Neuron

Temporal continuity shapes visual responses of macaque face patch neurons

Highlights

- Face patch neuronal responses depend on temporal continuity
- Neural selectivity is disrupted by randomly presenting snippets and static frames
- Onset responses to dynamic movie snippets are uncorrelated with original movie
- Post-onset responses are moderately disrupted by temporal context

Authors

Brian E. Russ, Kenji W. Koyano, Julian Day-Cooney, Neda Perwez, David A. Leopold

Correspondence

brian.russ@nki.rfmh.org

In brief

Although we experience the world as a continuous stream of information, our knowledge about the visual brain stems primarily from its responses to brief and isolated stimuli. Here, we show in the macaque object pathway that selectivity to natural movie content is strongly determined by visual continuity and temporal context.

Russ et al., 2023, Neuron *111*, 903–914
March 15, 2023 Published by Elsevier Inc.
https://doi.org/10.1016/j.neuron.2022.12.021

CellPress

Neuron

CellPress

Article

Temporal continuity shapes visual responses of macaque face patch neurons

Brian E. Russ,[1,2,3,4,6,*] Kenji W. Koyano,[1] Julian Day-Cooney,[1] Neda Perwez,[1] and David A. Leopold[1,5]

[1]Section on Cognitive Neurophysiology and Imaging, National Institute of Mental Health, Bethesda, MD 20814, USA
[2]Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute, Orangeburg, NY 10962, USA
[3]Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[4]Department of Psychiatry, New York University at Langone, New York City, NY 10016, USA
[5]Neurophysiology Imaging Facility, National Institute of Mental Health, National Institute of Neurological Disorders and Stroke, National Eye Institute, Bethesda, MD 20814, USA
[6]Lead contact
*Correspondence: brian.russ@nki.rfmh.org
https://doi.org/10.1016/j.neuron.2022.12.021

SUMMARY

Macaque inferior temporal cortex neurons respond selectively to complex visual images, with recent work showing that they are also entrained reliably by the evolving content of natural movies. To what extent does temporal continuity itself shape the responses of high-level visual neurons? We addressed this question by measuring how cells in face-selective regions of the macaque visual cortex were affected by the manipulation of a movie's temporal structure. Sampling a 5-min movie at 1 s intervals, we measured neural responses to randomized, brief stimuli of different lengths, ranging from 800 ms dynamic movie snippets to 100 ms static frames. We found that the disruption of temporal continuity strongly altered neural response profiles, particularly in the early response period after stimulus onset. The results suggest that models of visual system function based on discrete and randomized visual presentations may not translate well to the brain's natural modes of operation.

INTRODUCTION

We experience the world as a continually evolving sequence of events, scenes, and interactions. Humans and other animals use their vision to track and integrate a wide range of information, coordinating the inherent dynamics of the natural world with their own active sampling of important scene elements.[1,2] This integration provides a predictive context for interpreting the flow of incoming visual signals. What role might high-level visual neurons, such as those observed throughout the macaque inferior temporal cortex, play in such integration? For example, would the visual response profile of face-selective cortical neurons depend on the temporal continuity of a video? What might be the neural consequences of breaking up normal visual experience into discrete, randomized presentations?

It is difficult to make predictions on this point, since nearly all that is known about visual responses in the brain derives from the brief presentation of stimuli, often in highly reduced forms. For example, most studies of the macaque inferior temporal cortex have used isolated images of natural and artificial stimuli flashed briefly on the screen to study object selectivity among visual neurons.[3–7] By comparison, only a few single-unit studies have investigated how such neurons respond under more naturalistic conditions, for example, when a subject is allowed to freely gaze within an evolving visual scene.[8–10] Thus, much more is known about how the brain responds to flashed presentations of isolated stimuli than how it responds under the continuity and complexity of everyday vision.

In some areas, researchers have studied the brain's encoding of visual temporal sequences, such as in the processing of bodily actions.[4,11–14] These studies have linked the comprehension of bodily actions to the integration of form, movement, and location[15] and in some cases to the generation of one's own bodily actions.[11,16–18] Reductionistic investigations of sequence processing in the inferotemporal cortex has in some cases identified a marked temporal dependency, which has often been linked to Hebbian or other forms of statistical learning through repeated sequence presentation.[19–24] These studies provide evidence that object-selective neurons in the inferior temporal cortex exhibit significant temporal dependency in their responses.

A somewhat different facet of temporal integration involves the active sampling of a visual scene provided by the frequent redirection of gaze.[1,25,26] For humans and other primates, whose high-resolution vision is sharply concentrated in the fovea, the sequential gathering of important information across fixations is a central feature of visual cognition.[27] In normal vision, the temporal discontinuities introduced by this active sensing of the visual scene through gaze redirection is superimposed upon the temporal dynamics of actors and events in the external world. The stable perception of the world amid the constantly

shifting retinal stimulation is one of the enduring puzzles of visual neuroscience.[28–30]

Recently, the use of naturalistic stimulus paradigms has gained popularity and has begun to complement more conventional presentation methods (for a recent review, see Leopold and Park[31]). The human fMRI community has, in particular, taken up the use of movies as a new paradigm for tapping into various aspects of brain function,[32–48] with some monkey investigators following suit.[49–58] One single-unit study investigated responses in macaque anterior fundus face patch to repeated 5-min presentations of naturalistic movies full of dynamic social interaction.[8] Neurons in this area were strongly entrained by the movie timeline, with a high level of response repeatability across presentations of the same movie despite the free movement of the eyes. However, the role of temporal continuity in shaping neural responses to complex movie content is at present poorly understood. Although not yet studied in detail, this question may be important, since our understanding of high-level vision is at present shaped strongly by the brain's response to the presentation of discrete and isolated stimuli.

The current study addresses the role of temporal continuity on neural responses in the ventral visual pathway using a paradigm in which we systematically manipulated the temporal continuity of visual content stemming from a naturalistic movie. We recorded from cells in the macaque anterior fundus (AF) and the anterior medial (AM) face patches, in which responses to both static images and dynamic movies have previously been studied.[3,8,54,59–68] The goal of the present study was not specifically related to face processing but instead used the well-characterized responses in these areas to understand the influence of visual continuity (the uninterrupted viewing of stimulus) and temporal context (the effect of preceding meaningful visual content) over high-level visual responses. Our approach was to sample visual snippets from throughout a movie, thus breaking its natural continuity, and to observe whether neurons would maintain their selectivity when the movie content was delivered in discrete and randomized form. We found that this disruption of temporal continuity strongly and significantly affected neural responses to the movie content in both areas we recorded, as well as over multiple timescales. The most profound impact was on the response to the initial onset of the discrete presentation, for which neural responses were nearly uncorrelated to those observed in response to the same content of the intact movie.

RESULTS

Six macaque monkeys (*Macaca mulatta*) viewed intact movies and randomized movie clips ("snippets") as visual responses were recorded from single neurons in the AF and AM face patches. These regions were defined functionally using fMRI[63] and targeted using methods described previously[59,60] (Figure 1A, see method details). In total, we recorded from 163 neurons in the AF face patch (Subject T: 18; Subject R: 16; Subject S: 91; Subject SR: 38) and 72 neurons in the AM face patch (Subject D: 49; Subject M: 23). For the main analyses, we combined neurons across regions into a single population after determining that the results were similar in both regions. Corresponding analyses performed separately in the two face patches are presented in supplemental information. The animals viewed multiple repetitions of an intact 5-min movie. They also viewed randomized short snippets and static images extracted from the movie at each of 300 time points spaced by one second (see method details). Most of the analyses described below involve comparing the responses with 300 equivalent visual segments from the original, intact movie with either 800 ms movie snippets or 100 ms static images. Corresponding responses to two additional conditions (250 and 100 ms movie snippets) are also discussed. Five of the monkeys (T, R, S, D, and M) were free to direct their gaze while viewing the continuous movie and randomized snippets. In addition, two of the monkeys (S and SR) viewed the same movie and snippets under controlled fixation.

The recordings hinged on the longitudinal recording capabilities of the implanted microwire bundles since the data for individual neurons was amassed over 8–18 daily sessions of data collection[8,59](see Figure S1). Each of these sessions began with several minutes of conventional testing with flashed images from different stimulus categories, which was used both to establish each neuron's basic selectivity and as a tool for "fingerprinting" to identify the same neurons across subsequent recording sessions. This testing was followed by all other conditions presented in randomized blocks (Figure 1C). We first investigated the effects of temporal continuity on movie content during free viewing. In this condition, the intact movie was viewed between 8 and 24 times (median for viewings per neuron: monkey T = 8, monkey R = 10, monkey S = 15, monkey M = 24, and monkey D = 18). The randomized movie snippets were viewed between 7 and 37 times each (median viewings per neuron: 800 ms: T = 7, R = 10, S = 13, M = 13, and D = 12; 250 ms: T = 7, R = 11, S = 16, M = 14, and D = 17; 100 ms: T = 7, R = 9, S = 17, M = 13, and D = 20; static image: T = 13, R = 17, S = 37, M = 33, and D = 28).

Effects of manipulating temporal continuity during free viewing

Previous work has shown that neurons in the areas we recorded are reliably entrained to the movie under conditions of free gaze, when the animal actively samples the content of the movie. We first asked whether, under these conditions, the neural responses depend on the temporal continuity of the movie. We found that breaking up the movie into snippets or frames and randomizing their presentation strongly influenced the stimulus response selectivity of neurons in these areas. The differences were most obvious in the early responses, from 100–300 ms following stimulus onset, but were also evident in the later response period to video snippets.

