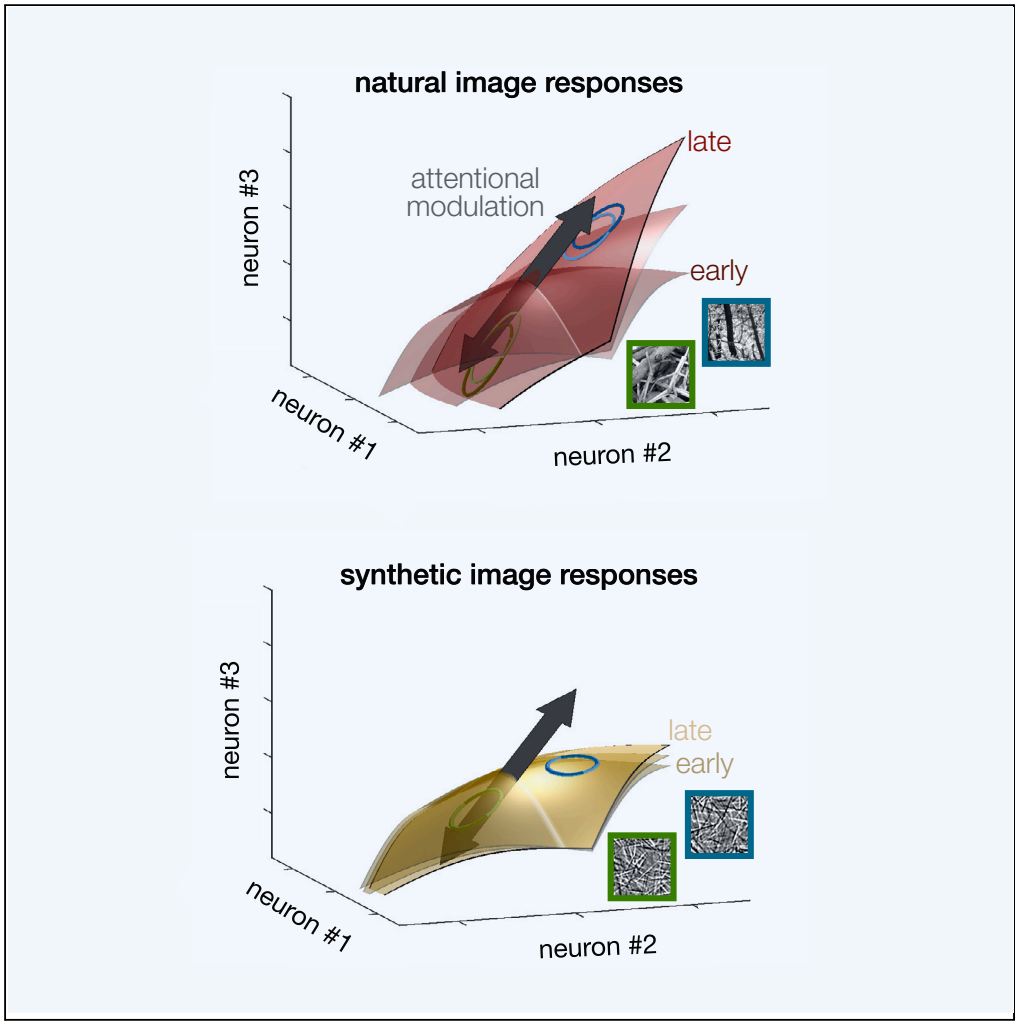
iScience

CellPress
OPEN ACCESS

Article

Paying attention to natural scenes in area V1



Andreea Lazar, Liane Klein, Johanna Klon-Lipok, Mihály Bányai, Gergő Orbán, Wolf Singer

andreea.lazar@esi-frankfurt.de (A.L.)
wolf.singer@esi-frankfurt.de (W.S.)

