Multiple objects evoke fluctuating responses in several regions of the visual pathway

Authors:

Meredith N. Schmehl[1-3], Valeria C. Caruso[4], Yunran Chen[5], Na Young Jun[1,3], Shawn M. Willett[6], Jeff T. Mohl[7], Douglas A. Ruff[8], Marlene Cohen[8], Akinori F. Ebihara[9], Winrich Freiwald[9], Surya T. Tokdar[3,5], Jennifer M. Groh[1-3,10-12]

Affiliations:

1. Department of Neurobiology, Duke University
2. Center for Cognitive Neuroscience, Duke University
3. Duke Institute for Brain Sciences, Duke University
4. Department of Psychiatry, University of Michigan
5. Department of Statistical Science, Duke University
6. Department of Ophthalmology, University of Pittsburgh
7. American Medical Group Association
8. Department of Neurobiology, University of Chicago
9. The Rockefeller University
10. Department of Psychology & Neuroscience, Duke University
11. Department of Computer Science, Duke University
12. Department of Biomedical Engineering, Duke University

Abstract

How neural representations preserve information about multiple stimuli is mysterious. Because tuning of individual neurons is coarse (for example, visual receptive field diameters can exceed perceptual resolution), the populations of neurons potentially responsive to each individual stimulus can overlap, raising the question of how information about each item might be segregated and preserved in the population. We recently reported evidence for a potential solution to this problem: when two stimuli were present, some neurons in the macaque visual cortical areas V1 and V4 exhibited fluctuating firing patterns, as if they responded to only one individual stimulus at a time. However, whether such an information encoding strategy is ubiquitous in the visual pathway and thus could constitute a general phenomenon remains unknown. Here we provide new evidence that such fluctuating activity is also evoked by multiple stimuli in visual areas responsible for processing visual motion (middle temporal visual area, MT), and faces (middle fundus and anterolateral face patches in inferotemporal cortex – areas MF and AL), thus extending the scope of circumstances in which fluctuating activity is observed. Furthermore, consistent with our previous results in the early visual area V1, MT exhibits fluctuations between the representations of two stimuli when these form distinguishable objects but not when they fuse into one perceived object, suggesting that fluctuating activity patterns may underlie visual object formation. Taken together, these findings point toward an updated model of how the brain preserves sensory information about multiple stimuli for subsequent processing and behavioral action.

Impact Statement

We find neural fluctuations in multiple areas along the visual cortical hierarchy that could allow the brain to represent distinct co-occurring visual stimuli.

Introduction

Neurons are broadly tuned. For example, even in primary visual cortex, neurons typically have receptive fields that are several times larger than spatial acuity at the perceptual level (Dow et al., 1981; Smith et al., 2001; Motter, 2009; Freeman & Simoncelli, 2011; Grill-Spector et al., 2017; Keliris et al., 2019). Due to this broad tuning, a given neuron may be capable of responding to more than one sensory object in a scene. When multiple stimuli are present, how does the brain preserve information about each item?

We recently reported a potential solution to this problem: When two stimuli are present, individual neurons may fluctuate between responding to only one or the other stimulus at a time, allowing dynamic populations of neurons to encode different stimuli at the same time. In our earlier work, we developed statistical approaches to evaluate whether the fluctuating activity patterns predicted from such a scheme exist (Caruso et al., 2018; Mohl et al., 2020; Glynn et al., 2021), and we reported evidence for this phenomenon in both auditory (Caruso et al., 2018) and visual (Caruso et al., 2018; Jun et al., 2022) areas. In particular, we found evidence for fluctuating activity in the inferior colliculus (Caruso et al., 2018), primary visual cortex (V1), area V4 (Jun et al., 2022), and the middle fundus (MF) face patch of inferotemporal (IT) cortex (Caruso et al., 2018) when two distinguishable stimuli were presented. We did not observe fluctuating activity in V1 when two stimuli fused to form a common object (i.e., two gratings superimposed to form a “plaid”) (Jun et al., 2022).

In short, we have identified activity fluctuation in a wide array of brain areas, from early auditory regions to early, middle, and late visual cortical regions. This breadth of brain areas suggests that fluctuating activity may reflect a general mechanism of neural representation supporting the encoding of multiple stimuli. To test this possibility, here we ask three critical questions: (a) Are

such fluctuations widespread within the visual pathway? (b) Do they occur only when the two stimuli are perceptually distinct? (c) Are they more prevalent when stimuli of a given type compete for representation in a neural subpopulation specialized for stimuli of that particular type? To address these questions, we investigated neural responses to pairs of stimuli in three cortical visual areas: middle temporal visual area (MT), the middle fundus face patch (MF), and the anterolateral face patch (AL).

We report that neuronal activity in all three areas fluctuated across trials between responses that appeared to be drawn from the response distributions corresponding to the respective individual stimuli. In MT, such fluctuating activity was present only when the two visual stimuli were distinct (adjacent gratings or dot patches), and not when they were superimposed and fused to form single objects (plaids). In MF and AL, fluctuating neuronal activity was present when two faces were presented (i.e., stimuli for which the patches are specialized) as well as when a face and a non-face object were presented, but in MF the fluctuations were more common when two faces were presented (i.e. two stimuli of the category for which the neurons appear specialized) than when a face and a non-face object (different stimulus categories) were presented.

Together, these findings suggest that fluctuating activity patterns are a widespread phenomenon connected to the computations underlying object formation/identification. Object separation appears generally necessary for fluctuating activity to occur, and specialization for certain types of stimuli can also affect the prevalence of fluctuations across the neural population.

Materials and Methods

Introduction to Datasets

We analyzed multiple datasets from two visual cortical regions (MT and IT cortex). All datasets were collected in awake, behaving macaques (MT: *Macaca mulatta;* IT: *Macaca mulatta* and *Macaca fascicularis*). The data were originally collected at the University of Pittsburgh (Ruff & Cohen, 2016; Ruff et al., 2016), Rockefeller University (Ebihara, 2015), and the German Primate Center in Goettingen (Li et al., 2016). All procedures were conducted in accordance with applicable US, German, and European guidelines or regulations and were approved by the relevant compliance committees at the relevant institutions (US guidelines: National Institutes of Health (NIH) Pub. No. 86-23, Revised 1985; Pittsburgh MT grating datasets: Institutional Animal Care and Use Committees of the University of Pittsburgh and Carnegie Mellon University (Protocol #: 20067560, PHS Assurance Number: D16-00118); IT datasets: The Rockefeller University Institutional Animal Care and Use Committee (Protocol #: 21104-H USDA, PHS Assurance Number: A3081-01); Goettingen MT dot patch dataset: [Niedersächsisches Landesamtfür Verbraucherschutzund Lebensmittelsicherheit (LAVES)] permit numbers 33.42502/08-07.02 and 33.14.42502-04-064/07). All surgical procedures were conducted aseptically and with suitable anesthetics and analgesics.

Our goal was to assess fluctuating neural activity in multiple visual areas and in response to multiple stimuli (**Table 1**). Monkeys were performing either fixation or other behavioral tasks in which the stimuli in question were unattended.

