ongoing sequences. We demonstrated here that neural populations in the vlPFC indeed encoded abstract global sequence violations.

A central hypothesis of Bayesian predictive processing models is a functional segregation of predictions that account for the causes of sensory inputs[65–67] or expectations and error representations, namely different neuronal populations representing internal models about sensory input and their update, respectively.[30,51,68–70] Our results confirm the existence of both error (mismatch) and expectation (or prediction) signals in the spiking activity of the PFC, as predicted by the theory and in agreement with macroscopic signals recorded with electrocorticography in the primate brain.[18,71] However, we traced global sequence context and its mismatches to overlapping population subspaces, indicating an integrative level of predictive processing in the PFC. This observation could imply a revision of current models of predictive-coding circuits that suggest distinguishable neuronal activity in higher-order cortical areas (including the PFC) for putative predictions and prediction errors.[72,73]

Frontal cortical areas have also been shown to provide contextual information in the rodent V1,[74] possibly through gain modulation of neuronal activity.[75] In our recordings, contextual and mismatch signals were coded by activity vectors (population subspaces) within the same neural population. Specifically, the

prefrontal neural population that emitted abstract responses to pure global deviations from a learned sequence also represented the global context. This convergence of predictive-coding computations at the population level may be a particular feature of PFC neurons, which are known for their role in integrative processes and mixed selectivity properties.[15,24] Interestingly, axons from the anterior cingulate cortex, a medial frontal cortical area, projecting to V1 in the rodent brain were not found to be modulated by deviant stimuli.[74] This discrepancy might point to a different role of the primate PFC, where we found responses to deviant stimuli, compared to rodent frontal cortex. Alternatively, this could indicate a functional specialization of primate frontal cortical areas, with the vlPFC more likely to reflect both types of responses (deviant and context). It is worth noting that, in an auditory oddball paradigm, context-dependent mismatch responses were indeed observed in the rat PFC.[76]

An abstract neural code for stimulus transitions

With our control experiment, we found that the neural subspace that conveyed local deviance responses was also activated for any stimulus change, independently of image identity. Such a response could stem from two distinct mechanisms: a higher-order process of prediction error, with neural responses representing the violation of a top-down prediction, i.e., the difference between what was observed and what was predicted,[18,46,74,76,77] or alternatively a passive feedforward process of SSA.[32,78,79] Our results falsify the hypothesis that SSA could explain the observed mismatch detection response, since a rare local deviant (last stimulus in XXXY trials) elicited a stronger population response than an equally rare stimulus that was not preceded by a predictable context (WXYZ trials; Figure 6). However, each stimulus in a WXYZ sequence still elicited a response of the "deviance population." As this code generalized across images (Figure 6B, light-gray lines and dashed gray lines), it can be interpreted as an abstract code for any stimulus change, thus encoding whether the present image was novel or was repeated. The response of this "change detector" population decreased with every transition (Figure 6A) and was weaker for a transition following repetitions (XXXY) than for the fourth transition of a sequence (WXYZ). Hence, the adaptation of neurons that encode abstract change information could serve as a building block to encode other structural variables, such as deviance or number.

Implications for perceptual inference and conscious perception

It has been suggested that PFC participates in a global neuronal workspace (GNW) that represents any conscious experience.[80,81] Indeed, several studies indicate that the contents of visual consciousness can be decoded from prefrontal neuronal activity and that PFC activity signals whether a stimulus was consciously perceived.[7,8,82,83] The present results are congruent with the GNW hypothesis, since they indicate that the PFC contains superimposed neural codes for all features of the perceived visual sequences, whether concrete or abstract. Some of these features encode expectations of the upcoming sequence, a finding that fits with prior evidence that PFC anticipatory signals may be critical for detecting perceptual ambiguity and biasing conscious perception through ongoing fluctuations[83–85] or for the provision of perceptual hypotheses.[46,86] Our results showing decoding of context from prefrontal populations indeed suggest a strong influence of expectation in the activity of prefrontal ensembles during conscious visual perception. Importantly, due to the passive viewing and the absence of a motor-related task, these prefrontal representations could be viewed under the prism of predictive processing theories of consciousness (e.g., neurorepresentationalism) that, in contrast to active inference,[77] do not consider behavioral output as critical for realizing phenomenal experience.[67,87–89]

Limitations of the study

Generalization of decoding provides strong evidence for the application of the same internal models to novel sensory sequences with the same structure, a signature of abstract processing. However, due to the rather low levels of complexity in the local-global task, it will be important to test the validity of our observations in sequences with higher levels of complexity at several steps of the hierarchy recently proposed.[9,90,91]

Furthermore, the local-global paradigm has limitations in probing conscious perception, since it does not include a condition where the sequences are perceptually suppressed and fall below the threshold for conscious perception. Including such a contrastive paradigm as well as testing whether image identification is influenced by features that are processed consciously or not could test the relation of the observed activity patterns to conscious perception. Also, a comparison between a no-report and report version of the paradigm could separate perceptual from post-perceptual effects in population decoding.

As only male animals participated in this study, the sex of the animals could also reflect a limitation of our study.

Another limitation of our study relates to potential confounds of anticipation of reward and orofacial movements on neural patterns[92] due to the connectivity of PFC with mouth representations in the primary motor cortex.[93] Changing the time of reward following the last stimulus offset could resolve this issue in future studies.

Finally, it remains to be determined how these PFC signals influence downstream or upstream cortical areas. Future studies, recording simultaneously from multiple cortical areas and layers, could probe how inter-areal communication contributes to predictive processing.

Conclusion

Our results reveal that the PFC, even in the absence of overt behavior, spontaneously encodes visual sequences at multiple levels, including abstract codes for serial position and/or task phase and sequence pattern. The neural code relies on a superimposition of multiple vector subspaces, each representing a specific dimension of the perceived sequence.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.113952.