Highlights

Attention enhances the specificity of population responses to natural stimuli in area V1

Neural subspace modulated by attention overlaps with that representing natural stimuli

Overlap is strongest for late responses, suggesting a gradual alignment

Overlap is limited for synthetic stimuli lacking higher-order regularities

Lazar et al., iScience 27, 108816
February 16, 2024 © 2024
https://doi.org/10.1016/j.isci.2024.108816

iScience

CellPress
OPEN ACCESS

Article

Paying attention to natural scenes in area V1

Andreea Lazar,[1,2,*] Liane Klein,[1,2] Johanna Klon-Lipok,[1,2] Mihály Bányai,[3] Gergő Orbán,[3,4] and Wolf Singer[1,2,4,5,*]

SUMMARY

Natural scene responses in the primary visual cortex are modulated simultaneously by attention and by contextual signals about scene statistics stored across the connectivity of the visual processing hierarchy. We hypothesized that attentional and contextual signals interact in V1 in a manner that primarily benefits the representation of natural stimuli, rich in high-order statistical structure. Recording from two macaques engaged in a spatial attention task, we found that attention enhanced the decodability of stimulus identity from population responses evoked by natural scenes, but not by synthetic stimuli lacking higher-order statistical regularities. Population analysis revealed that neuronal responses converged to a low-dimensional subspace only for natural stimuli. Critically, we determined that the attentional enhancement in stimulus decodability was captured by the natural-scene subspace, indicating an alignment between the attentional and natural stimulus variance. These results suggest that attentional and contextual signals interact in V1 in a manner optimized for natural vision.

INTRODUCTION

Neuronal circuits across the visual hierarchy make efficient use of limited resources to encode complex natural scenes by exploiting their structural regularities.[1,2] Growing evidence suggests that during perceptual inference the visual system employs a hierarchical internal model of the visual environment that integrates current sensory evidence with previously acquired knowledge of natural scene statistics.[3,4]

When visual attention is directed toward a specific spatial location, it is thought to facilitate the perception of the targeted sensory input by prioritizing its processing. The signatures of this process can already be observed in primary visual cortex (V1). In this area, spatial and object-centered attention modulate the firing rates of neurons responding to the selected stimuli.[5–9] Responses of V1 neurons are also shaped by regularities in visual stimuli extending beyond the receptive field of the neuron. These regularities reflect either complex visual structures[10,11] or in more general, the statistics of natural images[4,12–15] processed in ventral areas downstream of V1.[16] These modulations of V1 population activity by higher-order regularities can be summarized as contextual signals. Contextual signals that convey information stored in the hierarchically organized internal model of the world are essential for the parsing of visual scenes.[17,18] Thus, attentional mechanisms must interact with these other sources of top-down influences, raising the key question: how do these two processes — one that reflects the current allocation of attention and the other conveying information about higher-order regularities in natural environments — cooperate in modulating responses in V1.

Here we performed parallel recordings of neuronal responses in the primary visual cortex of two macaque monkeys engaged in an attention task. In this paradigm, attention was directed to one of two identical images by a spatial cue that was delayed from the stimulus onset. By measuring the multiunit activity of a population of neurons with receptive fields overlapping with one of the images, we tracked the contribution of the top-down effect of attention on the distributed stimulus representation. We found convincing enhancements in natural scene encoding with attention and hypothesized that these enhancements rely on refinements of the responses to high-level features present in natural images, such as regularities in spatial, contrast, and frequency structures or texture properties. To test this hypothesis, we independently manipulated a second form of top-down modulation, arising from the statistical structure of the input images. We constructed synthetic stimuli that lacked the higher-level features characteristic of natural scenes and found that, for these stimuli, the attentional benefits in stimulus encoding vanished; however, they could be recovered when the synthetic stimuli were modified to contain structured contours. To understand how attentional modulations might interact with the representation of stimuli, we investigated the geometry of population responses using a combination of dimension reduction methods and decoding analysis. Our analysis revealed that the evoked neuronal population responses to natural scenes converged to a compact low-dimensional representation, which coincided with the subspace where the dominant attentional signal could be identified.