Table 1: Comparison of Datasets

Dataset Name	Citation	Brain Area	Stimuli	Stimulus Locations	Task	Recording Method
MT adjacent gratings	(Ruff & Cohen, 2016)	MT	Drifting images of two gratings	Adjacent (3 degrees eccentric)	Detect orientation change of a different, ipsilateral stimulus	Array, single and multi-unit

MT dot patches	(Li et al., 2016)		Moving dots	Adjacent	Detect change in fixation stimulus	Single electrode, single unit
MT plaid	(Ruff et al., 2016)		Drifting images of gratings and plaids	Same location	Passive fixation	Array, single and multi-unit
IT face patch MF	(Ebihara, 2015; Caruso et al., 2018)	MF	Images of faces and objects	Adjacent	Passive fixation	Single electrode, single unit
IT face patch AL	(Ebihara, 2015)	AL	Images of faces and objects	Adjacent	Passive fixation	Single electrode, single unit

For area MT, we used three datasets. The first dataset included single- and multi-unit responses to drifting grating stimuli positioned adjacent to one another (**Figure 1A.1-2**) (Ruff & Cohen, 2016). The experiment involved simultaneous recordings from V1, V4, and MT. As we have previously identified fluctuating response patterns in V1 and V4 (Jun et al., 2022), this MT dataset provides an opportunity to compare the presence of fluctuating activity across visual areas under the same experimental conditions (and some of the same recording sessions) as in our previous work. However, this dataset was comparatively small (as detailed further below), making additional data desirable.

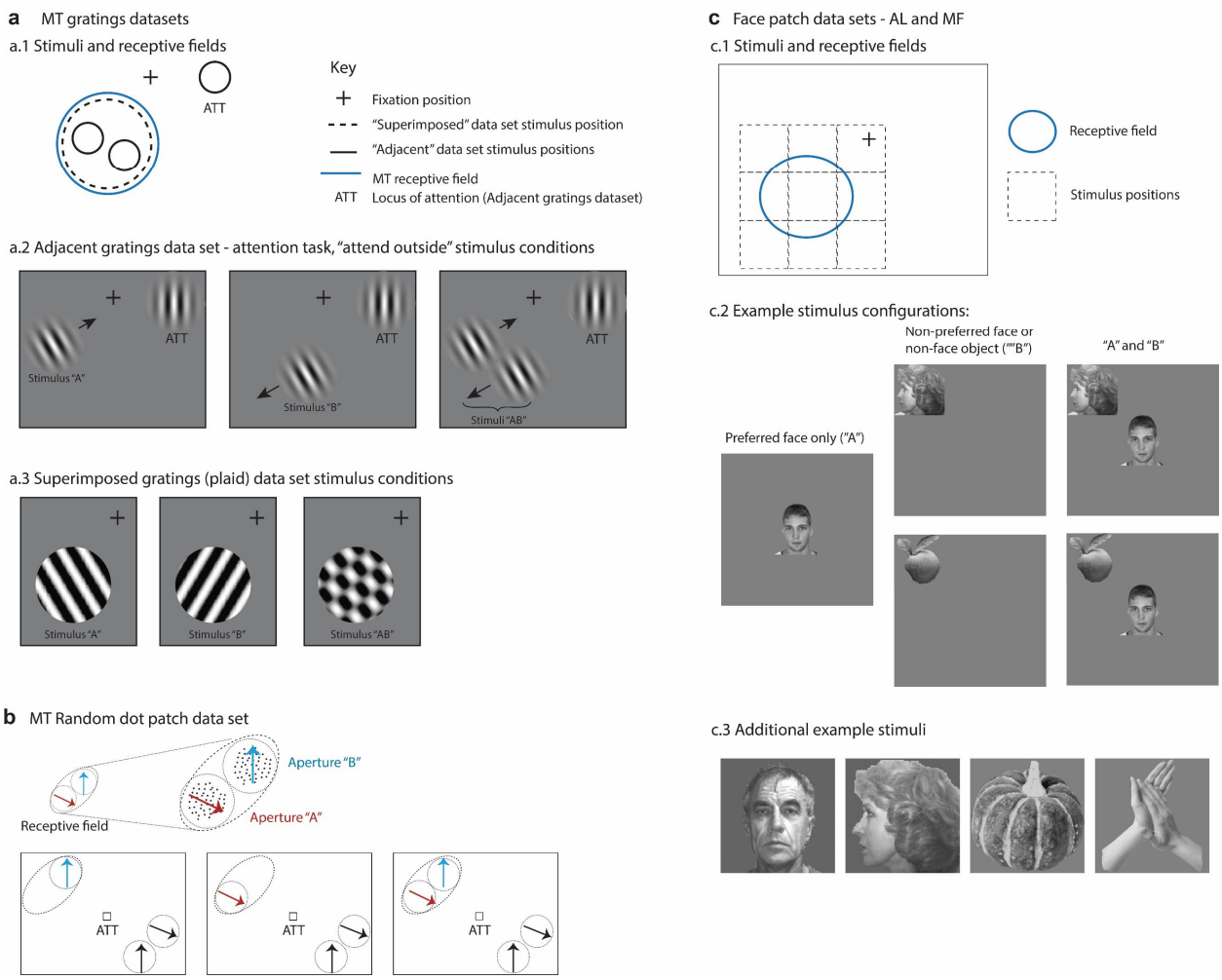


Figure 1: *Description of datasets. a) MT gratings datasets from (Ruff & Cohen, 2016) and (Ruff et al., 2016). a.1). For the adjacent gratings, the two stimuli were placed side by side and roughly within the MT neuron's receptive field. The superimposed gratings were larger and each was placed at the same position, also within the MT neuron's receptive field. a.2) Monkeys performed an attention task during the presentation of the adjacent gratings; only trials in which attention was successfully directed to the stimulus in the ipsilateral hemifield were included in this study. a.3) Monkeys performed a fixation task during presentation of the superimposed gratings. b) MT random dot patch dataset from (Li et al., 2016). Patches of moving dots were presented within apertures placed within the receptive field of an MT neuron. Attention was directed to the fixation target for the trials included in the present study. c. Face patch stimulus conditions (Ebihara, 2015). c.1) Monkeys performed a fixation task and stimuli were presented either at the center of the receptive field or from one of 8 locations surrounding it. c.2) "Preferred" faces were faces that evoked a strong response, and were presented in the center of the receptive field. Non-preferred faces or non-face objects were presented at one of the other locations. Combinations of stimuli involved the preferred face in the center and one of the non-preferred stimuli at one of the adjacent locations. c.3) Additional examples of stimuli used in the face patch datasets.*

Accordingly, we included a second dataset from a previously published study involving single- and multi-unit responses to adjacent random dot stimuli (**Figure 1B**) (Li et al., 2016). Both of these datasets were recorded while monkeys performed a behavioral task in which attention was directed away from the stimuli in the MT receptive field: toward the opposite hemifield (Ruff & Cohen, 2016) or to the fixation point (Li et al., 2016).

In both of these MT datasets, the stimuli can be considered as distinct and distinguishable objects from each other, regardless of whether they are gratings or patches of moving dots, because they are presented at separate locations. Our third MT dataset provides a comparison to these first two. In this dataset, individual gratings were presented at the same location in space, such that when they were presented together, they fused to form a plaid stimulus (**Figure 1A.**[1,3]). This dataset was originally published in (Ruff et al., 2016) and involved simultaneous V1 and V4 recordings previously analyzed for fluctuating activity in (Jun et al., 2022). In this dataset, monkeys performed a simple fixation task.