Figure 2 shows an example neuron (neuron AF094) responding to the same video content in continuous versus randomized snippet presentation. This neuron was typical in that it was face selective and entrained reliably by the movie. Figure 2A shows the raw raster responses during several representative 800 ms snippets and corresponding static frame presentations. From these example rasters, the repeatability across presentations and the effects of randomization are both evident. Figure 2B then shows the reconstructed time courses to the 800 ms snippets early onset responses (red, 100–300 ms following snippet onset) and the latter responses (blue, 300–900 ms following

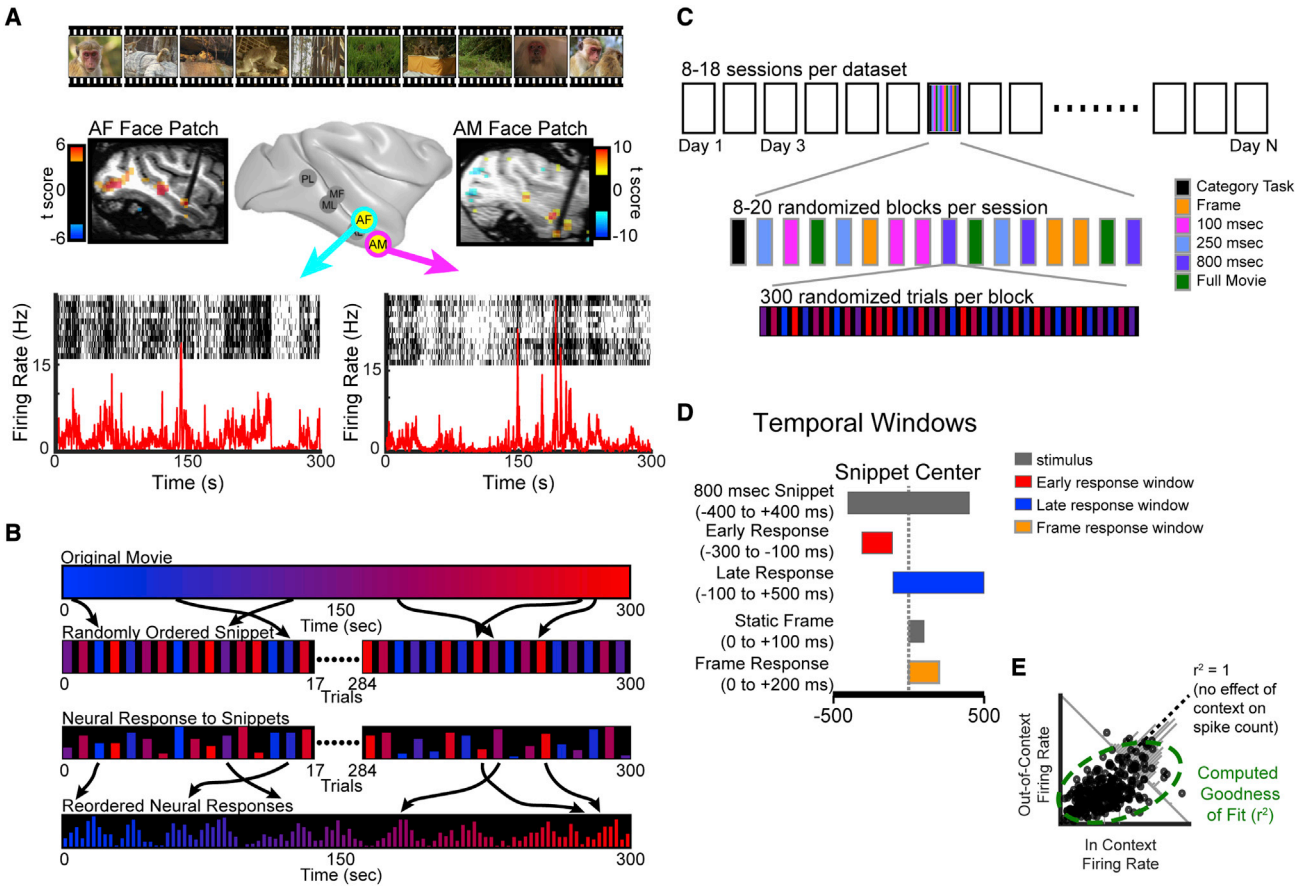

Figure 1. Experimental design

(A) Top row shows representative frames from the intact movie that was presented during the free-viewing paradigm. Middle row depicts the results of the fMRI face-patch localizer task used to target the AF (left) and AM (right) face patches. Functional localizer is overlaid on a T1w image taken after the electrodes where implanted (see method details). The bottom row depicts the spiking response from one AF (left) and one AM (right) neuron to the viewing of multiple trials of the intact movie. The black dots represent the time of a spike with 1 ms bins, and each row is a trial. The red line is the mean spike density function for all trials.

(B) Schematic of the creation of the individual snippets from the original movie, and the reconstructed response to those snippets.

(C) Schematic of the pseudorandomized task design that was used to collect the data within and across sessions.

(D) Schematic for the selection of response periods based on the individual snippet lengths.

(E) Schematic of the regression analysis between the snippet responses and the original movie. The green ellipse represents a hypothetical goodness-of-fit measure.

snippet onset) in comparison to the equivalent time periods from the continuous movie. The responses to the randomized presentation are reordered to match the native timeline of the movie (see Figure 1B, method details). Both reconstructed time courses revealed notable deviations in spiking during the corresponding analysis windows of the continuous presentation (gray time courses).

Figure 2C applies a similar analysis to spiking responses to the static single frames drawn from the movie, which showed little correspondence to the spiking activity in the continuous condition. Scatterplots showing the response correspondence for the 300 sample time points are shown to the right of each reconstruction. For this example neuron, the initial onset responses to both the snippet (red) and static frame (yellow) were uncorrelated with the responses to the equivalent visual content presented in the continuous condition (snippet: $r = 0.0665$, $p = 0.2506$; static frame: $r = -0.031$, $p = 0.5933$). The responses to content later in each snippet (blue) were affected by randomization but retained their correlation with the continuous responses ($r = 0.55$, $p = 0.01e{-}23$). Additional single-unit examples from both areas are shown in Figure S2.

Across the population, neural selectivity was strongly affected by the fragmented and randomized presentation of the movie content. Comparing the neural responses to the same movie frames presented in its original format, the most prominent changes in selectivity were observed in the response directly following the onset of the snippet or static frame. In fact, we found that for many visually responsive neurons, the selectivity measured following the onset of a snippet was statistically uncorrelated with that measured to the same frames during the native movie presentation. We quantified this relationship with a linear regression applied to the 300 samples compared across

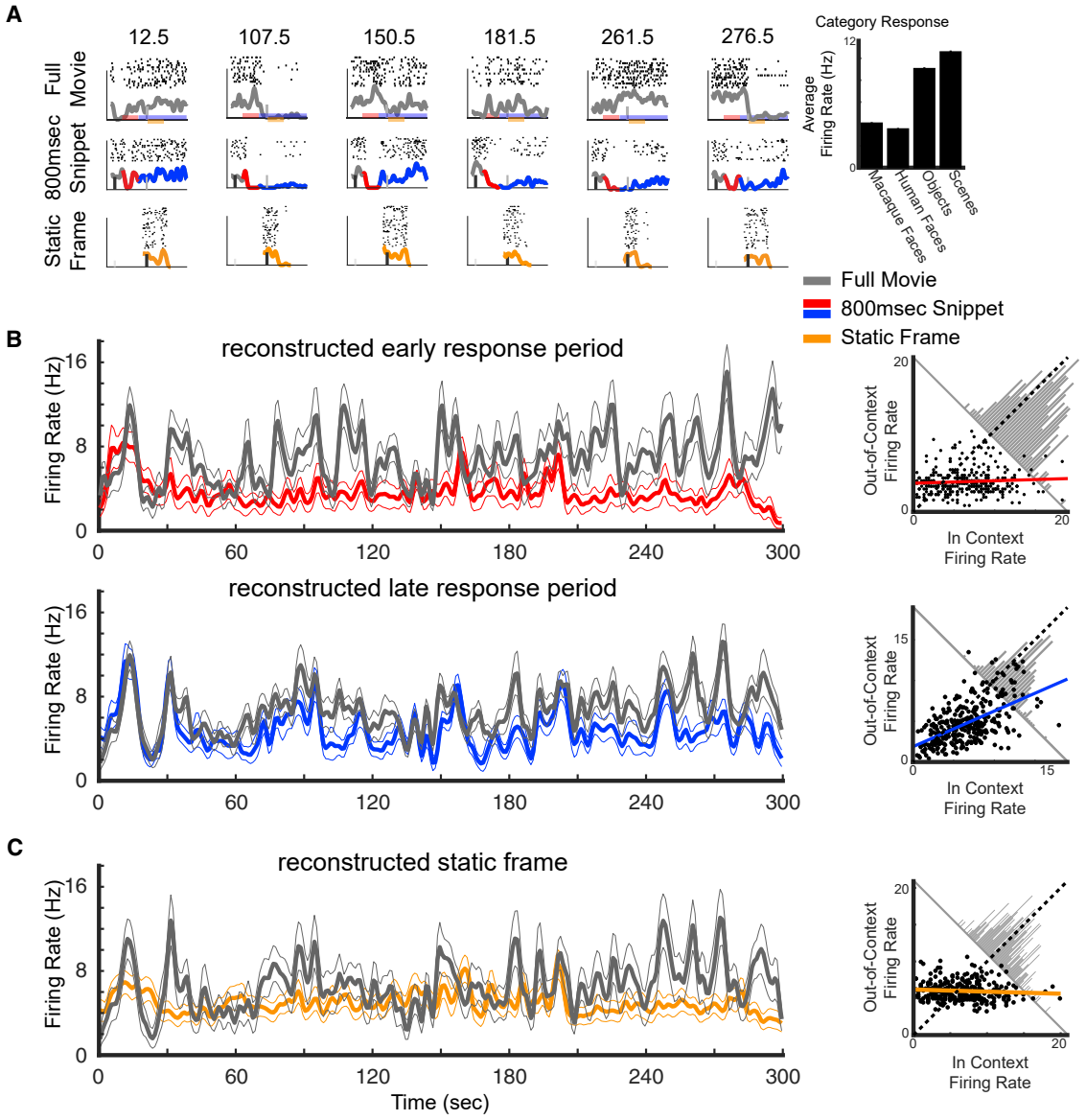

Figure 2. Representative response of a neuron from area AF

Six representative time points are featured (A), which show the diversity of response across time points and contexts. We compared stimulus-driven response during the intact movie (top row), 800 ms snippets (middle row), and 100 ms static center- frame (bottom row). The rows of spiking data show the raster plots and computed average spike density function during the three conditions for the featured time points. The mean firing rate (Hz) to the static categorical images is presented next to the snippet responses, highlighting the current neuron's face responsiveness. This neuron's average firing rate during the category task is shown in the right most inset. Responses to the category task were calculated over a 200 ms window starting at stimulus onset and averaged across all presentation within a given category.