ACKNOWLEDGMENTS

We thank Julien Lemaitre for veterinary care, Jean-Robert Deverre, Lionel Allirol, Fabrice Poupon, and the NeuroSpin support cells for help in setting up the lab, Abhilash Dwarakanath for his help with the data preprocessing pipeline, and Pierre-Louis Bellet for providing the brain graphic of Figure 1. We also thank the anonymous reviewers for their constructive comments on an earlier version of the manuscript and Andre Bastos for useful feedback. This research was supported by INSERM, CEA, Université Paris-Saclay, Collège de France, and Région Ile-de-France, and received specific funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under specific grant agreement no. 945539 (Human Brain Project SGA3). T.v.K. is supported by the European Union (ERC starting grant, TDVision, 101078667). T.I.P. is supported by the Templeton World Charity Foundation (TWCF0562).

AUTHOR CONTRIBUTIONS

Conceptualization, M.E.B., B.J., S.D., T.v.K., and T.I.P., experimental design, M.E.B., J.B., T.v.K., and T.I.P.; analyses, M.E.B. and J.B.; data acquisition, M.G., J.B., and T.I.P.; Utah array implantation, M.G., B.J., and T.I.P.; writing, M.E.B., T.v.K., S.D., and T.I.P.; visualization, M.E.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 9, 2022
Revised: June 6, 2023
Accepted: February 27, 2024
Published: March 13, 2024

REFERENCES

1. Bernardi, S., Benna, M.K., Rigotti, M., Munuera, J., Fusi, S., and Salzman, C.D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. Cell *183*, 954–967.e21. https://doi.org/10.1016/j.cell.2020.09.031.
2. Chafee, M.V., and Heilbronner, S.R. (2022). Prefrontal cortex. Curr. Biol. *32*, R346–R351. https://doi.org/10.1016/j.cub.2022.02.071.
3. Xie, Y., Hu, P., Li, J., Chen, J., Song, W., Wang, X.-J., Yang, T., Dehaene, S., Tang, S., Min, B., and Wang, L. (2022). Geometry of sequence working memory in macaque prefrontal cortex. Science *375*, 632–639. https://doi.org/10.1126/science.abm0204.
4. Miller, E.K., and Cohen, J.D. (2001). An Integrative Theory of Prefrontal Cortex Function. Annu. Rev. Neurosci. *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167.
5. Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956. https://doi.org/10.1038/35082081.
6. Bellet, J., Gay, M., Dwarakanath, A., Jarraya, B., Van Kerkoerle, T., Dehaene, S., and Panagiotaropoulos, T.I. (2022). Decoding rapidly presented visual stimuli from prefrontal ensembles without report nor post-perceptual processing. Neurosci. Conscious. *2022*, niac005. https://doi.org/10.1093/nc/niac005.
7. Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T.I., and Logothetis, N.K. (2022). Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports. Nat. Commun. *13*, 1535. https://doi.org/10.1038/s41467-022-28897-2.
8. Panagiotaropoulos, T.I., Deco, G., Kapoor, V., and Logothetis, N.K. (2012). Neuronal Discharges and Gamma Oscillations Explicitly Reflect Visual Consciousness in the Lateral Prefrontal Cortex. Neuron *74*, 924–935. https://doi.org/10.1016/j.neuron.2012.04.013.
9. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. Neuron *88*, 2–19. https://doi.org/10.1016/j.neuron.2015.09.019.
10. Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. Proc. Natl. Acad. Sci. USA *109*, 11860–11865. https://doi.org/10.1073/pnas.1204580109.
11. Wang, L., Uhrig, L., Jarraya, B., and Dehaene, S. (2015). Representation of Numerical and Sequential Patterns in Macaque and Human Brains. Curr. Biol. *25*, 1966–1974. https://doi.org/10.1016/j.cub.2015.06.035.
12. Badre, D., Bhandari, A., Keglovits, H., and Kikumoto, A. (2021). The dimensionality of neural representations for control. Curr. Opin. Behav. Sci. *38*, 20–28. https://doi.org/10.1016/j.cobeha.2020.07.002.
13. Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. Curr. Opin. Neurobiol. *37*, 66–74. https://doi.org/10.1016/j.conb.2016.01.010.
14. Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84. https://doi.org/10.1038/nature12742.
15. Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590. https://doi.org/10.1038/nature12160.
16. Saxena, S., and Cunningham, J.P. (2019). Towards the neural population doctrine. Curr. Opin. Neurobiol. *55*, 103–111. https://doi.org/10.1016/j.conb.2019.02.002.
17. Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. Proc. Natl. Acad. Sci. USA *106*, 1672–1677. https://doi.org/10.1073/pnas.0809667106.
18. Chao, Z.C., Takaura, K., Wang, L., Fujii, N., and Dehaene, S. (2018). Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. Neuron *100*, 1252–1266.e3. https://doi.org/10.1016/j.neuron.2018.10.004.

19. Dang, W., Jaffe, R.J., Qi, X.-L., and Constantinidis, C. (2021). Emergence of Non-Linear Mixed Selectivity in Prefrontal Cortex after Training. J. Neurosci. *41*, 7420–7434, JN-RM-2814-20. https://doi.org/10.1523/JNEUROSCI.2814-20.2021.

20. Dang, W., Li, S., Pu, S., Qi, X.-L., and Constantinidis, C. (2022). More Prominent Nonlinear Mixed Selectivity in the Dorsolateral Prefrontal than Posterior Parietal Cortex. eNeuro *9*, ENEURO.0517-0521.2022. https://doi.org/10.1523/ENEURO.0517-21.2022.

21. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. Roy. Stat. Soc. B *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

22. Baeg, E.H., Kim, Y.B., Huh, K., Mook-Jung, I., Kim, H.T., and Jung, M.W. (2003). Dynamics of Population Code for Working Memory in the Prefrontal Cortex. Neuron *40*, 177–188. https://doi.org/10.1016/S0896-6273(03)00597-X.

23. Ebitz, R.B., and Hayden, B.Y. (2021). The population doctrine in cognitive neuroscience. Neuron *109*, 3055–3068. https://doi.org/10.1016/j.neuron.2021.07.011.

24. Parthasarathy, A., Herikstad, R., Bong, J.H., Medina, F.S., Libedinsky, C., and Yen, S.-C. (2017). Mixed selectivity morphs population codes in prefrontal cortex. Nat. Neurosci. *20*, 1770–1779. https://doi.org/10.1038/s41593-017-0003-2.