The specific stimulus details for these MT datasets are as follows. In the adjacent gratings dataset, the grating sizes were selected to lie within the MT receptive field under study and did not exceed 1 degree in diameter. The gratings used in the superimposed dataset were larger, ranging between 7 and 13 degrees in diameter, and covered the MT receptive field. The sizes of the random dot patches are not available, but they were depicted at adjacent positions within the MT receptive field (Li et al., 2016).

For the IT cortex, we included datasets involving single-unit recordings in two areas of the face patch system, areas MF and AL, during a fixation task (**Figure 1C**) (Ebihara, 2015). We have

previously reported fluctuating activity in face patch MF (Caruso et al., 2018), but here we have re-analyzed the MF dataset (see “Analysis: response periods and inclusion criteria”) to provide detailed results for different types of stimuli (face-face vs. face-object pairs), allowing a comparison of results in the two face patches (MF and AL) in the same experimental conditions.

For both MF and AL, we assessed neural responses to image pairs drawn from three visual stimuli selected from a pool of faces and objects: a “preferred face,” (strong neural response), a “non-preferred face,” (little to no neural response), and a “non-preferred object” (little to no neural response). These stimuli were randomly presented either alone or in pairs, with each pair including the preferred face, positioned at the center of the neuron’s receptive field, and one of the non-preferred stimuli, presented adjacent to the preferred face at one of eight equidistant locations. Stimuli were 4x4 degrees in size.

As in (Caruso et al., 2018), we did not incorporate the exact location of the non-preferred face/object into the present analysis, but any excessive heterogeneity in the responses of the non-preferred stimuli due to location would likely result in exclusion of the condition from further analysis (see “Analysis: response periods and inclusion criteria”).

Analysis: response periods and inclusion criteria

<u>Response periods:</u>

For each dataset, we compared neural responses (see “Data Analysis”) using a spike counting window of 200 ms, the minimum stimulus presentation window across all datasets. This window was shifted by the approximate response latency of the brain region, specifically 30 ms for MT and 50 ms for IT, i.e., spikes were counted in a window from 30 to 230 ms after stimulus onset

for MT or 50 to 250 ms for IT. For the IT datasets, the stimuli were presented for 400 ms, and results were similar when we analyzed the data in a 50-450 ms time window.

Trial inclusion criteria:

Only correctly performed trials were analyzed. For fixation tasks, this meant that the monkey maintained fixation for the required interval (MT plaid dataset, IT datasets). For attention tasks, this meant that the monkey correctly identified the change in the attended stimulus, which was outside the receptive field of the neurons under study (MT gratings dataset: the motion change in a grating presented in the ipsilateral field; MT dot patch dataset: the change in color in the fixation target). In addition, trials were excluded from the MT gratings dataset if any microsaccades were identified during the stimulus presentation period. Microsaccades were defined as periods of time in which eye speed exceeded the mean eye speed in the fixation epoch by more than 6 standard deviations (Engbert & Kliegl, 2003).

"Triplet" definition and inclusion criteria:

Our analysis of fluctuating activity patterns relies on comparing the activity evoked on trials in which two stimuli, A and B, are presented simultaneously (AB) to trials in which only the A or only the B stimulus was presented. We refer to the corresponding A-alone, B-alone, and combined AB conditions as "triplets."

To be included for analysis, triplets were required to meet several criteria. First, the minimum number of included trials for each of the A, B, and AB components was set at 5 trials for each, in keeping with our prior work and assessments of the sensitivity of the analysis to the number of trails involved (Caruso et al., 2018; Mohl et al., 2020; Jun et al., 2022). The average numbers of trials of the included triplets across datasets are illustrated in **Supplementary Figure 1A**. The

two MT datasets involving adjacent stimuli had the lowest average trial counts (average of the A, B and AB trial counts was ~18 trials for the gratings and ~6 trials for the dot patches), whereas the MT “plaid” dataset had the most (average ~37 trials; all values computed after imposition of the full set of exclusion criteria, including those described below).

Second, spike count distributions for the A-alone and B-alone trials in a triplet were required to be sufficiently well-described by a Poisson distribution (with a variance roughly equal to the mean). The spike count distributions for A and B trials for triplets in all 5 datasets exhibited variance-to-mean ratios (Fano factors) with a mode of about 1, but there was some spread around that value (**Supplementary Figure 1B**). Accordingly, we discarded triplets for which the average variance-to-mean ratio in the A and B conditions was greater than 3. This value was chosen to exclude the worst outliers while preserving adequate data across all datasets to allow for appropriate comparisons. Screening with a more stringent criterion (variance-to-mean ratio < 2) reduced the overall quantity of data but did not change the key results we report here (**Supplementary Figure 2**). Notably, screening with the variance-to-mean ratio is a deviation from our previous approach involving an approximate chi-square goodness of fit test with Monte Carlo *p*-value calculation (Caruso et al., 2018; Jun et al., 2022). The former approach had the undesired effect of being more likely to exclude conditions with larger amounts of data, whereas the present method is neutral to dataset size.

Finally, spike count distributions for the individual A and B stimuli in a triplet were required to be sufficiently separable from one another. This is necessary so that it is possible to ascertain with some confidence how the AB spike count distributions compare to each of the corresponding single stimulus A and B distributions. We set the inclusion criterion as a 95% likelihood that the A and B spike counts were drawn from different Poisson distributions (specifically, the logarithm

of the intrinsic Bayes factor was required to be greater than 3, given an *a priori* 50-50 chance that the two distribution means are the same vs. different). The greater numbers of trials in some datasets than in others allowed for inclusion of triplets at smaller differences in the A vs. B response distributions (**Supplementary Figure 1C**), but this did not appear to impact the overall conclusions (see **Supplementary Figure 1D** and associated legend).

Data Analysis

A full description of our statistical method has been published previously (Caruso et al., 2018; Mohl et al., 2020; Jun et al., 2022). Briefly, we compare the distribution of spike counts in response to a stimulus pair (AB) to the distributions of spike counts in response to the component stimuli presented individually (A and B).

We consider four possibilities for the relationship between the AB response and the individual A and B responses (**Figure 2**). The first three attempt to fit the AB responses with a single Poisson distribution, as follows:

1. **Outside:** The mean of the distribution of AB responses might be higher than the maximum or lower than the minimum of A or B. This possibility includes either summation of the mean A and B responses or strong suppression of responses to one stimulus by the presence of the other.
2. **Single:** The mean of the distribution of AB responses might match the mean response to either A or B. This is akin to winner-take-all, with the same winner on every trial.
3. **Intermediate:** The AB responses might be well fit by a single Poisson distribution with a mean between A and B, as if the neuron performed an averaging operation on its inputs.

These possibilities are in contrast to the final possibility of interest:

4. **Mixture:** The AB responses match A on some trials and match B on other trials, creating a mixture of the two A and B distributions across trials (fluctuations or "multiplexing" from trial to trial). We model this possibility as a bimodal mixture of the A and B Poisson distributions as assessed on those single stimulus trials.