(B) Average intact movie response (gray) shown together with the reconstructed spike count responses to the early response (red, top), the late response (blue, bottom) using matching analysis windows relative to the center frame (see Figures 1B and 1D). Lighter lines represent the standard error of the mean.

(C) Analysis of response to static frames (orange), compared with reconstructed response in same format as (B).

the corresponding conditions (see scatterplots in Figure 2). This analysis provided a measure of goodness of fit (r^2), essentially summarizing the similarity in responses to the same movie content in continuous and discontinuous presentation modes.

Figures 3B and 3E show that for most neurons the visual selectivity to the initial presentation of the 800 ms snippets and static frames, respectively, had a poor correspondence to that observed during the continuous condition (Figure 3B, $r^2_{mean} = 0.11 +/- 0.12$; Figure 3E, $r^2_{mean} = 0.06 +/- 0.08$). In both cases, the r^2 distribution dropped sharply from a control analysis, in which a split-halves analysis was used to compute the r^2 distribution from the movies using the same number of trials (paired t test, Figure 3B, paired samples $t(103) = 25.99$, $p = 0.5e^{-48}$; Figure 3E, paired samples $t(103) = 22.18$, $p = 0.5e^{-42}$). These findings demonstrate that the

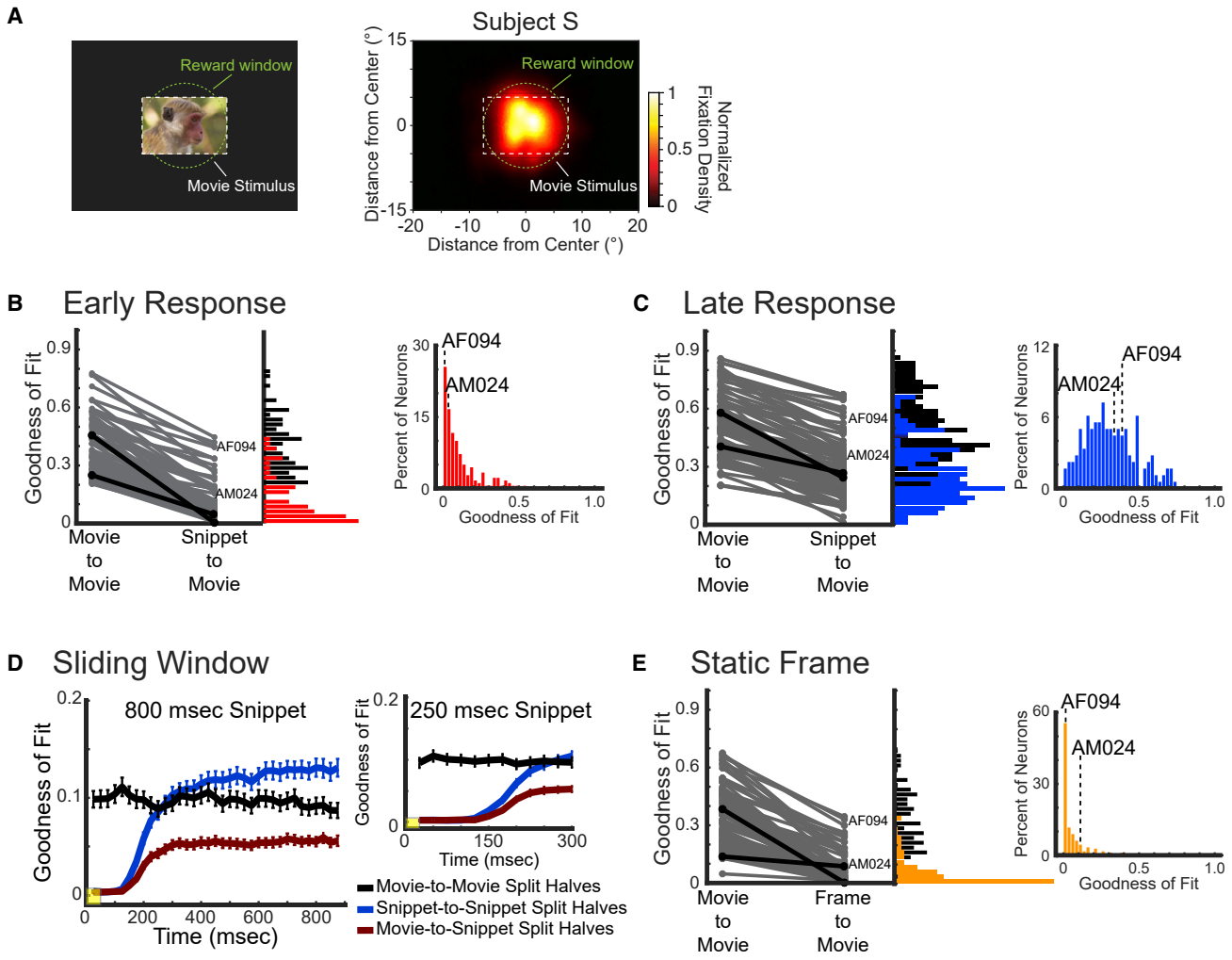

Figure 3. Population response changes across context and time

(A) Schematics of the free-viewing paradigm used during the movie-watching condition. Dotted areas represent the size of the stimulus (white) and reward windows (green). The reward window was not shown to the animals. The density plot shows the summed gaze positions during the free viewing of the within-context movie.

(B) Paired comparison within-context and between-context goodness-of-fits (r^2) of neurons in the AF and AM face patches during the onset period. Neurons in the pair comparison were filtered by an r^2 or 0.2 in the within-context condition. The black lines highlight the example neurons presented in Figures 2 and S2B. The vertical histograms represent the density of r^2-values for the split halves within context (black) and between context (red) for the full population. The horizontal histogram, on the right, shows the r^2 for all trials of the between context regression for the full population.

(C) Depicts the same data and analyses for the late period of the 800 ms snippets.

(D) Shows the evolution of the goodness of fit for the within-context (black), out-of-context (blue) and between context (red) regressions using a sliding window of 50 ms (yellow box) with 25 ms jumps, for both the 800 and 250 ms snippets. Error bars represent the standard-error of the mean of the population within each bin.

(E) Shows the same set of comparisons as (B) and (C) for the response to the presentation of the static frames.

abrupt appearance of a static image or dynamic clip on a blank background elicits visual responses that differ markedly from the responses to the same content presented within the context of the movie (see Figure S3 for analyses for separate analysis of AF and AM). In addition to the 800 ms movie snippet featured above, analysis of the onset period for the independent 250 and 100 ms snippet presentations yielded similar results (Figure S3).

Later in the response to the snippet, this disruption of visual selectivity was less pronounced (Figure 3C). Comparing visual responses during this period to the native movie allowed us to assess the effects of temporal continuity itself. By analyzing responses beginning 300 ms following the onset of the snippet, the transient effects were largely diminished (see Figure 3D), thus we could focus on potential effects of temporal randomization of responses to the movie content. Individual neurons varied in the level of disruption of their late period responses, with some neurons moderately affected by temporal reordering (e.g., Figures 2 and S2A) and others showing little if any influence (e.g., Figures S2B and S2C). Across the combined AF/AM population, the influence of temporal continuity was evident as a broad

distribution of r^2 measures (Figure 3C, r^2_{mean} =0.31 +/− 0.17), which differed markedly from the narrow r^2 distributions of the early responses (see Figure S3 for separate analysis of AF and AM). As above, the r^2 values were contrasted with those obtained from a within-session split-halves control condition. Importantly, responses to movie content late in the snippet period were also markedly disrupted by randomized presentation (paired samples t(148) = 30.78, p = $0.3e^{-66}$), although less so than during the early period (repeated measures ANOVA [context × time frame]: main effect of context F[1, 179] = 510.46; p = $0.3e^{-53}$; main effect of time frame F[1,179] = 862.39 p = $0.2e^{-70}$; but no interaction F[1, 179] = 0.56; p = 0.45). Thus, the visually selective responses in these face-selective areas are also affected by temporal continuity expressed over longer timescales.

Figure 3D plots the temporal evolution of selectivity for the snippets (800 and 250 ms) relative to the same content presented in the native movie (see method details). The running r^2 values show that the movie-to-snippet correlation is very low during the initial response period, climbing gradually and approaching its plateau after approximately 300 ms. The initially low r^2 values reflect the neural response to the abrupt onset of the visual snippet. Following this initial onset period, the movie-to-snippet r^2 remains lower than the other within-condition comparisons during the late period, reflecting the different temporal context resulting from snippet randomization. It should be noted that the absolute r^2 magnitude of the time course is lower than in the other panels because the correlation in the running analysis is computed over brief, 50 ms time windows. These results demonstrate two clear response phases following the onset of the movie snippets: namely, an early onset period in which response selectivity is severely disrupted, and a later period in which the onset effects have subsided, but the temporal reordering of the movie content continues to affect visual responses.

Effect of manipulating temporal continuity during controlled fixation

The experiments above demonstrate that when the monkey is allowed to freely direct its gaze the response to visual content depends on whether it is seen in the context of a continuous movie or isolated within an extracted snippet. Although this finding indicates that temporal continuity is critical for shaping high-level neural responses, one potentially confounding factor is the prospect that the animal's gaze position differs in the continuous versus snippet conditions and is possibly driving the context effects through different receptive field stimulation.[5–7,69–72] To test this possibility, we repeated the experiments above under conditions of controlled fixation. We recorded 55 neurons from two animals' AF face patches (17 new neurons from monkey S, and 38 neurons from task naive monkey SR), focusing on the repeated presentation of the full length 5-min movie and the randomized 800 ms snippets. In this condition, the animals were required to maintain their gaze continuously within 2 or 1.5 degrees of a central fixation point (Figures 4A and S4A), ensuring similar visual position of scene elements across trials and conditions.