25. Chen, S., Swartz, K.B., and Terrace, H.S. (1997). Knowledge of the Ordinal Position of List Items in Rhesus Monkeys. Psychol. Sci. *8*, 80–86. https://doi.org/10.1111/j.1467-9280.1997.tb00687.x.

26. Orlov, T., Yakovlev, V., Hochstein, S., and Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. Nature *404*, 77–80. https://doi.org/10.1038/35003571.

27. Terrace, H.S., Son, L.K., and Brannon, E.M. (2003). Serial Expertise of Rhesus Macaques. Psychol. Sci. *14*, 66–73. https://doi.org/10.1111/1467-9280.01420.

28. Kapoor, V., Besserve, M., Logothetis, N.K., and Panagiotaropoulos, T.I. (2018). Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. Commun. Biol. *1*, 215. https://doi.org/10.1038/s42003-018-0225-1.

29. Uhrig, L., Dehaene, S., and Jarraya, B. (2014). A Hierarchy of Responses to Auditory Regularities in the Macaque Brain. J. Neurosci. *34*, 1127–1132. https://doi.org/10.1523/JNEUROSCI.3165-13.2014.

30. Walsh, K.S., McGovern, D.P., Clark, A., and O'Connell, R.G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. Ann. N. Y. Acad. Sci. *1464*, 242–268. https://doi.org/10.1111/nyas.14321.

31. Garrido, M.I., Kilner, J.M., Stephan, K.E., and Friston, K.J. (2009). The mismatch negativity: A review of underlying mechanisms. Clin. Neurophysiol. *120*, 453–463. https://doi.org/10.1016/j.clinph.2008.11.029.

32. May, P.J.C., and Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology *47*, 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x.

33. Meyer, T., and Olson, C.R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. Proc. Natl. Acad. Sci. USA *108*, 19401–19406. https://doi.org/10.1073/pnas.1112895108.

34. Ruhnau, P., Herrmann, B., and Schröger, E. (2012). Finding the right control: The mismatch negativity under investigation. Clin. Neurophysiol. *123*, 507–512. https://doi.org/10.1016/j.clinph.2011.07.035.

35. Majaj, N.J., Hong, H., Solomon, E.A., and DiCarlo, J.J. (2015). Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. J. Neurosci. *35*, 13402–13418. https://doi.org/10.1523/JNEUROSCI.5181-14.2015.

36. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. Science *291*, 312–316. https://doi.org/10.1126/science.291.5502.312.

37. Nieder, A., Freedman, D.J., and Miller, E.K. (2002). Representation of the Quantity of Visual Items in the Primate Prefrontal Cortex. Science *297*, 1708–1711. https://doi.org/10.1126/science.1072493.

38. Hebb, D.O. (2005). The Organization of Behavior 0, ed. (Psychology Press). https://doi.org/10.4324/9781410612403.

39. Yuste, R. (2015). From the neuron doctrine to neural networks. Nat. Rev. Neurosci. *16*, 487–497. https://doi.org/10.1038/nrn3962.

40. Saez, A., Rigotti, M., Ostojic, S., Fusi, S., and Salzman, C.D. (2015). Abstract Context Representations in Primate Amygdala and Prefrontal Cortex. Neuron *87*, 869–881. https://doi.org/10.1016/j.neuron.2015.07.024.

41. Fu, Z., Beam, D., Chung, J.M., Reed, C.M., Mamelak, A.N., Adolphs, R., and Rutishauser, U. (2022). The geometry of domain-general performance monitoring in the human medial frontal cortex. Science *376*, eabm9922. https://doi.org/10.1126/science.abm9922.

42. Euler, M.J. (2018). Intelligence and uncertainty: Implications of hierarchical predictive processing for the neuroscience of cognitive ability. Neurosci. Biobehav. Rev. *94*, 93–112. https://doi.org/10.1016/j.neubiorev.2018.08.013.

43. Meirhaeghe, N., Sohn, H., and Jazayeri, M. (2021). A precise and adaptive neural mechanism for predictive temporal processing in the frontal cortex. Neuron *109*, 2995–3011.e5. https://doi.org/10.1016/j.neuron.2021.08.025.

44. Pinotsis, D.A., Buschman, T.J., and Miller, E.K. (2019). Working Memory Load Modulates Neuronal Coupling. Cerebr. Cortex *29*, 1670–1681. https://doi.org/10.1093/cercor/bhy065.

45. Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., and Hirsch, J. (2006). Predictive Codes for Forthcoming Perception in the Frontal Cortex. Science *314*, 1311–1314. https://doi.org/10.1126/science.1132028.

46. Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.-M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat. Neurosci. *11*, 1004–1006. https://doi.org/10.1038/nn.2163.

47. Gil-da-Costa, R., Stoner, G.R., Fung, R., and Albright, T.D. (2013). Nonhuman primate model of schizophrenia using a noninvasive EEG method. Proc. Natl. Acad. Sci. USA *110*, 15425–15430. https://doi.org/10.1073/pnas.1312264110.

48. Wilson, B., Marslen-Wilson, W.D., and Petkov, C.I. (2017). Conserved Sequence Processing in Primate Frontal Cortex. Trends Neurosci. *40*, 72–82. https://doi.org/10.1016/j.tins.2016.11.004.

49. Friston, K. (2005). A theory of cortical responses. Phil. Trans. R. Soc. B *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622.

50. Summerfield, C., and De Lange, F.P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. Nat. Rev. Neurosci. *15*, 745–756. https://doi.org/10.1038/nrn3838.

51. 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. https://doi.org/10.1038/4580.

52. Dora, S., Bohte, S.M., and Pennartz, C.M.A. (2021). Deep Gated Hebbian Predictive Coding Accounts for Emergence of Complex Neural Response Properties Along the Visual Cortical Hierarchy. Front. Comput. Neurosci. *15*, 666131. https://doi.org/10.3389/fncom.2021.666131.

53. Näätänen, R., Gaillard, A.W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. *42*, 313–329. https://doi.org/10.1016/0001-6918(78)90006-9.

54. Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clin. Neurophysiol. *118*, 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.

55. Pazo-Alvarez, P., Cadaveira, F., and Amenedo, E. (2003). MMN in the visual modality: a review. Biol. Psychol. *63*, 199–236. https://doi.org/10.1016/S0301-0511(03)00049-8.