We evaluated these four hypotheses by computing their posterior probabilities based on intrinsic Bayes factors (Berger & Pericchi, 1996; Caruso et al., 2018). We used equal priors (25%) for each of the four possibilities. For the purposes of this work, we defined "fluctuating responses" as responses that were classified as "mixture" with a posterior probability greater than 0.67 (i.e., mixture is at least twice as likely as the other three categories combined).

As noted previously (Caruso et al., 2018; Jun et al., 2022), these category boundaries may not always perfectly capture the presence of fluctuating activity. For example, if a neuron sometimes switches between responding to A and B stimuli in the middle of a stimulus presentation, it will tend to be classified in the "intermediate" bucket. Similarly, if a neuron responds to A on 90% of the trials, it is likely to be classified as "single" even if it responds to B on the remaining 10%.

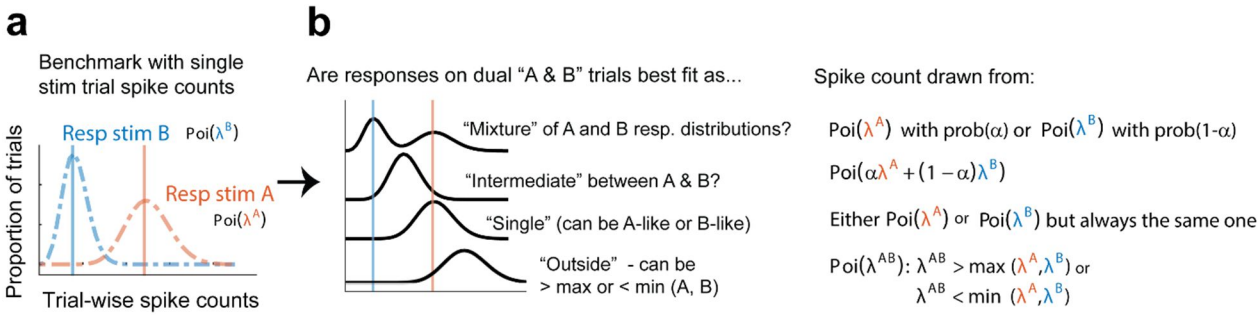


Figure 2: *Statistical comparisons to be made across datasets. We assign each trial into a "triplet" consisting of trials in which stimulus A (red), stimulus B (blue), or both stimulus A and B (AB, black) were presented. a) Within each triplet, spike count distributions are compared to determine how the AB distribution relates to the A and B distributions. b) The AB spike count distribution may be: a mixture of the A and B distributions, suggesting fluctuations between the responses to the two stimuli; an intermediate rate between the A and B distributions, suggesting averaging (or fluctuations on a faster time scale than the analysis period); a rate that matches either the single A or B distribution, suggesting a winner-take-all strategy, or a rate outside the range bounded by the A and B distributions, suggesting addition or subtraction of the responses. Spike counts are assumed to be drawn from Poisson distributions, with a weighting parameter alpha defining the relative contribution of the individual A and B distributions in "mixtures" and "intermediates."*

Results

Spiking activity fluctuates in MT when two stimuli are distinct, but not when they fuse into one object

We first evaluated the responses in visual area MT to stimuli that appeared as distinct objects. The first dataset involved adjacent drifting gratings, both of which were positioned within the receptive field of the MT neuron under study. These data were recorded simultaneously with recordings in V1, in which we have previously reported the presence of fluctuating activity in response to these stimuli (Jun et al., 2022). On the trials included for analysis, monkeys performed a task in which attention was directed elsewhere.

As in V1, we found that some units exhibited bimodal AB spike count distributions whose modes approximately match the individual A and B means, suggesting that the AB responses fluctuate between the A and B response rates across trials. An example of such a unit is shown in **Figure 3A**: the red and blue dashed lines indicate the means of the A-alone and B-alone response distributions, and the black curve provides a smoothed depiction of the distribution of spike counts across the combined AB trials. This distribution appears to reflect an admixture of the A-alone and B-alone response distributions, indicating that the unit fluctuates between each single response rate from trial to trial.

We next considered a second independent dataset from MT (Li et al., 2016) that shares similar properties to the adjacent gratings dataset. In this study, two separate patches of dots, each moving in different directions, were presented within the receptive field of the neurons under study. For the included trials, the monkey was performing a task in which attention was directed to a fixation point away from the dot patches. The combined-stimulus AB trials in this dataset were previously reported to show fluctuating activity patterns using a different analysis method (Li et al., 2016), and the dataset was made publicly available at (Li et al., 2017). As in the adjacent gratings MT dataset, we observed bimodal spike count distributions on combined AB trials with modes that appeared to match the mean responses observed on A-alone and B-alone trials (**Figure 3B**).

To evaluate the prevalence of such fluctuating activity patterns at the population level, we deployed the statistical comparison described in the Materials and Methods, in which spike count distributions on combined AB trials were evaluated in comparison to four possible options: “mixture,” “intermediate,” “single,” and “outside” (**Figure 2**). **Figure 3C** shows the results of this analysis for the adjacent gratings and dot patches datasets. In these two datasets, we only

identified “mixtures” and “intermediates,” but no “singles” or “outsides.” There was little difference between these two datasets (light blue vs. dark blue stacked bars in **Figure 3C**). Overall in these two datasets, “mixtures” constituted about 45% of the categorized conditions (**Figure 3E** bar labelled “distinct objects”).

In contrast, evidence of fluctuating activity was much lower in our third MT dataset, which involved gratings that were either presented individually (A-alone, B-alone) or superimposed to produce a fused percept of a “plaid” or checkerboard-like pattern. As in the first dataset, the stimuli were presented in the receptive field of the MT neuron under study, but this time they were presented at the same exact location, rather than adjacent to one another. The monkeys were performing a fixation task; no explicit attentional report was required. In this dataset, “mixture” responses were very rare, constituting only about 1.4% of the conditions that were well-categorized (**Figure 3D**, **E**). The majority of conditions were categorized as “outside” (~78%), with “single” and “intermediate” accounting for a further ~14% and ~6%, respectively.

The difference in the proportions of “mixture” response patterns in the first two MT datasets involving distinct objects (~45%) vs. this last dataset in which the two stimuli fuse to form a new third object (~1.4%) was statistically significant (**Figure 3E**, blue vs. green bars, chi-squared test $p<10^{-9}$). This finding is consistent with the interpretation that the prevalence of fluctuating activity is related to whether two stimuli are perceptually distinguishable vs. fuse to form a single object.

Spiking activity fluctuates in face patches MF and AL when two stimuli are distinct

To further explore the generality of these fluctuating activity patterns, we next turned to a higher order area of the visual pathway – the face patch system of inferotemporal (IT) cortex. Face patches are exquisitely sensitive to the identity of individual faces, and respond preferentially to

faces as compared to other objects. This affords the opportunity to test whether fluctuating activity is evoked solely when the two stimuli are both faces, as might be expected because the non-face objects are largely thought to be encoded by neural populations in other regions of IT, or if both face-face and face-object pairs evoke fluctuating activity. We previously identified fluctuating activity patterns in the MF face patch (Caruso et al., 2018), but we did not separate face-face and face-object stimulus conditions. Here, we revisited that dataset to conduct that comparison. We also sought to compare the results in MF to those in face patch AL. Recent work has suggested that MF is “earlier” in the face-patch hierarchy than AL, and that AL shows properties such as view-invariance that are less evident in MF (Moeller et al., 2008; Freiwald & Tsao, 2010; Ebihara, 2015). Thus, any differences between the results in MF and AL could shed light on how fluctuating activity varies within the processing hierarchy at this late stage of visual processing.