The results from the controlled fixation condition were clear. Just as during the free-viewing condition, the disruption of temporal continuity strongly and significantly altered the selectivity to specific visual content (Figure 4). As above, we measured this change through the decreased goodness of fit between a direct comparison of neural responses during the within-context (i.e., movie) and out-of-context (i.e., snippet) conditions, restricted to periods in which the animals were fixating continuously. For both the early period (100–300 ms; t(54) = 11.8, p = 0.15^{e-15}; Figure 4B) and the later period (300–900 ms; t(54) = 13.501, p = 0.6^{e-18}, Figure 4C) of the snippet response, responses were markedly affected by the removal of the video content from the natural temporal context. When we calculated the sliding window as above (Figure 4D), we again found that the correspondence between the within and out-of-context responses began to increase shortly after 100 ms into the snippets and plateaued around 400 ms, never reaching the r^2 of the within-context responses. In contrast to the free-viewing analysis, here the snippet-to-snippet comparison showed the highest overall values, after an initial rising phase and then plateau at the same level as the movie-to-movie analysis. We interpret this increased correspondence as reflecting a brief, albeit potentially important, consequence of controlled fixation during initial snippet onset period. Finally, as above, the strongest changes in response selectivity were observed for the flashed static frames extracted from the movie. This finding was observed even when blank periods were eliminated between presentations of the static frames (Figure 4E). Thus, the nature and magnitude of the context effects closely resembled those measured during the original free-viewing condition for all the comparisons, suggesting that differences in gaze position did not underly the differences observed during continuous versus discrete and randomized stimulus presentation. This conclusion was further supported by analysis of gaze position during the free-viewing experiment described earlier (Figure S4).

Relationship to face selectivity

Although neural face selectivity measured with static, isolated images did not figure prominently into the design of the present study, we did confirm that many of the neurons we tested in and around the AF and AM face patches exhibited a preference for faces over other images (Figures 5A and S3 for separate analysis of AF and AM), in accordance with many previous studies.[3,62,73–75] Specifically, we found that 88 of the 180 recorded neurons, or 49% of our free-viewing population, had a *d'* greater than 0.65 which is a criterion similar to that used by previous studies.[73,75]

To test whether a cell's category selectivity had bearing on its observed dependency on temporal context during naturalistic stimulus presentation, we compared each neuron's d' score for face selectivity with the magnitude of its same context goodness-of-fit (Figure 5B) and its between context goodness-of-fit values (Figure 5C). The results show that although face-selective neurons did not maintain their selectivity during the presentation of movies (Figure 5B, black data), those having a higher level of face selectivity maintained somewhat higher selectivity under during the early phase of the snippets (Figure 5B, blue data). In addition, the results from the between context goodness-of-fits (Figure 5C) indicate a weak relationship, between the face selectivity d' value and context r^2 value in the early and late response periods. Thus, although the majority neurons were disrupted by temporal discontinuity, regardless of their

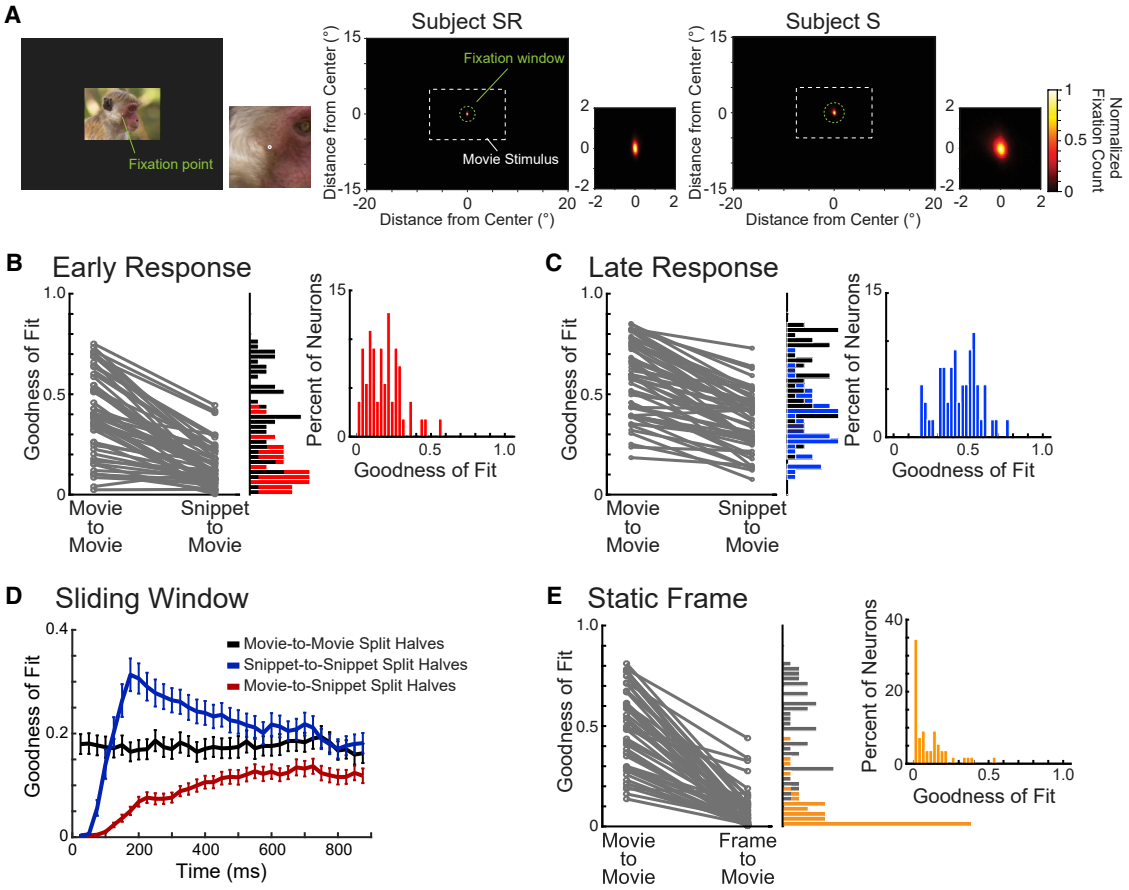

Figure 4. Population response during gaze control

Data from the AF face patch although subjects maintained their gaze on a central fixation point.

(A) Schematics of the fixation paradigm used during the gaze control condition. Dotted areas represent the size of the stimulus (white) and fixation windows (green). The fixation window was not shown to the animals. Small inset windows show the summed density of gaze positions throughout the within-context movie presentation.

(B and C) Show the paired comparison of within-context and between-context goodness-of-fits (r^2) of neurons in the during the early and late response periods, respectively, in the same format as Figure 3.

(D) Show the evolution of the goodness of fit for the within-context (black), out-of-context (blue), and between context (red) regressions using a sliding window of 50 ms (yellow box) with 25 ms jumps. Error bars represent the standard-error of the mean of the population within each bin.

(E) Paired comparison of within and between context goodness-of fit, in the same format as Figure 3B. The static frames were presented for 300 ms without ISI in this experiment.

category selectivity, the early responses of face-selective neurons showed a slightly higher preservation of selectivity during the snippets than other neurons.

DISCUSSION

In the current study, we found that the responses of neurons in anterior face-selective regions of the macaque inferior temporal cortex differ depending on temporal context. Neurons responded differently to the same movie content when it was presented continuously versus when it was divided into brief snippets and shown out of context in randomized presentation. The disruption in visual responses was particularly prominent in the period shortly following the onset of the snippet. The results were similar during free viewing and controlled fixation. In the following sections, we discuss how visual operations may differ between the continuous, free viewing of a movie and the brief, randomized presentation of its content. We also discuss the role of perceptual history in the establishment of spatial and temporal context during normal vision, as well as the bearing of these results on the activity of face-selective neurons.

Visual selectivity altered by abrupt onset

The most striking effect we observed following the presentation of snippets and static frames extracted from the movie was during the initial phase of the response. During the first 200 ms of activity following the onset of the stimulus, the neural response profile to the content extracted from the movie had virtually no correspondence to the same neurons' responses to the

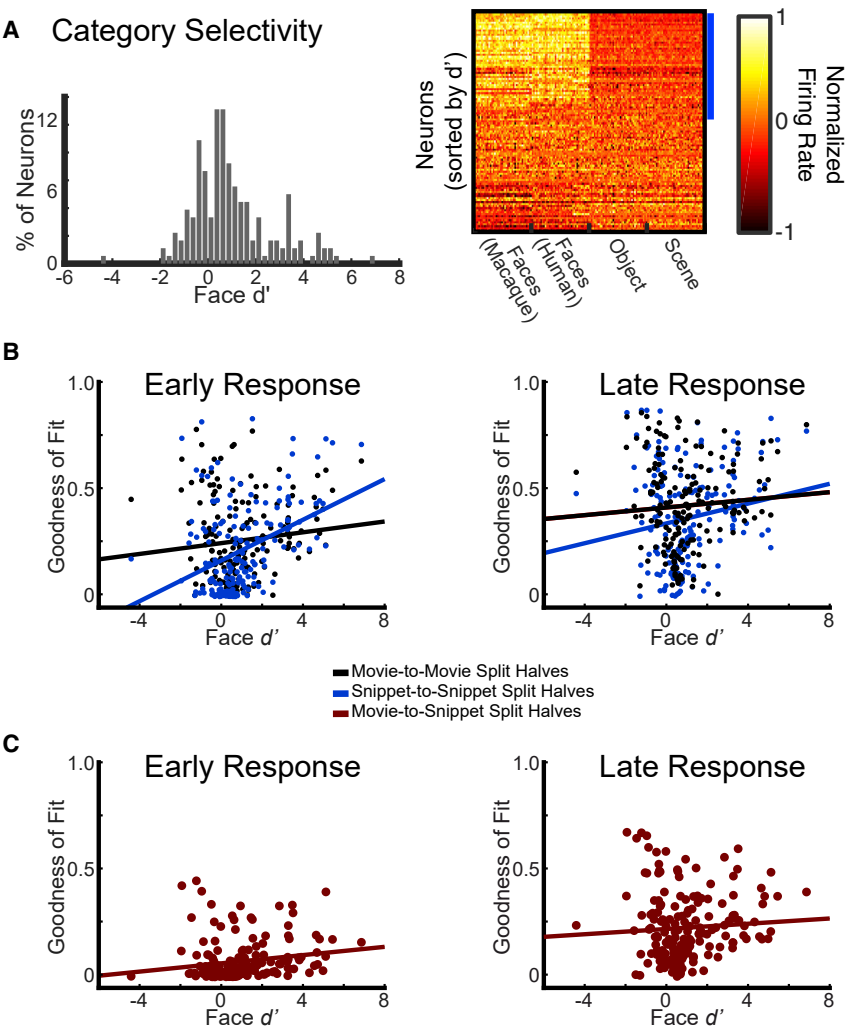

Figure 5. Interaction between face selectivity and context responses

(A) Selectivity of the combined face-patch neurons to faces over scenes and objects. The left histograms depicting the density of d' values for all neurons. Values of d' greater than zero represent a preference for larger responses to faces. The right heatmaps show normalized firing rate to the maximum response of a given cell, for all neurons within the populations sorted by their d' score. The blue bar on the side of the heatmap shows neurons that had significant d' values and thus would be categorized as face selective.