56. Keller, G.B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. Neuron *74*, 809–815. https://doi.org/10.1016/j.neuron.2012.03.040.

57. Zmarz, P., and Keller, G.B. (2016). Mismatch Receptive Fields in Mouse Visual Cortex. Neuron *92*, 766–772. https://doi.org/10.1016/j.neuron.2016.09.057.

58. Hamm, J.P., and Yuste, R. (2016). Somatostatin Interneurons Control a Key Component of Mismatch Negativity in Mouse Visual Cortex. Cell Rep. *16*, 597–604. https://doi.org/10.1016/j.celrep.2016.06.037.

59. Wacongne, C., Changeux, J.-P., and Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. J. Neurosci. *32*, 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012.

60. Winkler, I. (2007). Interpreting the Mismatch Negativity. J. Psychophysiol. *21*, 147–163. https://doi.org/10.1027/0269-8803.21.34.147.

61. Deouell, L.Y. (2007). The Frontal Generator of the Mismatch Negativity Revisited. J. Psychophysiol. *21*, 188–203. https://doi.org/10.1027/0269-8803.21.34.188.

62. Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H.-J., Kirsch, H.E., Dalal, S.S., Deouell, L.Y., and Knight, R.T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc. Natl. Acad. Sci. USA *113*, 6755–6760. https://doi.org/10.1073/pnas.1525030113.

63. Shalgi, S., and Deouell, L.Y. (2007). Direct evidence for differential roles of temporal and frontal components of auditory change detection. Neuropsychologia *45*, 1878–1888. https://doi.org/10.1016/j.neuropsychologia.2006.11.023.

64. El Karoui, I., King, J.-R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., et al. (2015). Event-Related Potential, Time-frequency, and Functional Connectivity Facets of Local and Global Auditory Novelty Processing: An Intracranial Study in Humans. Cerebr. Cortex *25*, 4203–4212. https://doi.org/10.1093/cercor/bhu143.

65. von Helmholtz, H. (1867). Handbuch der Physiologischen Optik (Voss).

66. Gregory, R.L. (1980). Perceptions as hypotheses. Phil. Trans. Roy. Soc. Lond. B *290*, 181–197. https://doi.org/10.1098/rstb.1980.0090.

67. Pennartz, C.M.A. (2015). The Brain's Representational Power: On Consciousness and the Integration of Modalities (The MIT Press). https://doi.org/10.7551/mitpress/9780262029315.001.0001.

68. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical Microcircuits for Predictive Coding. Neuron *76*, 695–711. https://doi.org/10.1016/j.neuron.2012.10.038.

69. Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. *11*, 127–138. https://doi.org/10.1038/nrn2787.

70. Friston, K. (2018). Does predictive coding have a future? Nat. Neurosci. *21*, 1019–1021. https://doi.org/10.1038/s41593-018-0200-7.

71. Dürschmid, S., Reichert, C., Hinrichs, H., Heinze, H.-J., Kirsch, H.E., Knight, R.T., and Deouell, L.Y. (2019). Direct Evidence for Prediction Signals in Frontal Cortex Independent of Prediction Error. Cerebr. Cortex *29*, 4530–4538. https://doi.org/10.1093/cercor/bhy331.

72. Bell, A.H., Summerfield, C., Morin, E.L., Malecek, N.J., and Ungerleider, L.G. (2016). Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. Curr. Biol. *26*, 2280–2290. https://doi.org/10.1016/j.cub.2016.07.007.

73. Bastos, A.M., Lundqvist, M., Waite, A.S., Kopell, N., and Miller, E.K. (2020). Layer and rhythm specificity for predictive routing. Proc. Natl. Acad. Sci. USA *117*, 31459–31469. https://doi.org/10.1073/pnas.2014868117.

74. Hamm, J.P., Shymkiv, Y., Han, S., Yang, W., and Yuste, R. (2021). Cortical ensembles selective for context. Proc. Natl. Acad. Sci. USA *118*, e2026179118. https://doi.org/10.1073/pnas.2026179118.

75. Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science *345*, 660–665. https://doi.org/10.1126/science.1254126.

76. Casado-Román, L., Carbajal, G.V., Pérez-González, D., and Malmierca, M.S. (2020). Prediction error signaling explains neuronal mismatch responses in the medial prefrontal cortex. PLoS Biol. *18*, e3001019. https://doi.org/10.1371/journal.pbio.3001019.

77. Parras, G.G., Nieto-Diego, J., Carbajal, G.V., Valdés-Baizabal, C., Escera, C., and Malmierca, M.S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nat. Commun. *8*, 2148. https://doi.org/10.1038/s41467-017-02038-6.

78. Fishman, Y.I., and Steinschneider, M. (2012). Searching for the Mismatch Negativity in Primary Auditory Cortex of the Awake Monkey: Deviance Detection or Stimulus Specific Adaptation? J. Neurosci. *32*, 15747–15758. https://doi.org/10.1523/JNEUROSCI.2835-12.2012.

79. Kaliukhovich, D.A., and Vogels, R. (2014). Neurons in Macaque Inferior Temporal Cortex Show No Surprise Response to Deviants in Visual Oddball Sequences. J. Neurosci. *34*, 12801–12815. https://doi.org/10.1523/JNEUROSCI.2154-14.2014.

80. Dehaene, S., and Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. Neuron *70*, 200–227. https://doi.org/10.1016/j.neuron.2011.03.018.

81. Dehaene, S., Kerszberg, M., and Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. Proc. Natl. Acad. Sci. USA *95*, 14529–14534. https://doi.org/10.1073/pnas.95.24.14529.

82. Levinson, M., Podvalny, E., Baete, S.H., and He, B.J. (2021). Cortical and subcortical signatures of conscious object recognition. Nat. Commun. *12*, 2930. https://doi.org/10.1038/s41467-021-23266-x.

83. Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P.R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. Science *360*, 537–542. https://doi.org/10.1126/science.aar7186.

84. Dwarakanath, A., Kapoor, V., Werner, J., Safavi, S., Fedorov, L.A., Logothetis, N.K., and Panagiotaropoulos, T.I. (2023). Bistability of prefrontal states gates access to consciousness. Neuron *111*, 1666–1683.e4. https://doi.org/10.1016/j.neuron.2023.02.027.