Accordingly, we assessed MF and AL face patch responses with both face-face and face-object stimulus combinations. In face-face combinations, one face evoked the strongest response from among a stimulus set (“preferred face”), and the other face evoked the weakest response from the same set. The preferred face was presented at the center of the neuron’s receptive field, while the non-preferred face was presented at a different location (**Figure 1C**). Similarly, the object stimuli were also chosen to evoke weak responses and were presented at a different location than the preferred face. The stimulus conditions were randomly interleaved, and the data were collected while monkeys performed simple fixation tasks.

Both face patch regions and both types of stimulus combinations elicited fluctuating activity. **Figure 4A-D** show the responses of four example neurons, each of which exhibited spike

counts that fluctuated across trials between appearing to respond to the preferred vs. non-preferred face (**Figure 4A, C**) or to the face vs. the object (**Figure 4B, D**).

To evaluate these results at the population level, we used the same statistical comparison described above for the MT analysis. As shown in **Figure 4E**, we identified “mixtures” in face patch MF for both face-face (dark brown bar) and face-object (mustard-colored bar) stimulus combinations. We also observed “intermediates” and “singles,” but no “outsides.” Overall, “mixtures” constituted about 32.5% of well-categorized conditions (**Figure 4G**, light brown bar). The proportion of “mixtures” was higher for face-face stimulus combinations (43.6%, **Figure 4G** dark brown bar) than for face-object stimulus combinations (22.7%, **Figure 4G** mustard-colored bar). The difference in these proportions was just barely significant by chi-squared test ($p=0.043$), but we note that when using stricter Fano factor inclusion criteria for the dataset (variance-to-mean ratio of single-stimulus responses less than 2 rather than 3, see *Materials and Methods: “Triplet” definition and inclusion criteria*), the reduction in the amount of included data increases this p value to $p=0.054$ (see **Supplementary Figure 2**). Accordingly, this aspect of the results should be viewed as of borderline significance.

“Mixtures” were even more prevalent in face patch AL (**Figure 4F**), constituting about 66% of the conditions for both face-face and face-object pairs (**Figure 4G**). The approximate doubling in the prevalence of “mixtures” in AL vs. MF (65.9 vs. 32.5%) was statistically significant by chi-squared test ($p=0.00002$). Like face patch MF, AL responses had no “outsides,” but “singles” and “intermediates” were still present. (In MF, we previously showed that such “intermediates” could show fluctuating activity within the timescale of individual trials (Caruso et al., 2018)).

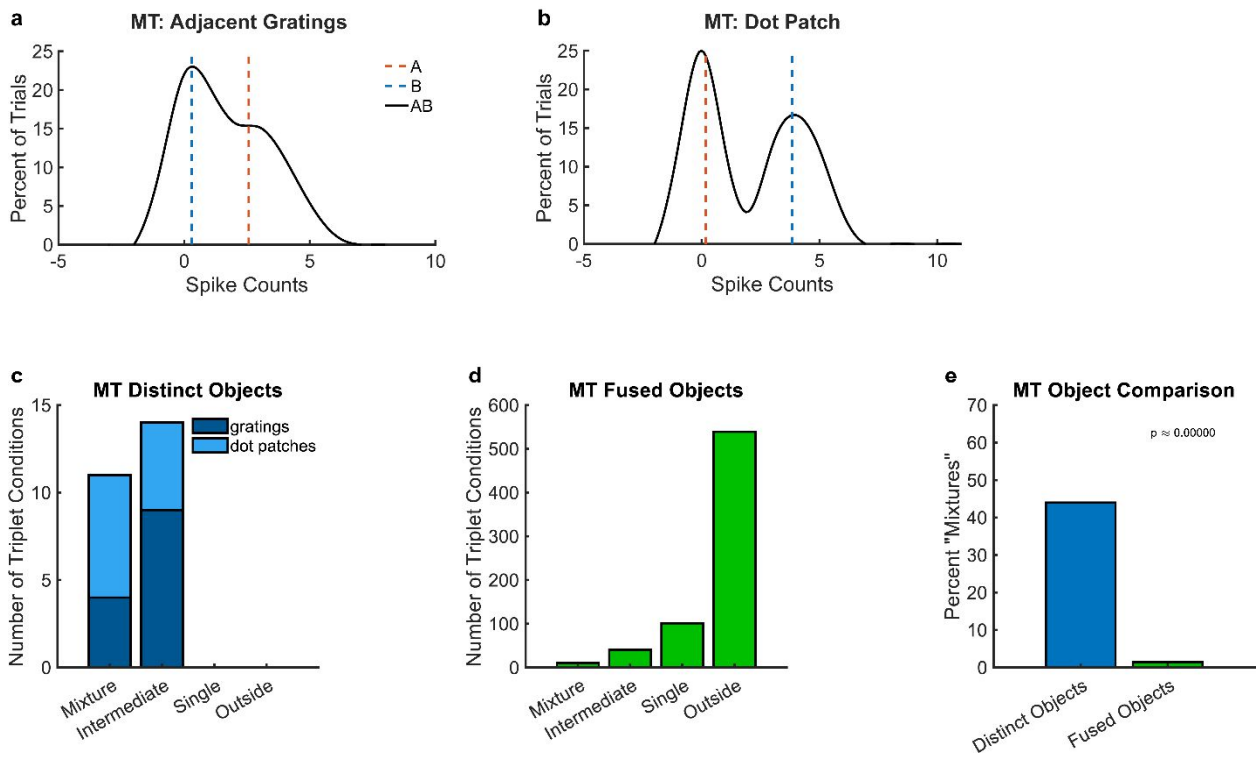


Figure 3: *Example individual unit responses and population results for area MT. a) An example unit's response to two simultaneous adjacent gratings (AB, black solid lines) in comparison to the mean response to each grating individually (A and B, red and blue dotted lines). b) A different example unit's response to simultaneous dot patches. In both (a) and (b), distributions were smoothed via convolution with a 3-point sliding window of which the middle point is weighted twice as much as either of the two outer points, followed by cubic spline interpolation. c) Number of triplets classified into each category (as shown in* ***Figure 2****) when responses to two simultaneous gratings (dark blue) or dot patches (light blue) are recorded in MT. Only triplets for which the winning model garnered a posterior probability of 0.67 or greater are shown. d) Same as (c) but for fused objects, created by overlaying two gratings to form a plaid. e) Comparison of (c) and (d) showing that fluctuating ("mixture") responses are only present when two objects are distinct, and not when two objects are fused into one. Results in this figure only include triplet conditions for which the average variance-to-mean ratio (Fano factor) of activity on the A-only and B-only trials was less than 3. See* ***Supplementary Figure 2*** *for corresponding findings using stricter inclusion criteria.*