(B and C) Show the relationship between face selectivity and our goodness-of-fit (r^2) measure for the early and late response periods for the different contexts. (B) Depicts relationship, and lines of best fit, for the two movie-to-movie (black; early: Pearson's correlation = 0.1109, p = 0.1383; late: Pearson's correlation = 0.0698, p = 0.3517) and snippet-to-snippet (blue; early: Pearson's correlation = 0.3852, p = 0.9e−7; late: Pearson's correlation = 0.1708, p = 0.0219) goodness-of-fit measure and d′, for the early (left) and late (right) response periods). (C) Depicts the same comparison for the between context goodness-of-fits and a given neurons d' score (red; early: Pearson's correlation = 0.1706, p = 0.022; late: Pearson's correlation = 0.0738, p = 0.3248). All p values are Bonferroni corrected for multiple comparisons, 0.05 alpha = 0.0083.

same content appearing in the movie itself. Is this result to be expected? In some ways, the answer may be yes, since onset responses have long been shown to cause nonspecific transients throughout the visual system.[76–78] However, nonspecific transients were not the only factor reshaping neural selectivity in the early snippet responses observed in the present study. For some neurons, the response to snippet content, including the onset, was markedly decreased compared with the same content in the full movie (e.g., response at 150.5 s in Figure 2A). Likewise, the change in selectivity cannot be attributed to long response latencies. The average response latency for both populations (combined = 118 ms; AF = 115 ms; AM = 120 ms), as calculated relative to the static frame onset (see method details), preceded the plateau observed in the goodness-of-fit time courses (Figures 3D and 4D). The low correspondence of the responses during the snippet onset presentation instead suggests that the visual continuity itself is important.

Visual continuity is an aspect of natural experience that has been absent in most experimental assays of neural selectivity but which evidently has a significant influence on neural responses. One interpretation of the differences observed between continuous and discrete presentation conditions invokes neural adaptation or fatigue, which can affect neural processing at all levels of the visual system. Adaptation can have many manifestations, including alterations in stimulus feature tuning, and is usually studied through continued or repetitive presentation of stimuli.[79,80,81] In our study, the most straightforward way adaptation might affect responses is to diminish the magnitude of responses during continuous presentation compared with the discrete flashed presentation.[82,83] However, this was not the case. Instead, we found the magnitude changes were stronger overall during the visually continuous presentation (see Figures S5A and S5B). This finding, combined with the clear differences in stimulus selectivity in the two conditions, does not lead to a straightforward explanation based on neural fatigue. At the same time, it is likely that some form of adaptation contributes to the difference. To systematically investigate this topic, future studies will need to vary the temporal dynamics in the presentation of movie segments and to employ carefully designed stimuli whose specific adaptational influences on subsequent presentations can be assessed.

Although the neural populations recorded in this study were from two fMRI-determined face patches, our study did not aim specifically to investigate category selectivity nor its disruption with abrupt presentation. Our use of natural movie stimuli differs from the more conventional use of isolated objects in that the images depicted scenes containing multiple individuals and

objects within a complex spatial layout.[52] Because the movies were not designed to balance visual features or movie content equally, they are not well suited, for example, to systematically compare the effects of temporal continuity for faces versus other categories. Thus, more work is required to understand whether, for example, the observed effects of temporal context are related to the complex scene content or whether, for example, similar disruptions in tuning would be observed for continuous viewing of stimuli presented in isolated form or in the absence of spatial context.

Nonetheless, it bears emphasis that our understanding of high-level feature and category selectivity derives almost exclusively from the presentation of flashed, isolated stimuli. Thus, if the response selectivity observed during flashed presentation does not represent the visual preferences of neurons under more continuous conditions, current notions of feature coding and category representation in the brain may be inaccurate.

Temporal integration of category-selective neurons

Beyond the striking changes in the early responses, there were also smaller but significant changes that emerged from the temporal reordering of scene content. These changes were observed in both the free viewing and controlled fixation conditions. As most studies to date have intentionally randomized stimuli that have no particular temporal relationship, relatively little is known about whether feature-selective neurons in the temporal cortex are sensitive to temporal or behavioral context. Some work, however, has suggested that where an animal is looking[9,84,85] or what an animal has recently seen[13] can strongly affect the responses of IT neurons. Such neurons have been shown to modify their responses based on contextual information, usually in the form of a cue or memory component.[9,86–93] Neurons in the superior temporal sulcus are associated with motion stimuli,[12,14,16,94–96] in particular biological motion, for which temporal continuity may be particularly important. Recently, it has been shown that neurons in macaque IT can acquire selectivity for temporal context when isolated stimuli are played in a learned temporal sequence.[22,23,79] Increasingly, researchers are using naturalistic stimuli as a complementary approach to conventional stimulus presentation (for a review, see Leopold and Park[31]).

Our results provide some insight into timescales over which neurons integrate visual information by systematically varying the amount of temporal contextual information that was presented, in the form of brief movie snippets lasting 800, 250, and 100 ms snippets, and static frames presented for 100 ms, in each case extracted from a longer movie. We found that a subset of neurons within both the AF and AM face patches were significantly affected by the temporal reordering of movie content, even in the late snippet period, suggesting that they integrate some types of visual information at least over timescales of seconds.

In the ventral stream, few previous studies that have similarly addressed temporal integration, albeit with somewhat different methods. Hasson et al.[34] used fMRI in a naturalistic viewing paradigm somewhat similar to the one we implemented to investigate the length of temporal receptive fields throughout the brain. They found that temporal receptive windows, the length of coherent temporal context necessary to reliably stimulate a portion of the brain, increased systematically as information moved from early visual cortex through the visual system into the prefrontal cortex. In the human temporal cortex, they found 4–8 s spans of coherent temporal information were required to create reliable and repeatable visual responses. Although there are differences in methodology and experimental design, these findings match our results, likewise suggesting that neurons within the temporal lobe may incorporated greater than one second worth of contextual information in determining their response to a given stimulus.[8,31,48,76–78]

Conclusions

It is interesting to consider that during periods of visual free viewing, inferior temporal cortical neurons, including those face patch neurons reported here, enter into a mode of operation marked by sequential target selection, continuity of visual content and attentional processes, and anticipation of events. In this exploratory mode, and in contrast to discrete modes of image presentation, basic visual responses may be modified by neuromodulatory inputs, along with a range of brain areas involved in memory, spatial perception, motor planning, and executive function. As experimental work continues to explore the frontiers of natural visual experience, our portrait of visual brain organization may need to evolve and expand. A central feature of natural experience that is explicitly removed from most experimental approaches is that of temporal continuity. This study demonstrates that for populations of well-studied feature-selective neurons in the macaque high-level visual cortex, two aspects of temporal continuity, namely visual continuity, the uninterrupted viewing of a stimulus, and temporal context, the orderly progression of visual events in a scene, critically determines how the brain interprets its visual input at a given moment in time.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Surgical procedures
  - Visual stimuli and behavioral monitoring
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Longitudinal neural recording
  - Data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2022.12.021.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Institute of Mental Health (ZIAMH002838 and ZIAMH002898) to D.A.L.

Functional and anatomical MRI scanning was carried out in the Neurophysiology Imaging Facility Core (NIMH, NINDS, and NEI). This work utilized the computational resources of the NIH HPC Biowulf cluster http://hpc.nih.gov).

AUTHOR CONTRIBUTIONS

B.E.R., conceptualization, methodology, investigation, formal analysis, data curation, writing – original draft, writing – review and editing, visualization, and supervision; K.W.K., investigation, formal analysis, writing – review and editing, and visualization; J.D.-C., investigation and writing – review and editing; N.P., investigation and writing – review and editing; D.A.L., conceptualization, methodology, writing – original draft, writing – review and editing, supervision, and funding acquisition.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 22, 2022
Revised: September 9, 2022
Accepted: December 13, 2022
Published: January 10, 2023