[1]Ernst Strüngmann Institute, Frankfurt am Main, Germany
[2]Max-Planck Institute for Neuroscience, Frankfurt am Main, Germany
[3]HUN-REN Wigner Research Center for Physics, Budapest, Hungary
[4]These authors contributed equally
[5]Lead contact
*Correspondence: andreea.lazar@esi-frankfurt.de (A.L.), wolf.singer@esi-frankfurt.de (W.S.)
https://doi.org/10.1016/j.isci.2024.108816

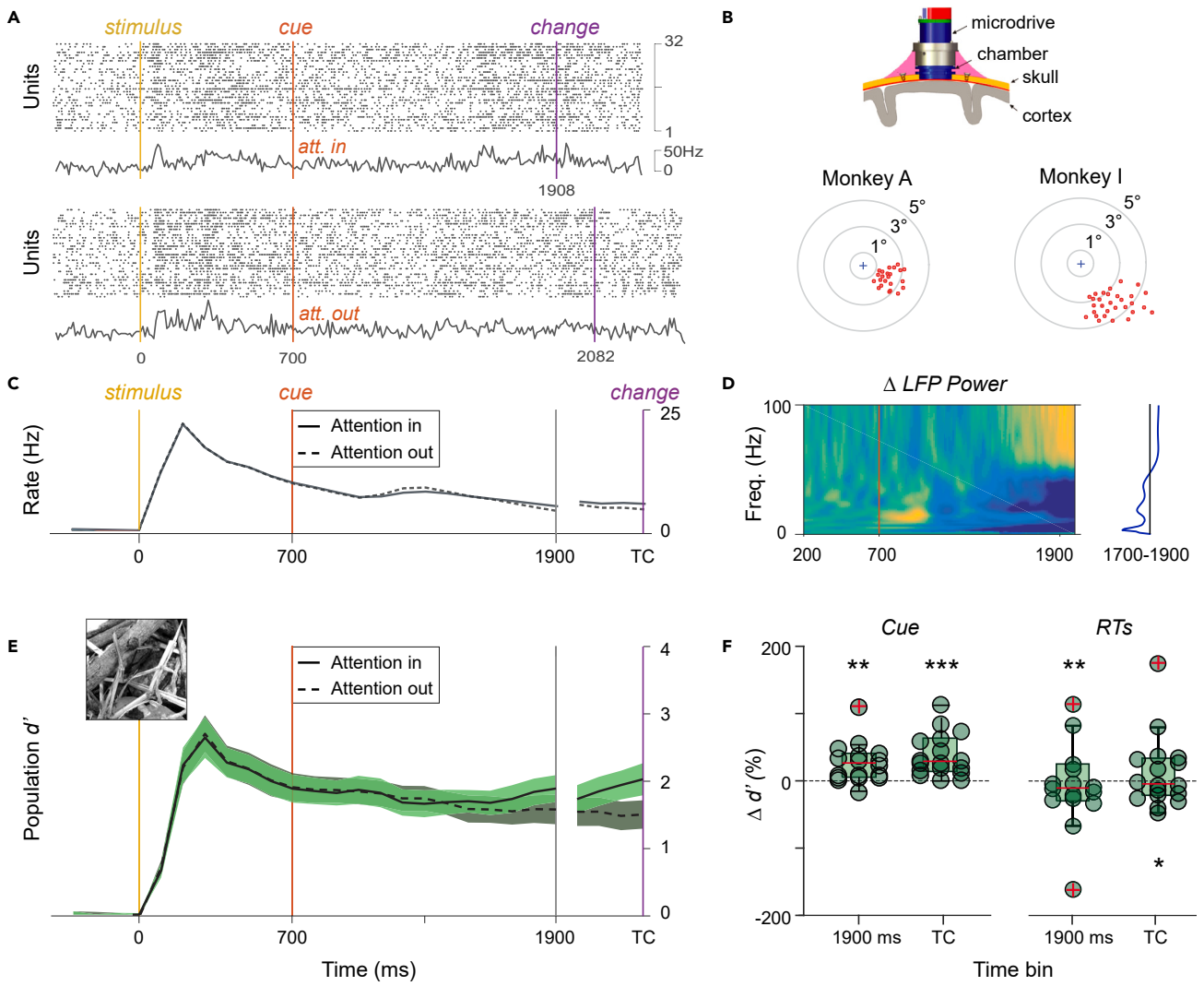


Figure 1. Neuronal population recordings from primary visual cortex during an attention task with natural scene stimuli

(A) Multi-unit responses to the same stimulus in two example trials, corresponding to the attention in and attention out conditions. Colored lines represent timing of stimulus onset (yellow), attention cue (red) and stimulus change (purple).