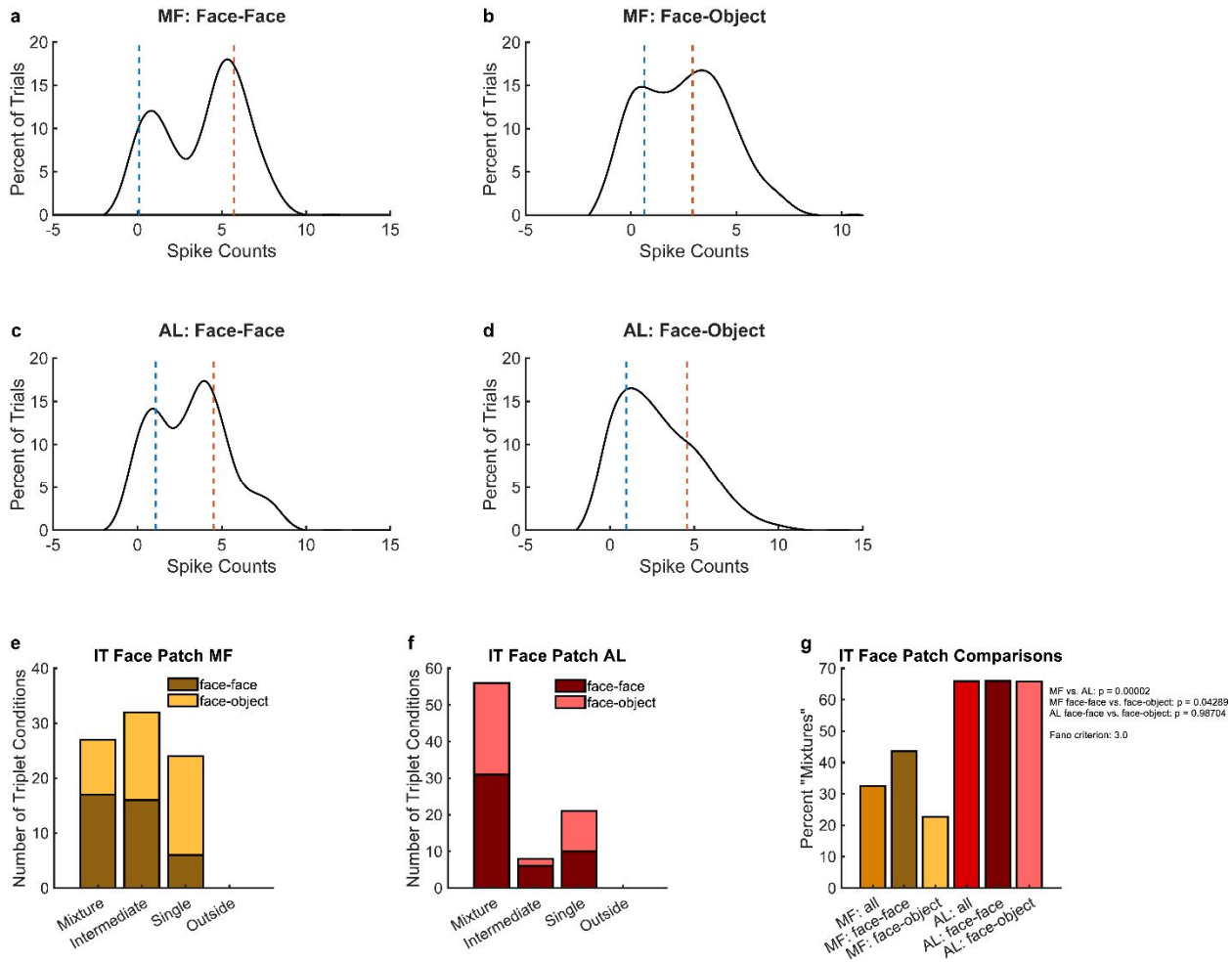


Figure 4: *Example individual unit responses and population results for IT face patches MF and AL. a) An example single MF unit's response to two simultaneous faces (AB, black solid lines) in comparison to the mean response to each face individually (A and B, red and blue dotted lines). b) Same as (a) but in response to one face and one object. c) Same as (a) but recorded in AL. d) Same as (b) but recorded in AL. In (a) through (d), distributions were smoothed via convolution with a 3-point sliding window of which the middle point is weighted twice as much as either of the two outer points, followed by cubic spline interpolation. e) Number of triplets classified into each category (as shown in Figure 2) when responses to a face-face pair (dark brown) or a face-object pair (mustard color) are recorded in face patch MF. Only triplets for which the winning model garnered a posterior probability of 0.67 or greater are shown. f) Same as (e) but recorded in face patch AL. g) Comparison of (e) and (f) showing that more conditions display a fluctuating ("mixture") pattern in AL than in MF. In MF, there were more "mixtures" among face-face pairs than among face-object pairs. Results in this figure only include triplet conditions for which the average variance-to-mean ratio (Fano factor) of activity on the A-only and B-only trials was less than 3. See* ***Supplementary Figure 2*** *for corresponding findings using stricter inclusion criteria.*

Discussion

A key problem in neuroscience is how the brain retains information about multiple co-occurring stimuli. We present a potential solution to this problem: When multiple stimuli are present, neurons across levels of visual processing may fluctuate, on the whole trial time scale, between responding to individual stimuli in distinct time windows. This fluctuating pattern may be a way for the brain to preserve information about multiple stimuli in complex sensory environments.

We report here that fluctuating activity was evoked by pairs of distinguishable stimuli in MT and areas MF and AL in the face patch system of IT cortex. This work extends our previous findings in several key ways. First, it expands the range of brain regions and stimulus conditions in which this phenomenon has been observed. In MT, fluctuating activity was observed in two independent datasets involving different stimulus conditions and collected by different labs. In the face patch system, fluctuating activity was observed in both face patches included in this study. This suggests that fluctuating activity evoked by combinations of stimuli is a robust phenomenon within the visual pathway and occurs across multiple stages of the visual hierarchy.

Second, another of our goals was to evaluate whether fluctuating activity is more prevalent when the two stimuli involved are competing for representation in the same neural population. The use of face-face and face-object stimulus combinations to probe signals in the face patch system provides an opportunity to address this question in a unique way. If the face patch system serves as a series of structures whose specific role is to represent faces and not non-face objects, one might expect that fluctuating activity would be more evident when the two stimuli were both faces than when only one of the stimuli is a face. Specifically, one might expect winner-take-all for the face stimulus, which would be classified as “single” in our

analysis. Instead, we observed fluctuating activity for both face-face and face-object pairs, although there was a modestly higher incidence of fluctuating activity for face-face stimuli than face-object stimuli in MF. In AL, fluctuating activity was very common for both types of stimulus combinations. This suggests that fluctuating activity can occur even when somewhat distinct populations of neurons are available to respond to the two stimuli in question; neurons outside the face patch system are in principle available to encode the presence and properties of the non-face object, yet fluctuating activity nevertheless occurred.

The presence of fluctuations evoked by face-object stimuli is reminiscent of a related aspect of our previous findings in V1 (Jun et al., 2022). Specifically, we observed fluctuating activity in V1 even though the two stimuli involved were usually not located within the same receptive field and in principle could have evoked activity in largely distinct neural subpopulations. One possible explanation for this is that the activity of the whole population is necessary to specify the presence or identity of a stimulus, even if some members of the population are silent. A second possibility is that fluctuations occur “on spec” regardless of the particular stimuli involved due to the possibility that they may require activity in overlapping populations of neurons to be encoded. Supporting these possibilities, we found fluctuating activity not only in the two MT datasets, which involved adjacent stimuli in the same neural receptive field, but also in the IT and earlier V1 datasets, which largely did not.