REFERENCES

1. Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. Curr. Opin. Neurobiol. *20*, 172–176. https://doi.org/10.1016/j.conb.2010.02.010.
2. Melcher, D. (2001). Persistence of visual memory for scenes. Nature *412*, 401. https://doi.org/10.1038/35086646.
3. Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., and Livingstone, M.S. (2006). A cortical region consisting entirely of face-selective cells. Science *311*, 670–674. https://doi.org/10.1126/science.1119983.
4. Jellema, T., and Perrett, D.I. (2003). Perceptual history influences neural responses to face and body postures. J. Cogn. Neurosci. *15*, 961–971. https://doi.org/10.1162/089892903770007353.
5. Perrett, D.I., Hietanen, J.K., Oram, M.W., and Benson, P.J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. *335*, 23–30. https://doi.org/10.1098/rstb.1992.0003.
6. Desimone, R., Albright, T.D., Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci. *4*, 2051–2062. https://doi.org/10.1523/JNEUROSCI.04-08-02051.1984.
7. Gross, C.G., Rocha-Miranda, C.E., and Bender, D.B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. J. Neurophysiol. *35*, 96–111.
8. McMahon, D.B.T., Russ, B.E., Elnaiem, H.D., Kurnikova, A.I., and Leopold, D.A. (2015). Single-unit activity during natural vision: diversity, consistency, and spatial sensitivity among AF face patch neurons. J. Neurosci. *35*, 5537–5548. https://doi.org/10.1523/JNEUROSCI.3825-14.2015.
9. Sheinberg, D.L., and Logothetis, N.K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. J. Neurosci. *21*, 1340–1350. https://doi.org/10.1523/JNEUROSCI.21-04-01340.2001.
10. Mosher, C.P., Zimmerman, P.E., and Gothard, K.M. (2014). Neurons in the monkey amygdala detect eye contact during naturalistic social interactions. Curr. Biol. *24*, 2459–2464. https://doi.org/10.1016/j.cub.2014.08.063.
11. Keysers, C., and Perrett, D.I. (2004). Demystifying social cognition: a Hebbian perspective. Trends Cogn. Sci. *8*, 501–507. https://doi.org/10.1016/j.tics.2004.09.005.
12. Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G.A. (2005). Observing others: multiple action representation in the frontal lobe. Science *310*, 332–336. https://doi.org/10.1126/science.1115593.
13. Singer, J.M., and Sheinberg, D.L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. J. Neurosci. *30*, 3133–3145. https://doi.org/10.1523/JNEUROSCI.3211-09.2010.
14. Vangeneugden, J., Pollick, F., and Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. Cereb. Cortex *19*, 593–611. https://doi.org/10.1093/cercor/bhn109.
15. Jellema, T., Maassen, G., and Perrett, D.I. (2004). Single cell integration of animate form, motion and location in the superior temporal cortex of the macaque monkey. Cereb. Cortex *14*, 781–790. https://doi.org/10.1093/cercor/bhh038.
16. Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G.A. (2011). Action observation circuits in the macaque monkey cortex. J. Neurosci. *31*, 3743–3756. https://doi.org/10.1523/JNEUROSCI.4803-10.2011.
17. Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. *2*, 661–670. https://doi.org/10.1038/35090060.
18. Ferrari, P.F., Gallese, V., Rizzolatti, G., and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. Eur. J. Neurosci. *17*, 1703–1714.
19. Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature *335*, 817–820. https://doi.org/10.1038/335817a0.
20. 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.
21. Meyer, T., Ramachandran, S., and Olson, C.R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. J. Neurosci. *34*, 9332–9337. https://doi.org/10.1523/JNEUROSCI.1215-14.2014.
22. Schwiedrzik, C.M., and Freiwald, W.A. (2017). High-level prediction signals in a low-level area of the macaque face-processing hierarchy. Neuron *96*, 89–97.e4. https://doi.org/10.1016/j.neuron.2017.09.007.
23. Ramachandran, S., Meyer, T., and Olson, C.R. (2017). Prediction suppression and surprise enhancement in monkey inferotemporal cortex. J. Neurophysiol. *118*, 374–382. https://doi.org/10.1152/jn.00136.2017.
24. Perrett, D.I., Xiao, D., Barraclough, N.E., Keysers, C., and Oram, M.W. (2009). Seeing the future: natural image sequences produce "anticipatory" neuronal activity and bias perceptual report. Q. J. Exp. Psychol. (Hove) *62*, 2081–2104. https://doi.org/10.1080/17470210902959279.
25. Leszczynski, M., and Schroeder, C.E. (2019). The role of neuronal oscillations in visual active sensing. Front. Integr. Neurosci. *13*, 32. https://doi.org/10.3389/fnint.2019.00032.
26. Henderson, J.M. (2003). Human gaze control during real-world scene perception. Trends Cogn. Sci. *7*, 498–504. https://doi.org/10.1016/j.tics.2003.09.006.
27. Mitchell, J.F., and Leopold, D.A. (2015). The marmoset monkey as a model for visual neuroscience. Neurosci. Res. *93*, 20–46. https://doi.org/10.1016/j.neures.2015.01.008.
28. Efron, R. (1967). The duration of the present. Ann. N. Y. Acad. Sci. *138*, 713–729. https://doi.org/10.1111/j.1749-6632.1967.tb55017.x.
29. Melcher, D., and Colby, C.L. (2008). Trans-saccadic perception. Trends Cogn. Sci. *12*, 466–473. https://doi.org/10.1016/j.tics.2008.09.003.
30. Wurtz, R.H., Joiner, W.M., and Berman, R.A. (2011). Neuronal mechanisms for visual stability: progress and problems. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 492–503. https://doi.org/10.1098/rstb.2010.0186.

31. Leopold, D.A., and Park, S.H. (2020). Studying the visual brain in its natural rhythm. NeuroImage *216*, 116790. https://doi.org/10.1016/j.neuroimage.2020.116790.

32. Bartels, A., and Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. Hum. Brain Mapp. *21*, 75–85. https://doi.org/10.1002/hbm.10153.

33. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. Science *303*, 1634–1640. https://doi.org/10.1126/science.1089506.

34. Hasson, U., Yang, E., Vallines, I., Heeger, D.J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. J. Neurosci. *28*, 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008.

35. Hasson, U., Malach, R., and Heeger, D.J. (2010). Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. *14*, 40–48. https://doi.org/10.1016/j.tics.2009.10.011.

36. Huth, A.G., Nishimoto, S., Vu, A.T., and Gallant, J.L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron *76*, 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014.

37. Hasson, U., and Honey, C.J. (2012). Future trends in neuroimaging: neural processes as expressed within real-life contexts. NeuroImage *62*, 1272–1278. https://doi.org/10.1016/j.neuroimage.2012.02.004.

38. Wang, H.X., Freeman, J., Merriam, E.P., Hasson, U., and Heeger, D.J. (2012). Temporal eye movement strategies during naturalistic viewing. J. Vis. *12*, 16. https://doi.org/10.1167/12.1.16.

39. Çukur, T., Nishimoto, S., Huth, A.G., and Gallant, J.L. (2013). Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. *16*, 763–770. https://doi.org/10.1038/nn.3381.

40. Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., and Castellanos, F.X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage *122*, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069.

41. Franchak, J.M., Heeger, D.J., Hasson, U., and Adolph, K.E. (2016). Free viewing gaze behavior in infants and adults. Infancy *21*, 262–287. https://doi.org/10.1111/infa.12119.

42. Vodrahalli, K., Chen, P.-H., Liang, Y., Baldassano, C., Chen, J., Yong, E., Honey, C., Hasson, U., Ramadge, P., Norman, K.A., and Arora, S. (2017). Mapping between fMRI responses to movies and their natural language annotations. NeuroImage *180*, 223–231. https://doi.org/10.1016/j.neuroimage.2017.06.042.

43. Finn, E.S., and Bandettini, P.A. (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. Neuroimage *235*, 117963. https://doi.org/10.1016/j.neuroimage.2021.117963.

44. Finn, E.S., Glerean, E., Khojandi, A.Y., Nielson, D., Molfese, P.J., Handwerker, D.A., and Bandettini, P.A. (2020). Idiosynchrony: from shared responses to individual differences during naturalistic neuroimaging. Neuroimage *215*, 116828. https://doi.org/10.1016/j.neuroimage.2020.116828.

45. Vanderwal, T., Eilbott, J., and Castellanos, F.X. (2019). Movies in the magnet: naturalistic paradigms in developmental functional neuroimaging. Dev. Cogn. Neurosci. *36*, 100600. https://doi.org/10.1016/j.dcn.2018.10.004.

46. Nastase, S.A., Goldstein, A., and Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. Neuroimage *222*, 117254. https://doi.org/10.1016/j.neuroimage.2020.117254.

47. Meer, J.N.V. van der, Breakspear, M., Chang, L.J., Sonkusare, S., and Cocchi, L. (2020). Movie viewing elicits rich and reliable brain state dynamics. Nat. Commun. *11*, 5004. https://doi.org/10.1038/s41467-020-18717-w.

48. Nentwich, M., Leszczynski, M., Russ, B.E., Hirsch, L., Markowitz, N., Sapru, K., Schroeder, C.E., Mehta, A., Bickel, S., and Parra, L.C. (2022). Semantic novelty modulates neural responses to visual change across the human brain. Preprint at biorxiv. https://doi.org/10.1101/2022.06.20.496467.

49. Shepherd, S.V., Steckenfinger, S.A., Hasson, U., and Ghazanfar, A.A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. Curr. Biol. *20*, 649–656. https://doi.org/10.1016/j.cub.2010.02.032.

50. Mantini, D., Hasson, U., Betti, V., Perrucci, M.G., Romani, G.L., Corbetta, M., Orban, G.A., and Vanduffel, W. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. Nat. Methods *9*, 277–282. https://doi.org/10.1038/nmeth.1868.

51. Mantini, D., Corbetta, M., Romani, G.L., Orban, G.A., and Vanduffel, W. (2013). Evolutionarily novel functional networks in the human brain? J. Neurosci. *33*, 3259–3275. https://doi.org/10.1523/JNEUROSCI.4392-12.2013.

52. Russ, B.E., and Leopold, D.A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. Neuroimage *109*, 84–94. https://doi.org/10.1016/j.neuroimage.2015.01.012.

53. Russ, B.E., Kaneko, T., Saleem, K.S., Berman, R.A., and Leopold, D.A. (2016). Distinct fMRI Responses to Self-Induced versus Stimulus Motion during Free Viewing in the Macaque. J. Neurosci. *36*, 9580–9589. https://doi.org/10.1523/JNEUROSCI.1152-16.2016.

54. Park, S.H., Russ, B.E., McMahon, D.B.T., Koyano, K.W., Berman, R.A., and Leopold, D.A. (2017). Functional subpopulations of neurons in a macaque face patch revealed by single-unit fMRI mapping. Neuron *95*, 971–981.e5. https://doi.org/10.1016/j.neuron.2017.07.014.

55. Sliwa, J., and Freiwald, W.A. (2017). A dedicated network for social interaction processing in the primate brain. Science *356*, 745–749. https://doi.org/10.1126/science.aam6383.

56. Ortiz-Rios, M., Balezeau, F., Haag, M., Schmid, M.C., and Kaiser, M. (2021). Dynamic reconfiguration of macaque brain networks during natural vision. Neuroimage *244*, 118615. https://doi.org/10.1016/j.neuroimage.2021.118615.

57. Sakon, J.J., and Suzuki, W.A. (2021). Neural evidence for recognition of naturalistic videos in monkey hippocampus. Hippocampus *31*, 916–932. https://doi.org/10.1002/hipo.23335.

58. Ainsworth, M., Sallet, J., Joly, O., Kyriazis, D., Kriegeskorte, N., Duncan, J., Schüffelgen, U., Rushworth, M.F.S., and Bell, A.H. (2021). Viewing ambiguous social interactions increases functional connectivity between frontal and temporal nodes of the social brain. J. Neurosci. *41*, 6070–6086. https://doi.org/10.1523/JNEUROSCI.0870-20.2021.