85. Moutard, C., Dehaene, S., and Malach, R. (2015). Spontaneous Fluctuations and Non-linear Ignitions: Two Dynamic Faces of Cortical Recurrent Loops. Neuron *88*, 194–206. https://doi.org/10.1016/j.neuron.2015.09.018.

86. Weilnhammer, V., Fritsch, M., Chikermane, M., Eckert, A.-L., Kanthak, K., Stuke, H., Kaminski, J., and Sterzer, P. (2021). An active role of inferior frontal cortex in conscious experience. Curr. Biol. *31*, 2868–2880.e8. https://doi.org/10.1016/j.cub.2021.04.043.

87. Seth, A.K., and Bayne, T. (2022). Theories of consciousness. Nat. Rev. Neurosci. *23*, 439–452. https://doi.org/10.1038/s41583-022-00587-4.

88. Pennartz, C.M.A. (2018). Consciousness, Representation, Action: The Importance of Being Goal-Directed. Trends Cognit. Sci. *22*, 137–153. https://doi.org/10.1016/j.tics.2017.10.006.

89. Pennartz, C.M.A. (2022). What is neurorepresentationalism? From neural activity and predictive processing to multi-level representations and consciousness. Behav. Brain Res. *432*, 113969. https://doi.org/10.1016/j.bbr.2022.113969.

90. Planton, S., Van Kerkoerle, T., Abbih, L., Maheu, M., Meyniel, F., Sigman, M., Wang, L., Figueira, S., Romano, S., and Dehaene, S. (2021). A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. PLoS Comput. Biol. *17*, e1008598. https://doi.org/10.1371/journal.pcbi.1008598.

91. Al Roumi, F., Planton, S., Wang, L., and Dehaene, S. (2023). Brain-imaging evidence for compression of binary sound sequences in human memory. Elife *12*, e84376. https://doi.org/10.7554/eLife.84376.

92. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brain-wide activity. Science *364*, 255. https://doi.org/10.1126/science.aav7893.

93. Miyachi, S., Hirata, Y., Inoue, K.i., Lu, X., Nambu, A., and Takada, M. (2013). Multisynaptic projections from the ventrolateral prefrontal cortex to hand and mouth representations of the monkey primary motor cortex. Neurosci. Res. *76*, 141–149. https://doi.org/10.1016/j.neures.2013.04.004.

94. Hill, D.N., Mehta, S.B., and Kleinfeld, D. (2011). Quality Metrics to Accompany Spike Sorting of Extracellular Signals. J. Neurosci. *31*, 8699–8705. https://doi.org/10.1523/JNEUROSCI.0971-11.2011.

95. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Ketamine	Virbac	Ketamine 1000 (100 mg/mL)
Dexmedetomidine	Orion Pharma	Dexdomitor 0.5 mg/mL
Methylprednisolone	Pfizer	Solu-Medrol 120 mg
Dexamethasone	MSD	Dexadreson 2 mg/ml
Cefazoline	Mylan	Cefazoline 1 g
Buprenorphine	Axience	Buprecare 0.3 mg/mL
Deposited data		
PFC data from local-global experiment	This study, monkey multielectrode array	https://doi.org/10.6084/m9.figshare.24311767.v1
PFC data from control experiment	This study, monkey multielectrode array	https://doi.org/10.6084/m9.figshare.24311929.v1
Experimental models: Organisms/strains		
Rhesus macaques	Austria and China	VI225 and 412061
Software and algorithms		
Code for analysis and figures	This study	https://zenodo.org/records/10665315
MATLAB	MathWorks	Version 2020
Python	Python software foundation	Version 3.10.9
UltraMegasort2000 toolbox	https://github.com/danamics/UMS2K	Version of February 2012
Other		
Stereotactic frame	KOPF Instruments	Model 1530
Microelectrode array	Blackrock Neurotech	Standard Utah array with 96 channels
Cerebus neural signal processor system	Blackrock Microsystems	Cerebus System Digital
Pneumatic inserter	Blackrock Microsystems	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Theofanis I. Panagiotaropoulos (theofanis.panagiotaropoulos@cea.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Preprocessed data have been deposited at Figshare and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Two adult male rhesus macaques (A and H, 9–10 kg, 19 and 16 years old, respectively) participated in this study. Both animals were pair-housed. They had previously been implanted with a custom-made skull-form-specific titanium headpost and trained on a passive visual fixation task with liquid reward in a primate chair. Daily water access was controlled during the experimental period. All

procedures including maintenance and care of the animals were conducted in accordance with the European convention for animal care (86–406) and the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Animal studies were approved by the institutional (CEA) Ethical Committee (CETEA protocol #A18_028).

METHOD DETAILS

Sequence paradigm

We used an adaptation of the local-global paradigm[17] with visual stimuli. The stimuli were 10 colored images of objects, matched in luminance (Figure 1D). Fixed pairs of images were used in every experiment, here denoted as stimulus A and B. The paradigm consisted in the presentation of binary visual sequences composed of 4 items. Each item was displayed for 300 ms, with an inter-stimulus interval (isi) of 300 ms (stimulus onset asynchrony, SOA - 600ms). The sequence could be one of four: aaaa or bbbb, denoted as xxxx; and aaaB or bbbA, denoted as xxxY, where the capital letter indicates a *local deviant* item. One sequence was presented per trial which were organized in blocks of 200 trials. During each block, one sequence was used as the frequent, global standard sequence which was established during 50 habituation trials at the beginning of the block. 80% of the remaining 150 trials were global standards and 20% were global deviants, which differed in the last position compared to the standard. Each of the four sequence types was used as the global standard sequence in one block. We will denote a global context according to its global standard sequence: xxxx block for aaaa-frequent and bbbb-frequent or xxxY block for aaaB-frequent and bbbA-frequent. We will furthermore denote trials according to their context: xY|xx indicates a trial with a local deviant that occurs in a block of frequent xxxx sequences. The four trial types are thus xx|xx, xY|xY (global standards) and xx|xY, xY|xx (global deviants).