Finally, the evidence for fluctuating activity in MT when the two stimuli were perceptually distinct but not when they fused to form a common object lends support to the idea that there is an interplay between segregation of the scene into separate objects and the occurrence of fluctuating activity patterns. We previously found a similar pattern in V1 although not in V4 (Jun

et al., 2022). It is not yet clear what might account for the discrepancy with V4, and this matter bears further investigation.

There are both strengths and limitations to the use of multiple datasets across different laboratories, brain areas, and experimental conditions. The key strength is that the presence of fluctuating activity under a wide variety of circumstances indicates that this property of neural representations is a robust one that does not require specialized circumstances to uncover. Across this and our two previous studies (Caruso et al., 2018; Jun et al., 2022), we have now shown fluctuating activity occurring in the context of behavioral tasks requiring a report of both stimuli (IC, (Caruso et al., 2018)), strictly fixation (MF, AL, V4 - this study; (Caruso et al., 2018; Jun et al., 2022)), and attend-elsewhere tasks (MT, V1; this study and (Jun et al., 2022)). As mentioned above, fluctuating activity can be observed when two stimuli are within the same receptive field (MT) as well as when they are not (V1). A key limitation is that comparisons between datasets are likely to be affected by parameters such as the amount of data acquired. Future work in which conditions are held constant as much as possible across datasets will be needed before strong claims can be made about matters such as whether fluctuating activity is more vs. less prevalent across different behavioral tasks, brain areas, and even sensory systems.

Another area for future exploration is the extent to which fluctuating responses could allow for preservation of more than two stimuli. The current study exclusively used datasets in which two stimuli were presented, either alone or in pairs. It will be sensible to extend these analysis methods to more complex but realistic stimulus conditions involving three or more stimuli.

In sum, neural switching appears to be a general phenomenon throughout a range of areas in the visual system and could support the ability to preserve information about multiple sensory

components. Fluctuating activity may relate to object formation, a key perceptual ability whose underlying mechanism(s) remain poorly understood. Future experimentation is needed to understand how neural fluctuations may support such cognitive phenomena.

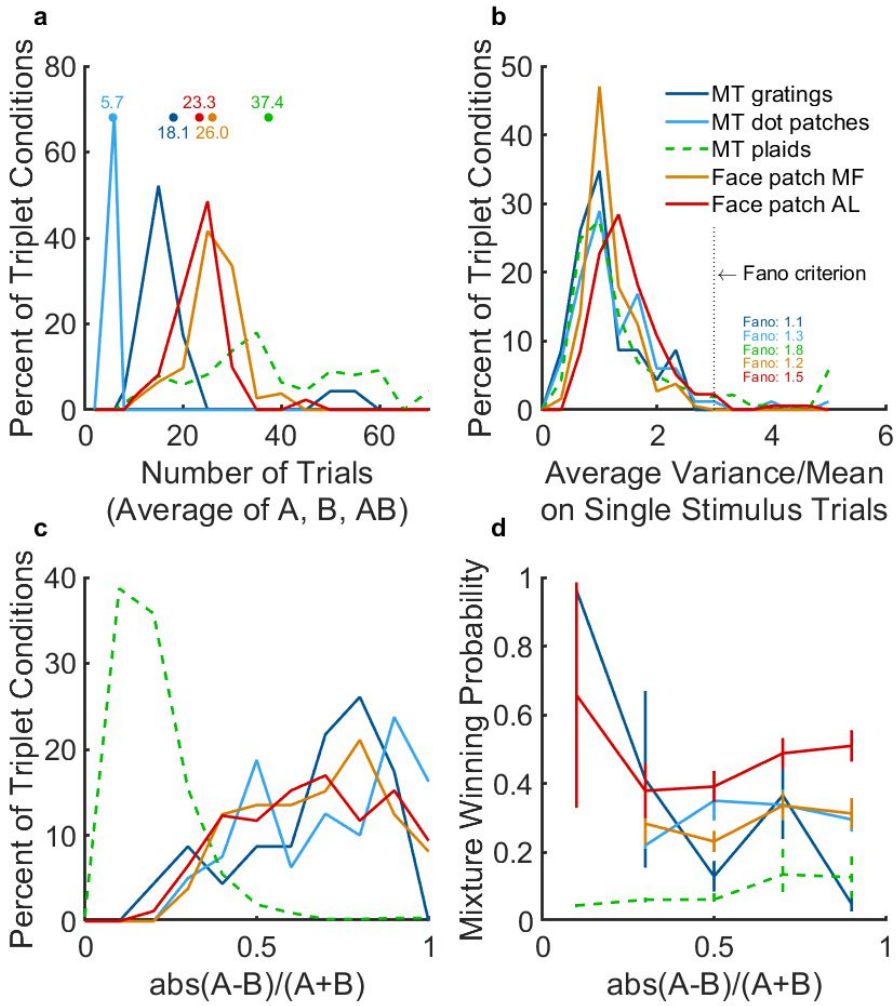
Acknowledgements

We thank Justine Shih, Tingan Zhu, Gelana Tostaeva, Justine Griego, and Sara Gannon for helpful discussions about these topics. This work was supported by: F31 fellowship to MNS; National Institutes of Health grant nos. R00EY020844 (MRC), R01EY022930 (MRC); Core Grant P30 EY008098s (MRC); R01DC013906 (JMG, STT); and R01DC016363 (JMG, STT); and by support to MRC from the McKnight, Whitehall, Sloan and Simons Foundations.

Competing Interests

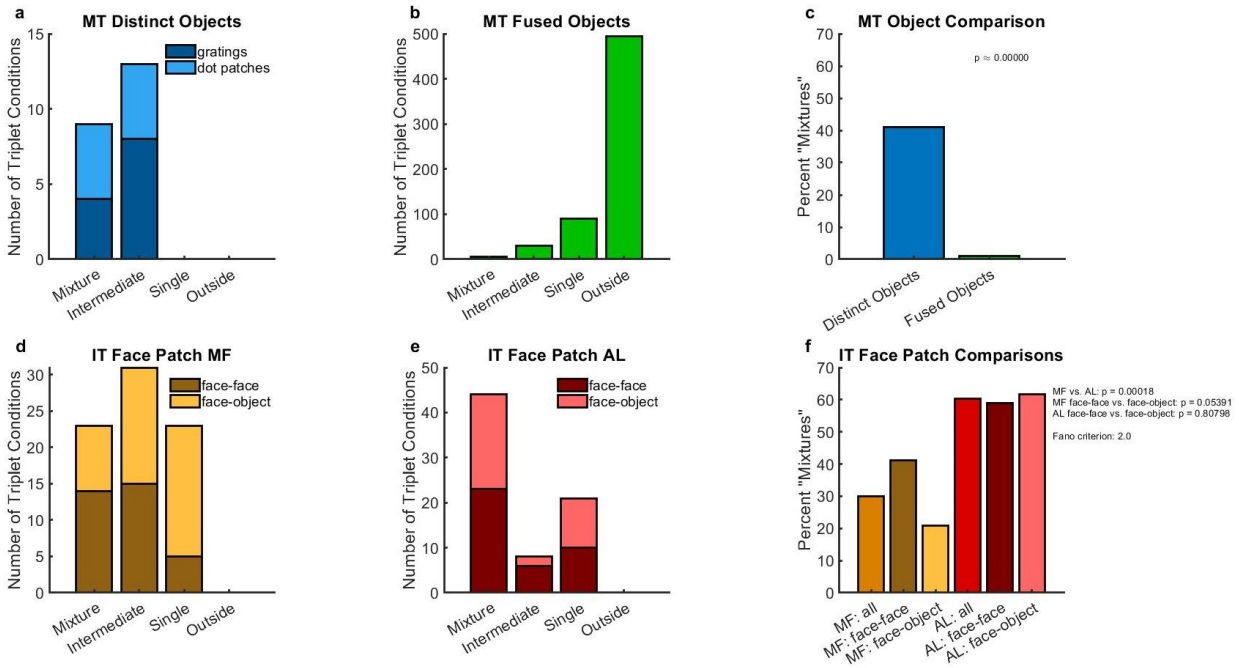
JMG is a reviewing editor at *eLife*.