59. McMahon, D.B.T., Bondar, I.V., Afuwape, O.A.T., Ide, D.C., and Leopold, D.A. (2014). One month in the life of a neuron: longitudinal single-unit electrophysiology in the monkey visual system. J. Neurophysiol. *112*, 1748–1762. https://doi.org/10.1152/jn.00052.2014.

60. McMahon, D.B.T., Jones, A.P., Bondar, I.V., and Leopold, D.A. (2014). Face-selective neurons maintain consistent visual responses across months. Proc. Natl. Acad. Sci. USA *111*, 8251–8256. https://doi.org/10.1073/pnas.1318331111.

61. Koyano, K.W., Jones, A.P., McMahon, D.B.T., Waidmann, E.N., Russ, B.E., and Leopold, D.A. (2021). Dynamic suppression of average facial structure shapes neural tuning in three macaque face patches. Curr. Biol. *31*, 1–12.e5. https://doi.org/10.1016/j.cub.2020.09.070.

62. Freiwald, W.A., Tsao, D.Y., and Livingstone, M.S. (2009). A face feature space in the macaque temporal lobe. Nat. Neurosci. *12*, 1187–1196. https://doi.org/10.1038/nn.2363.

63. Tsao, D.Y., Moeller, S., and Freiwald, W.A. (2008). Comparing face patch systems in macaques and humans. Proc. Natl. Acad. Sci. USA *105*, 19514–19519. https://doi.org/10.1073/pnas.0809662105.

64. Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., and Tootell, R.B.H. (2003). Faces and objects in macaque cerebral cortex. Nat. Neurosci. *6*, 989–995. https://doi.org/10.1038/nn1111.

65. Freiwald, W.A., and Tsao, D.Y. (2014). Neurons that keep a straight face. Proc. Natl. Acad. Sci. USA *111*, 7894–7895. https://doi.org/10.1073/pnas.1406865111.

66. Tsao, D.Y., Schweers, N., Moeller, S., and Freiwald, W.A. (2008). Patches of face-selective cortex in the macaque frontal lobe. Nat. Neurosci. *11*, 877–879. https://doi.org/10.1038/nn.2158.

67. Russ, B.E., Petkov, C.I., Kwok, S.C., Zhu, Q., Belin, P., Vanduffel, W., and Hamed, S.B. (2021). Common functional localizers to enhance NHP & cross-species neuroscience imaging research. Neuroimage *237*, 118203. https://doi.org/10.1016/j.neuroimage.2021.118203.

68. Popivanov, I.D., Jastorff, J., Vanduffel, W., and Vogels, R. (2012). Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. NeuroImage *63*, 723–741. https://doi.org/10.1016/j.neuroimage.2012.07.013.

69. Maunsell, J.H., and Van Essen, D.C.V. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J. Neurophysiol. *49*, 1127–1147.

70. Maunsell, J.H., and Newsome, W.T. (1987). Visual processing in monkey extrastriate cortex. Annu. Rev. Neurosci. *10*, 363–401. https://doi.org/10.1146/annurev.ne.10.030187.002051.

71. Boussaoud, D., Desimone, R., and Ungerleider, L.G. (1991). Visual topography of area TEO in the macaque. J. Comp. Neurol. *306*, 554–575. https://doi.org/10.1002/cne.903060403.

72. Taubert, J., Wardle, S.G., Tardiff, C.T., Patterson, A., Yu, D., and Baker, C.I. (2022). Clutter substantially reduces selectivity for peripheral faces in the macaque brain. J. Neurosci. *42*, 6739–6750. https://doi.org/10.1523/JNEUROSCI.0232-22.2022.

73. Aparicio, P.L., Issa, E.B., and DiCarlo, J.J. (2016). Neurophysiological organization of the middle face patch in macaque inferior temporal cortex. J. Neurosci. *36*, 12729–12745. https://doi.org/10.1523/JNEUROSCI.0237-16.2016.

74. Bell, A.H., Malecek, N.J., Morin, E.L., Hadj-Bouziane, F., Tootell, R.B.H., and Ungerleider, L.G. (2011). Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. J. Neurosci. *31*, 12229–12240. https://doi.org/10.1523/JNEUROSCI.5865-10.2011.

75. Popivanov, I.D., Jastorff, J., Vanduffel, W., and Vogels, R. (2014). Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. J. Neurosci. *34*, 95–111. https://doi.org/10.1523/JNEUROSCI.2748-13.2014.

76. Schiller, P.H., and Malpeli, J.G. (1977). Properties and tectal projections of monkey retinal ganglion cells. J. Neurophysiol. *40*, 428–445. https://doi.org/10.1152/jn.1977.40.6.1443-s.

77. Müller, J.R., Metha, A.B., Krauskopf, J., and Lennie, P. (2001). Information conveyed by onset transients in responses of striate cortical neurons. J. Neurosci. *21*, 6978–6990. https://doi.org/10.1523/JNEUROSCI.21-17-06978.2001.

78. Oram, M.W., and Perrett, D.I. (1992). Time course of neural responses discriminating different views of the face and head. J. Neurophysiol. *68*, 70–84. https://doi.org/10.1152/jn.1992.68.1.70.

79. Feuerriegel, D., Vogels, R., and Kovács, G. (2021). Evaluating the evidence for expectation suppression in the visual system. Neurosci. Biobehav. Rev. *126*, 368–381. https://doi.org/10.1016/j.neubiorev.2021.04.002.

80. Kohn, A., and Movshon, J.A. (2004). Adaptation changes the direction tuning of macaque MT neurons. Nat. Neurosci. *7*, 764–772. https://doi.org/10.1038/nn1267.

81. Jeyabalaratnam, J., Bharmauria, V., Bachatene, L., Cattan, S., Angers, A., and Molotchnikoff, S. (2013). Adaptation shifts preferred orientation of tuning curve in the mouse visual cortex. PLoS ONE *8*, e64294. https://doi.org/10.1371/journal.pone.0064294.

82. Williams, N.P., and Olson, C.R. (2022). Contribution of individual features to repetition suppression in macaque inferotemporal cortex. J. Neurophysiol. *128*, 378–394. https://doi.org/10.1152/jn.00475.2021.

83. Williams, N.P., and Olson, C.R. (2022). Independent repetition suppression in macaque area V2 and inferotemporal cortex. J. Neurophysiol. *128*, 1421–1434. https://doi.org/10.1152/jn.00043.2022.

84. DiCarlo, J.J., and Maunsell, J.H. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. Nat. Neurosci. *3*, 814–821. https://doi.org/10.1038/77722.

85. DiCarlo, J.J., and Maunsell, J.H.R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. J. Neurophysiol. *89*, 3264–3278. https://doi.org/10.1152/jn.00358.2002.

86. Anderson, B., Mruczek, R.E.B., Kawasaki, K., and Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. Cereb. Cortex *18*, 2540–2552. https://doi.org/10.1093/cercor/bhn015.

87. Woloszyn, L., and Sheinberg, D.L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. Neuron *74*, 193–205. https://doi.org/10.1016/j.neuron.2012.01.032.

88. Miller, E.K., Li, L., and Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. Science *254*, 1377–1379. https://doi.org/10.1126/science.1962197.

89. Rolls, E.T., Baylis, G.C., Hasselmo, M.E., and Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. Exp. Brain Res. *76*, 153–164. https://doi.org/10.1007/BF00253632.

90. Kourtzi, Z., and DiCarlo, J.J. (2006). Learning and neural plasticity in visual object recognition. Curr. Opin. Neurobiol. *16*, 152–158. https://doi.org/10.1016/j.conb.2006.03.012.

91. Meyer, T., Walker, C., Cho, R.Y., and Olson, C.R. (2014). Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. Nat. Neurosci. *17*, 1388–1394. https://doi.org/10.1038/nn.3794.

92. Vinken, K., Op de Beeck, H.P., and Vogels, R. (2018). Face repetition probability does not affect repetition suppression in macaque inferotemporal cortex. J. Neurosci. *38*, 7492–7504. https://doi.org/10.1523/jneurosci.0462-18.2018.

93. Kuravi, P., and Vogels, R. (2017). Effect of adapter duration on repetition suppression in inferior temporal cortex. Sci. Rep. *7*, 3162. https://doi.org/10.1038/s41598-017-03172-3.

94. Nelissen, K., Vanduffel, W., and Orban, G.A. (2006). Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. J. Neurosci. *26*, 5929–5947. https://doi.org/10.1523/JNEUROSCI.0824-06.2006.

95. Polosecki, P., Moeller, S., Schweers, N., Romanski, L.M., Tsao, D.Y., and Freiwald, W.A. (2013). Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. J. Neurosci. *33*, 11768–11773. https://doi.org/10.1523/JNEUROSCI.5402-11.2013.

96. Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B.B., and Ungerleider, L.G. (2012). Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. J. Neurosci. *32*, 15952–15962. https://doi.org/10.1523/JNEUROSCI.1992-12.2012.

97. Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436.

98. Chaure, F.J., Rey, H.G., and Quian Quiroga, R.Q. (2018). A novel and fully automatic spike-sorting implementation with variable number of features. J. Neurophysiol. *120*, 1859–1871. https://doi.org/10.1152/jn.00339.2018.

99. Hwang, J., Mitz, A.R., and Murray, E.A. (2019). NIMH MonkeyLogic: behavioral control and data acquisition in MATLAB. J. Neurosci. Methods *323*, 13–21. https://doi.org/10.1016/j.jneumeth.2019.05.002.

100. Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. *16*, 1661–1687. https://doi.org/10.1162/089976604774201631.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
MATLAB	Mathworks	https://uk.mathworks.com/products/matlab.html
Analysis Code	Custom Matlab pipeline	https://github.com/beruss/Temporal_Context https://doi.org/10.5281/zenodo.7419223
Psychtoolbox	Brainard[97]	http://psychtoolbox.org
Wave_clus	Chaure et al.[98]	https://github.com/csn-le/wave_clus
NIMH MonkeyLogic	Hwang et al.[99]	https://monkeylogic.nimh.nih.gov/index.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for data can be directed to either the lead contact, Brian E. Russ (brian.russ@nki.rfmh.org), or the senior author, David A. Leopold (leopoldd@mail.nih.gov).