This two-by-two design enabled us to study effects of lower-order (local) and higher- order (global) sequence regularity. Consider for instance a single xxxY sequence: it ends with a local deviant, an image that violates the repeated structure of the previous three images. However, assuming that monkeys quickly detected the global sequence regularity in a block, the same local deviant, occurring within a block of similar xY global standard trials, is predictable and may no longer generate a global surprise (xY|xY, "predicted local deviant"). Conversely, a rare xx trial, which does not violate the local context, may elicit a global surprise when presented among many xY sequences (xx|xY, hereafter called "pure global deviant").

Each trial started with the display of a black fixation spot (diameter of 0.3°) at the center of the screen. After 300 ms of fixation by the monkey, the fixation point disappeared and the sequence was displayed centrally with stimuli at a size of 8° per visual angle. The animals had to maintain the gaze within a window of 8 degrees of visual angle centered on the stimulus. A liquid reward was given for trial completion 100 ms after offset of the last item.

For monkey H, 5 stimulus pairs were used during a total of 10 experiments and for monkey A, 4 stimulus pairs were used during a total of 6 experiments.

Repetition vs. change control experiment

We performed two sessions of an additional experiment with monkey A during which we showed four different types of sequence chunks containing each 4 images. The types of sequences were XXXX, XXXY, XYYZ and WXYZ, where letters indicate any of 948 grayscale images of objects from the Brainscore database,[35] randomly changing with each sequence presentation. Sequence types were uniformly distributed across a recording session so that there was no global context. All stimuli could occur in any position and were presented 1 to 5 times. This experiment served to control for stimulus-specific adaptation that could underlie the deviance response to the last stimulus in XXXY trials. As we used a broad range of stimuli, the images used as WXY are expected to be on average as distant from Z than X is from Y in XXXY trials, thereby leading to the same amount of stimulus-specific adaptation for the population responding to Y (in XXXY) as to Z in (WXYZ). The timing of the stimuli were the same as in the local-global experiment and the reward was given at 100 ms after offset of the last stimulus.

Implantation of microelectrode arrays

The macaques were tranquilized in their cage by intramuscular injection of ketamine 1000 (3 mg/kg) and dexmedetomidine (0.015 mg/kg). Once in the operating room, they were placed in a stereotactic frame and deeply anesthetized (assisted respiration) by inhalation of oxygen (20%) and sevoflurane (1.5–2%). An intravenous catheter was installed for the administration of physiological fluids (NaCl with 5% glucose, 10 mL/kg/h). A steroidal anti-inflammatory drug, the methylprednisolone (solumedrol 1 mg/kg i.m.) or dexamethasone (dexazone 0.5 mg/kg i.m.) is administered to prevent swelling of the cortex as well as, an antibiotic (cefazoline 50 mg/kg i.m.) and a morphine derivative (buprecare 0.02 mg/kg i.m.). All surgical procedures were performed aseptically, and recordings of heart rate, respiration patterns, blood pressure and body temperature were monitored throughout the surgery.

Macaques were given a methylprednisolone (monkey A, 1 mg/kg) or dexamethasone dose (monkey H, 0.1 mg/kg) the day before the implantation to avoid brain edema. Monkey H received another dose of dexamethasone (0.5 mg/kg) the day of implantation. The implantation of the gas sterilized multielectrode array began with a longitudinal incision in the skin. The skin and underlying muscle was retracted and a craniotomy was performed over the lateral prefrontal cortex using a surgical drill. The bone flap was removed and then a U-shaped opening in the dura mater was made to expose the cortex. Hyperventilation after dura opening was used to reduce intracranial pressure and avoid swelling of the cortex. The Utah microelectrode array was implanted into the inferior convexity of the prefrontal cortex, 1–2 mm anterior to the bank of the arcuate sulcus and below the ventral bank of the principal sulcus, using a

pneumatic inserter (Blackrock Microsystems). The dimensions of the array was 4 × 4 mm in a 10 × 10 electrode configuration (96 active recording electrodes) resulting in an electrode-to-electrode distance of 400 μm. The electrode length was 1 mm. The titanium connector that can be connected to the electrophysiological recording device was implanted on the skull with titanium screws. Then the dura mater was sewn back together, the bone flap was reinserted and secured by a thin titanium strip. Finally, the skin was sutured. After the electrode array implantation, injections of antibiotics (cefazoline 50 mg/kg i.m.) were given for 10 days and buprenorphine (0.015 mg/kg i.m.) for 3–5 days depending on the pain level.

Data preprocessing

The recorded broadband signals were preprocessed using MATLAB (MathWorks). Broadband neural signals (0.1–30 kHz) were recorded with a Cerebus neural signal processor system (Blackrock Microsystems) and bandpass filtered offline between 0.6 and 3 kHz using a 2nd order Butterworth filter. Spikes were detected with an amplitude threshold set at five times the median absolute deviation and spike events larger than 50 times the mean absolute deviation were discarded. Further, spike events with an inter-spike interval of less than the refractory period of 0.5 ms were also discarded. Spike times were aligned to the onset of the photodiode signal indicating the actual time of presentation of the last item in a sequence. All further analyses were performed with Python. Firing rates of individual sites were computed from the spike times in non-overlapping bins of 25 ms and smoothed with a Gaussian kernel corresponding to 50 ms standard deviation. For the data shown in Figure 6 and Figure S5, firing rates were computed with a moving average window of 50 ms and a step size of 10 ms in order to obtain a better temporal resolution.

Spike sorting

To isolate spikes from individual neurons we performed dimensionality reduction of the waveforms using Uniform Manifold Approximation and Projection (UMAP), in Python. We then employed the HDBSCAN clustering algorithm to cluster the reduced data and obtain cluster labels for each spike. To ensure that the waveforms in a cluster all belong to a unique neuron, we only retained clusters for which the timing of the spikes respected a refractory period of 2 ms while firing above 5 Hz. To this extent, we used the function "rpv_contamination" of the UltraMegasort2000 toolbox.[94] The function returns an estimate of the proportion of contamination by rogue spikes from other neurons. We selected only neurons for which the proportion of contaminating spikes was significantly lower than 5%. Finally, if any two clusters had more than 30% of their spike times overlapping within a 2 ms window, we selected the largest cluster to avoid selecting multiple instances of a waveform that crossed the threshold at multiple points. This spike sorting procedure was performed on each channel independently.