Supplementary Figures



Supplementary Figure 1: *Properties of the datasets and application of exclusion criteria. a) Average A, B, and AB trial counts for each included triplet (after Poisson, trial count, and response separation screening). b) The distribution of variance-to-mean ratios, or Fano factors, on A alone and B alone trials across the different datasets. Results shown involve the average A and B Fano factors for a given triplet (computed after the screening for minimum trial count and A and B response separation). c) A and B response separations, computed as an index of the difference between A and B responses as a fraction of their sum, for included triplets (after Poisson, trial count, and response separation screening). Note that the MT plaid dataset differs from the other 4 datasets in that many of the included triplets have much smaller A and B response separations than those that are included for other datasets. This is likely due to the larger numbers of trials providing greater confidence that the A and B responses did indeed differ at smaller difference levels. d) Average "winning probability" for the "mixture" model (see Materials and Methods: Data Analysis) as a function of the response separation. Average "mixture" winning probability was higher for all 4 datasets involving distinguishable stimuli (MT*

adjacent gratings and dot patches, the face patch datasets) than for the MT plaid dataset, across all levels of A vs. B response separation.



Supplementary Figure 2: *Same as population results in* ***Figure 3*** *and* ***Figure 4****, but using a Fano factor criterion of 2 instead of 3. All trends in the data are maintained compared to the comparisons with a criterion of 3, other than the loss of significance between face-face and face-object comparisons in MF.*

References

Berger, J. O., & Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association*, *91*(433), 109–122. https://doi.org/10.1080/01621459.1996.10476668

Caruso, V. C., Mohl, J. T., Glynn, C., Lee, J., Willett, S. M., Zaman, A., Ebihara, A. F., Estrada, R., Freiwald, W. A., Tokdar, S. T., & Groh, J. M. (2018). Single neurons may encode simultaneous stimuli by switching between activity patterns. *Nature Communications*, *9*(1), 1–16. https://doi.org/10.1038/s41467-018-05121-8

Dow, B. M., Snyder, A. Z., Vautin, R. G., & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. *Experimental Brain Research*, *44*(2), 213–228. https://doi.org/10.1007/BF00237343

Ebihara, A. F. (2015). Normalization Among Heterogeneous Population Confers Stimulus Discriminability on the Macaque Face Patch Neurons. *Student Theses and Dissertations*. https://digitalcommons.rockefeller.edu/student_theses_and_dissertations/278

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035–1045. https://doi.org/10.1016/S0042-6989(03)00084-1

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889

Freiwald, W. A., & Tsao, D. Y. (2010). Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science*, *330*(6005), 845–851. https://doi.org/10.1126/science.1194908

Glynn, C., Tokdar, S. T., Zaman, A., Caruso, V. C., Mohl, J. T., Willett, S. M., & Groh, J. M. (2021). Analyzing second order stochasticity of neural spiking under stimuli-bundle exposure. *The Annals of Applied Statistics*, *15*(1), 41–63. https://doi.org/10.1214/20-AOAS1383

Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The Functional Neuroanatomy of Human Face Perception. *Annual Review of Vision Science*, *3*(1), 167–196. https://doi.org/10.1146/annurev-vision-102016-061214

Jun, N. Y., Ruff, D. A., Kramer, L. E., Bowes, B., Tokdar, S. T., Cohen, M. R., & Groh, J. M. (2022). Coordinated multiplexing of information about separate objects in visual cortex. *ELife*, *11*, e76452. https://doi.org/10.7554/eLife.76452

Keliris, G. A., Li, Q., Papanikolaou, A., Logothetis, N. K., & Smirnakis, S. M. (2019). Estimating average single-neuron visual receptive field sizes by fMRI. *Proceedings of the National Academy of Sciences*, *116*(13), 6425–6434. https://doi.org/10.1073/pnas.1809612116

Li, K., Kozyrev, V., Kyllingsbæk, S., Treue, S., Ditlevsen, S., & Bundesen, C. (2016). Neurons in Primate Visual Cortex Alternate between Responses to Multiple Stimuli in Their Receptive Field. *Frontiers in Computational Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fncom.2016.00141

Li, K., Kozyrev, V., Kyllingsbæk, S., Treue, S., Ditlevsen, S., & Bundesen, C. (2017). *Data from: Neurons in primate visual cortex alternate between responses to multiple stimuli in their receptive field* (Version 1, p. 24585489 bytes) [Data set]. Dryad. https://doi.org/10.5061/DRYAD.88PV1

Moeller, S., Freiwald, W. A., & Tsao, D. Y. (2008). Patches with Links: A Unified System for Processing Faces in the Macaque Temporal Lobe. *Science*, *320*(5881), 1355–1359. https://doi.org/10.1126/science.1157436

Mohl, J. T., Caruso, V. C., Tokdar, S. T., & Groh, J. M. (2020). Sensitivity and specificity of a Bayesian single trial analysis for time varying neural signals. *Neurons, Behavior, Data Analysis, and Theory*, *3*(1), 1–19. https://doi.org/10.48550/arXiv.2001.11582

Motter, B. C. (2009). Central V4 Receptive Fields Are Scaled by the V1 Cortical Magnification and Correspond to a Constant-Sized Sampling of the V1 Surface. *Journal of Neuroscience*, *29*(18), 5749–5757. https://doi.org/10.1523/JNEUROSCI.4496-08.2009

Ruff, D. A., Alberts, J. J., & Cohen, M. R. (2016). Relating normalization to neuronal populations across cortical areas. *Journal of Neurophysiology*, *116*(3), 1375–1386. https://doi.org/10.1152/jn.00017.2016

Ruff, D. A., & Cohen, M. R. (2016). Attention Increases Spike Count Correlations between Visual Cortical Areas. *Journal of Neuroscience*, *36*(28), 7523–7534. https://doi.org/10.1523/JNEUROSCI.0610-16.2016

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex. *Cerebral Cortex*, *11*(12), 1182–1190. https://doi.org/10.1093/cercor/11.12.1182