Materials availability

The current study did not generate new unique reagents or materials.

Data and code availability

- The data reported in this manuscript will be shared upon request by the lead contact or senior author.
- Analysis code for the comparison of temporal context effects have been deposited on Github (https://github.com/beruss/Temporal_Context; https://doi.org/10.5281/zenodo.7419223).
- The lead contact will provide any additional information required to reanalyze the data reported upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Six rhesus macaques (*macaca mulatta*) participated in the current experiments, including 2 females (ages 4-13 years at the time of the study) and 4 males (ages 4-10 years). All procedures were approved by the Animal Care and Use Committee of the National Institute of Mental Health, and followed the guidelines of the United States National Institutes of Health.

METHOD DETAILS

Surgical procedures

In an initial surgery, each animal was implanted with a fiberglass head-post, for immobilizing the head during recording sessions. After behavioral training, described below, microwire bundle electrodes (Microprobes) and adjustable microdrives were implanted into the face patches during a second surgical procedure (see Figure 1A). The electrodes, microdrives, and surgical procedures are described in previous publications.[59,60] In four subjects (two female, three male) the electrodes were directed to the anterior fundus face patch (AF): Subject T received bilateral 32 channel electrodes bundles; Subject R received a 64 channel bundle in the right hemisphere; and Subject S and Subject SR received a 64 channel bundle in the left hemisphere. In two other subjects (both male), electrodes were directed to the anterior medial (AM) face patches: Subject M received one 64 channel microwire bundle in the left AM; Subject D was implanted bilaterally with 128 channel microwire bundles, with the left bundle located within AM and the right bundle just adjacent to AM.

Following electrode implantation, the subjects participated in a series of behavioral tasks. During this period, the subjects' access to liquid was controlled, such that they were motivated to receive liquid reward in exchange for participating in the experimental tasks. In the case that they did not receive sufficient liquid reward in a given session, they were supplemented with liquids in their home cage. Hydration and weights were monitored throughout the experimental period to ensure the animals' health and well-being.

Visual stimuli and behavioral monitoring

Subjects viewed a range of intact movies, randomized snippets and static frames extracted from the movies, and flashed categorical images. Image presentation and gaze fixation monitoring were under computer control using a combination of custom written QNX[9]

and Psychtoolbox[97] or a MonkeyLogic[99] code. Neurophysiological signals were recorded using TDT hardware and software. The subjects' horizontal and vertical eye position was continuously monitored using a 60 Hz Eyelink (SR Research) system, which converted position to analog voltages that were passed to the neurophysiological and behavioral recording software. At the start of each session, the subject's eye position was centered and calibrated using a nine-position grid with 6° visual angle separation. This calibration procedure was rerun repeatedly throughout each session to ensure good calibration during all experimental trials. Blocks of intact movie presentation, randomized snippet presentation, and categorical image presentation were randomly interleaved (Figure 1C) and took the following form.

Intact movie presentation

The monkeys repeatedly viewed multiple repetitions of a 5 min movie containing a wide range of interactions, mostly between other monkeys.[8,52,53] The movies were presented on an LCD screen with a horizontal dimension of 15 dva. While there was no fixation, subjects were required to maintaining their gaze within the movie frame, for which they received a drop of juice reward each 2 - 2.5 s.

Randomized snippet presentation

The content from the movie above was broken down into short movie snippets and static frames, which the monkey viewed in randomized order. The original 5-min movie was first divided into three hundred 1 s segments, which were subsequently used to generate snippets of different lengths. The snippet durations of 800, 250, and 100 ms spanned the center frame of each 1 s movie segment. For example, the 800 ms snippet from the 151st second would present the frames of the movie from 150.10 to 150.90 s, whereas the 100 ms snippet from the same segment would display the frames from 150.45 to 150.55 s. The static frame condition presented the center frame for 100 ms. For each snippet length and the static frame condition, all 300 stimuli were presented repeatedly and in randomized order (Figure 1B). Other aspects of the stimulus were the same as in the intact movie condition, other than the 200 ms blank black screen now inserted between each presentation. As above, the subjects were permitted view the entire snippet content within the brief presentation period and were rewarded with a drop of juice and a 1.0 s break after completing approximately 2 s of trials, depending slightly on the snippet lengths. For an inappropriate fixation break, the animal would forfeit its juice for that period and incur an additional 750 ms timeout followed by repetition of the previous stimulus. For each block of the snippet task, the stimulus duration was held constant until all 300 stimuli were presented.

Categorical image presentation

In these blocks, the monkeys serially viewed brief image presentations drawing from four categories, with 40 images from each category. The image categories were monkey faces, human faces, objects, natural scenes. The stimuli were randomized and presented at 12 dva for 100 ms with a 200 ms inter-stimulus interval. As above, the subjects were permitted to direct their gaze anywhere within the image. If gaze was maintained in the window for seven successive images, the subject received a liquid reward and a 1000 ms intertrial break. An inappropriate fixation break led to forfeit of reward and an additional 750 timeout. This condition served the dual role of evaluating category selectivity and establishing a ''fingerprint'' for tracking the identity of single neurons across sessions.[60]

QUANTIFICATION AND STATISTICAL ANALYSIS

Longitudinal neural recording

Given the large number of conditions and presentations required in the current study, our recording paradigm hinged on the capacity to accurately record the same neurons across multiple sessions. This longitudinal tracking of individual cells required particular care in spike sorting and rules for determining the continuous isolation of neurons across sessions.

Spike sorting

Following each session of recording, channels were cleaned and then submitted to an automatic spike sorting algorithm. The cleaning procedure, run separately on each bank of 16 channels sharing a connector, involved regressing out the first principal component across channels. This procedure was important in our experience to combat certain types of movement and chewing-related artifacts that were shared across channels. Following this cleaning procedure, the automatic sorting program wave_clus[98,100] was applied to identify and sort action potentials on individual channels. Candidate spike waveforms were first marked as any voltages exceeding four standard deviations of the noise. These candidate waveforms were then sorted automatically using PCA space, rendering a sequence of timestamps for between 0 and 4 usable waveforms from each microwire channel.

Tracking neurons across days

Verifying the isolation of a given neuron across days required a combined consideration of spike waveform, as determined by the automated spike sorting algorithm,[98,100] and an independent comparison of the response fingerprint, or selectivity to the flashed images (see Figure S1 for example of isolation and fingerprinting). A cell was determined to be the same as that isolated on a previous day if it (1) was recorded from the same channel, (2) had the same fingerprint signature of response selectivity, and (3) had a similar waveform. Of these three criteria, the third was the only one that was subject to flexibility, since it is well known that small positional changes in the relationship between an electrode and a neuron can change the action potential shape. However, any neuron recorded from a different electrode or having a different selectivity fingerprint across days were considered to be a different neuron. Applying these criteria, we excluded from analysis neurons that were held for fewer than 75% of the sessions, and therefore had insufficient numbers of trials for our statistical analyses. All visually responsive neurons, defined as those giving statistically significant responses in any one of the conditions, were investigated for the effects of temporal continuity, with a separate analysis investigating the specific factor of face selectivity.

Data analysis

The main data analysis focused on comparing the neural responses to the intact movie with those elicited by the same visual content presented during the randomized 800 ms snippets. Based on pilot analysis, we divided this analysis into two distinct time windows (Figure 1D). The first analysis window analyzed spiking from 100-300 ms following the onset of the snippet, targeting the contribution of the initial response. The second analysis window then focused on the delayed, or extended, response between 300-900 ms.

For each time window, we compared the mean spike count over multiple presentations of the 300 snippets to the corresponding mean spike counts elicited by the corresponding frames in the intact movie. In the main analysis, we compared the two conditions using a linear regression, from which we derived the goodness-of-fit (r^2) to assess the effects of context (see Figure 1E). We also reordered the randomized snippet responses into their original sequence in order to visualize any temporal disruption (see Figure 1B).

For the main analysis, the r^2 parameter provided insight into the disruption in visual selectivity introduced by randomized snippet presentation. For example, r^2 values of 1.0 would indicate that a neuron exhibits identical responses to the same visual content across the continuous versus randomized snippet presentation conditions, whereas an r^2 value of 0.0 would correspond to a complete disruption of the stimulus selectivity, where the neural responses to the specific content were uncorrelated. For intermediate values, the r^2 indicates the tightness of the stimulus selectivity relationship across conditions, or response variance explained.

In addition to the 800 ms snippets, we also analyzed neural responses to equivalent snippets of 250 and 100 ms, as well as static images of the center frame presented for 100 ms. In those cases, the presentation times were too short to allow for a separate consideration of selectivity before and after the initial stimulus onset, thus a standard response window of 50 to 250 ms was used for evaluation of the onset response.

We calculated a gaze position dependent goodness-of-fit for the early and late period responses of the 800ms snippets. For each epoch, we then evaluated the dispersion of eye positions across presentations based on the Euclidean distance to the mean of the within context movie. This created a set of neural responses that were median split into clustered, or near, and spread-out, or far-away, fixations, as defined by a trials eye positions proximity to the other presentations of the same scene. Randomized snippet trials were then sorted based on the spatial proximity of the mean gaze to that observed during continuous viewing of the same content. This allowed us to divide the snippet responses into epochs exhibiting near and far-away gaze positions relative to that observed during the continuous movie. From this data, we calculated goodness-of-fit of neural responses separately for the snippet trials in which the gaze did or did not approximate that observed during the continuous movie.

We used the same response windows for each clip to calculate a Context Index. Here we define the Context Index as the difference between the within context response and the out-of-context response divided by the sum of both context's responses. Therefore, if a cell responded strongly to the within context condition and not at all the out-of-context response, the Context Index would be equal to 1, showing a strong preference for the within context condition. However, if the response to both conditions was nearly equal, the Context Index would equal 0, or no preference for either condition.

Neuronal latencies were estimated from their responses to the static frames using the following procedures. We first calculated the average and standard deviation of the firing rate during the 50ms prior to stimulus onset. Then for each of the 300 snippets, we calculated a response onset to their average spike rates. Using a sliding window, we found the moment were spiking activity changed (increased or decreased) by 3std of the baseline firing rate and was maintained for a minimum of 15ms. We then averaged the obtained latencies for all 300 snippets to find the mean latency for a given neuron.