QUANTIFICATION AND STATISTICAL ANALYSIS

All of the below analyses were performed in Python.

Mixed selectivity analysis

To assess selectivity of single neurons in PFC, we performed an ANOVA on the activity of individual neurons in all six recording sessions of monkey A. The independent variables were stimulus identity (A or B), local deviance and global deviance as well as pairwise interactions. The dependent variable was the average firing rate during 1 s after onset of the last stimulus. As the sample size was unbalanced for global standard and deviant trials, we used subsampling of global standards. We visually inspected and confirmed normal distribution. ANOVA was performed for a total of 100 random subsamples. The resulting p values for the coefficients of the independent variables were corrected for multiple comparisons across all recording sessions, separately for each sample, using false discovery rate (FDR)[21];. Neurons with a significant sample-averaged coefficient ($p < 0.01$) for only one variable were considered to have classical selectivity, those with significant coefficients for at least two variables but not for interactions as linear mixed selective and those with significant interaction effect as non-linear mixed selective. For statistics, see Table S1 and the text of the Results section.

Effect of context in individual channels

We separately assessed the effect of context on the responses of individual sites in order to determine whether a reduction of firing alone could explain the difference between xx and xY blocks. We only included channels in the analysis that were significantly modulated by the sequence presentation. As a criterion for sequence modulation, we tested if there was a difference in firing rate of individual channels during the 300 ms fixation period prior to sequence onset and the first 300 ms after presentation of the last sequence stimulus. We used pairwise two-tailed *t* test per recording site and recording session and FDR across all sites and sessions within an animal to correct for multiple comparisons. Sites with a corrected p value <0.01 were regarded as being modulated by sequence presentation. We then used, for each of these channels, independent two-tailed *t* test on the response averaged over the duration of the first three stimuli of a trial. P-values were also corrected with FDR and sites with a p value <0.01 were considered as being modulated by sequence context (xx vs. xY block). Given the sign of the t-value, we could infer whether a channel showed reduced or increased firing for xY blocks compared to xx blocks. Values for trial N and mean values are given in the respective result section.

Population analyses

Multiple linear regression

We assessed how the sequences were represented on the neural population level by computing the axes across the MUA space that carried most information about the variables *stimulus identity*, *global context*, *local* and *global deviance*, without pre-selection of recording sites. For this, we used multiple linear regression.[14] We performed two separate analyses, one to study the representation of the sequences prior to the last stimulus, and one for the time after onset of the last stimulus, in order to measure responses to deviants. For the time before the last stimulus onset, the variables stimulus identity and global context were considered, whereas the trial condition after onset of the last stimulus was defined by the variables stimulus identity, local and global deviance. As deviance responses following the last stimulus might be dynamic, we performed a separate regression per time bin between 0 until 1.4 s after onset of the last stimulus.

The multiple linear regression was performed separately for each recording channel with the above-mentioned sequence variables as independent variables and the MUA (r) of channel i (in a time bin t) as dependent variable:

$$r_{t,I} = \beta_{t,i}^{stim} \times stimulus + \beta_{t,i}^{loc} \times local + \beta_{t,i}^{glob} \times global + \boldsymbol{\varepsilon}_{t,i} \quad \text{(Equation 1)}$$

The above equation holds for time bins t after presentation of the last stimulus. For the analysis of sequence structure representation before the last stimulus, the responses between 0 and 1.8 s after sequence onset were averaged per trial and a single regression was performed per recording channel.

$$r_i = \beta_i^{stim} \times stimulus + \beta_i^{ctx} \times context + \boldsymbol{\varepsilon}_i \quad \text{(Equation 2)}$$

ε is a noise parameter per channel (and time bin). r is a vector of dimension *Ntrials*, as are the independent variables stimulus, local, global and global context. Those were dummy variables, with A = −1, B = 1 for the stimulus variable; local or global standards = −1, local or global deviants = 1; xx block = −1 and xY block = 1. This approach results in a coefficient β per channel, variable, (and time bin) that indicates how much the firing of a channel was influenced by a certain variable.

The set of the 96 coefficients (number of recording channels) for one sequence variable k (and time point t) constitutes a 96−dimensional vector (k) (or (k)t) that we denote as the population axis representing this sequence variable. Note that Mante et al.[14] orthogonalized these axes and denoised them using principal component analysis. We chose not to add these steps after regression in order to measure orthogonality resulting directly from the regression and because the data did not require further denoising. The detailed N values for each recording session and condition are given in Table S2.

Decoding from the population trajectories

In order to use the resulting population axes for decoding, the MUA of all channels was projected onto the population axis of each sequence variable k, respectively.

$$r_{j,t}^{(k)} = R_{j,t} \cdot \beta^{(k)} \quad \text{(Equation 3)}$$

or

$$r_{j,t}^{(k)} = R_{j,t} \cdot \beta_t^{(k)} \quad \text{(Equation 4)}$$

for time-varying population axes.

j is the trial index. Rj,t is a 96−dimensional vector of the population firing rate in a single trial j and time bin t. rj,t is a scalar and corresponds to the dimensionality-reduced population activity in one trial j and time bin t in the subspace carrying most information about a sequence variable k. This trial-by-trial projection was then used to classify trials according to each sequence variable. The sign of these projections was dependent on the definition of the independent variables (see above). A positive activation along the axis coding of stimulus identity, e.g., corresponded to stimulus B, whereas a negative activation corresponded to stimulus A. As a measure of decoding performance, we computed the area under the ROC curve (AUROC) by varying the decision boundary for classification.

Cross-validation

The decoding performance was cross-validated both within and across sessions. Within each session, we used 10-fold cross-validation, meaning that 90% of all data was included for the regression and the remaining 10% for projecting test data onto the obtained axes. This was repeated 10 times so that all trials were used for testing. We shuffled the data prior splitting and ensured a balance of trial conditions in the training data. The reported performance within a session is the AUROC across all tested trials.

For the cross-validation across sessions for the variables *global context*, *local and global deviance*, we used the population axes from one random training fold of each session and projected all trials from all other sessions onto these axes. We then computed the AUROCs for each pair of training and test sessions and reported the performance separately for pairs that had the same or different stimulus pairs.

Decoding of serial position

We used multinomial logistic regression to predict the item position in a sequence based on the neural population responses. This was a classification with 4 target classes (item 1–4). The 300 ms after onset of a stimulus, shifted by 100 ms, were labeled with the item number of the most recent stimulus. The activity of each channel was averaged in these intervals, resulting in 4 values per trial.

The 4 items from all trials were pooled and used to train the classifiers, using 10-fold cross-validation. Only xx trials were used for training. For testing, the activity in each test trial and time bin between 100 ms prior to sequence onset until 1.4 s after sequence offset was passed through the trained classifier. This resulted in predictive probabilities for item 1 to 4 over time and allowed us to study the dynamic encoding of item position throughout a trial. We also assessed item position classification on incomplete trials. The monkeys could break fixation at any time during a trial by moving the gaze outside of the 6° fixation window which aborted the presentation of the sequence. We computed the predictions for item position 1 to 4 for trials interrupted after presentation of the first, second or third stimulus. Note that the fixation break could have occurred at any time between onset of one stimulus and onset of the next stimulus, meaning that the time the monkeys perceived the last stimulus varied within one condition.

Cross-condition decoding of global deviance

To test for encoding of global deviance irrespective of local deviance, we trained a separate binary classifier to predict global deviance for the time after last stimulus onset, on xx trials only and tested on xY trials. We used logistic regression on the pooled data from all sessions, per animal. This was done to reduce the impact of the block structure of the paradigm within each session, which could have been problematic in this case, as the classifier was trained from global deviants and standards from separate blocks (e.g., rare xx in an xY block vs. frequent xx in an xx block). Decoding performance was again measured as AUROC for each time bin, separately for each session.

Decoding of deviance or change in control data

We used logistic regression to predict whether a sequence chunk was XXXY or XXXX (deviance decoder), based on the activity after the last stimulus. We used a time-varying decoder in time bins of 50 ms, and with a step size of 10 ms. Image identities were balances in both conditions, i.e., we only included stimuli for training that occured in XXXY and XXXX chunks, resulting in 757 unique images. Decoder performance was cross-validated by leaving trials with one image out for testing. We hence trained 757 different classifiers. Images that did not occur in both conditions (191 different images) were only used once for testing but not included in the training set.

We additionally trained a decoder to detect any change from a repetition by contrasting the response to alternations and repeats in the second and third position of sequence chunks (i.e., the second stimulus in WXYZ and XYYZ trials vs. the second stimulus in XXXY trials as well as the third stimulus in WXYZ vs. the third stimulus in XYYZ trials). The same cross-validation approach was used to test this decoder.

Participation ratio

To evaluate the fraction of sites that contributed to the encoding of each subspace, we introduced a measure of participation ratio. It represents the percentage of sites that have significant coefficients for each subspace, respectively. We performed 100 random permutations of trials and recomputed the coefficients using the above described population analysis. As a time point, we chose the time bin of maximum decoding performance for each sequence variable. The fraction of absolute permutation coefficients that were larger than the true absolute coefficients resulted in a p value for each subspace. P-values were corrected across all subspaces using FDR. A p value <0.05 was regarded as significant. The participation ratio was defined as the fraction of sites with significant p value for each subspace.

Assessment of learning effects during the habituation period

We measured whether the code for global context evolved over the course of the habituation period. For this, we projected the activity of single habituation trials (0–1.6 s after sequence onset) onto the population vector that separated xx from xY blocks during the test trials within the same session. As we had used 10-fold cross-validation to obtain those vectors, we also obtained 10 projections of the same habituation trials. We averaged those projections across folds to obtain one activation value for the global context trajectory per trial. To assess learning, we tested the evolution of the difference between xY and xx blocks over the 50 habituation trials using a linear regression of this difference onto trial number. For this, we first averaged the projections from the two blocks of the same type within a session, then computed the difference between block types and averaged the difference over of all sessions for each animal. Statistics are given in the main text.

Statistics

Statistical tests for single-neuron and single-channel analyses are described in the respective paragraph (see above).

Random permutation test

To test the significance of decoding performance from population trajectories, we used a random permutation test with cluster-based correction for multiple comparisons.[95] After estimating the population axes and projecting single trials onto these axes, we generated 100 surrogate datasets by shuffling the trial conditions of test trials. We then computed the AUROCs for the different sequence variables based on the trajectories with the shuffled trial labels. We averaged the true AUROCs across recording sessions (10 in monkey H, 7 in monkey A) and likewise obtained 100 surrogate session-averages. The true AUROCs per sequence variable were transformed into t-values by subtracting the average over the permutations and dividing by their standard deviation, separately for each time point. Absolute t-scores that passed a threshold of 3 standard deviations were candidates for significant clusters in time. This procedure corresponds to a two-tailed test. A correction for multiple comparisons across time was performed by comparing the sum of t-values within each true cluster with the sum of t-values within surrogate clusters. Those surrogate clusters

were obtained by transforming each of the 100 permutation samples into t-values by subtracting the mean of the remaining 99 samples and dividing by their standard deviation. If a true cluster had a sum of absolute t-values larger than 95% of the largest surrogate clusters, it passed the threshold for significance which was set to a type-1 error of 5%. For the test of decoding performance across sessions, the same procedure was followed. First, we averaged for each test session the performance based on the decoder trained on the different possible training sessions (same or different stimulus pair). Second, we averaged the 10 or 7 test sessions. The same was done for the surrogate AUROCs based on shuffled trial labels. The same test was also performed for the cross-condition decoding of global deviance and the effect of deviance onto eye movements (see below).

Analysis of eye movements

The eye velocity (v) was measured from the non-calibrated horizontal (x) and vertical (y) eye position recording by computing the difference between time bins (t).

$$vt = |xt - xt - 1| + |yt - yt - 1| \quad \text{(Equation 5)}$$

To test for an effect of local or global novelty on eye movements, the median smoothed velocity (20 ms moving average) in each condition was computed across trials from all sessions and tested using a random permutation test (see above). We then controlled for eye movements in the time window during which there was a significant effect of deviants, +- 100 ms to be more inclusive, by removing deviant trials with an average velocity in this time period above the median eye velocity in the standard trials. The effect of deviance in the neural data was visualized for all trials vs. the controlled case (Figures S5A and S5B).