(B) Top: Chronic implantation device with movable electrodes (GrayMatter probe). Bottom: Receptive field locations of multiunits in the two monkeys shown in red. Cross indicates the fixation point.

(C) Grand average firing rates in the attention-in and attention-out conditions (baseline substracted, 18 sessions). Data aligned on stimulus onset or time of change (TC). Colored lines as in A.

(D) Spectrogram of the change of LFP power between the attention-in and attention-out conditions (baseline substracted). A gradual shift in power from low to high frequencies starts ≈ 300 ms after the attentional cue (marked by vertical line).

(E) Stimulus discriminability along the trial, quantified by population d' (200 ms sliding windows, baseline substracted, 18 sessions). Data aligned on stimulus onset or time of change (TC). Shading indicates standard error of the mean.

(F) Effects of attention on stimulus discriminability shown across sessions (reported in % change). Left panel: trials are separated based on cue information. Right panel: trials from the attention-in condition are sorted based on reaction time (RT) and separated in two-halves. y axis shows percent increase in d' for *in* vs. *out* trials (Cue) or *fast* vs. *slow* trials (RTs). (* p-val < 0.05, ** p-val < 0.01, *** p-val < 0.001).

RESULTS

Attention-induced enhancement of natural scene representation

We obtained parallel multisite recordings of multi-unit activity (MUA) (example trials in Figure 1A) and local field potentials (LFP) (32 channels, chronically implanted microdrive, Figure 1B) from area V1 of two awake behaving macaques (*Macaca mulatta*), while they performed a spatial attention task. Trials were initiated by a lever press while the monkey maintained fixation. After 500 ms two identical stimuli were presented at symmetrical locations on either side of the vertical meridian (distance from fixation spot 2.3–3.2° of the visual angle), one of

which covered the receptive fields (RFs) of the recorded units (example of RF centers in the two monkeys in Figure 1B). The fixation spot changed color, 700 ms after stimulus onset, cueing the monkey toward the "target" stimulus. Monkeys were rewarded if they responded to the rotation of the target stimulus, which occurred 1900–2400 ms after stimulus onset. If they responded to the rotation of the non-cued stimulus, the distractor, reward was withheld and the trial was aborted. We performed analysis in 200 ms long sliding windows (100 ms sliding resolution) for the response interval from stimulus onset to 1900 ms (stimulus-aligned data) and for the 500 ms interval preceding the stimulus change (change-aligned data, time of change (TC)), thus excluding any activity evoked by the stimulus rotation or the motor response (further details in STAR Methods).

We found that attention increased the average firing rates only modestly in V1, and the changes were significant only for the change-aligned and not the stimulus aligned data (Figures 1C and S1, Wilcoxon signed-rank test; [1700-1900 ms], p = 0.1330 n.s.; [TC-200, TC], p = 0.0279; n = 18 sessions in 2 monkeys). In spite of the small firing rate changes, we observed a strong shift in LFP power from low to high frequency oscillations with attention, which started ~ 300 ms after the cue and increased gradually with the expectancy of the stimulus change (Figure 1D, individual animals in S4). Across trials, reaction times in the attention-in condition were positively correlated with the LFP power in the theta and beta bands and negatively correlated with the LFP power in the gamma band (Spearman correlation; theta $r = 0.18$, $p = 1e-11$; beta $r = 0.08$, $p = 0.002$; gamma $r = -0.12$, $p = 1e-05$), suggesting a strong task-related effect.

Our initial goal was to examine the impact of attention on the stimulus specificity of the sustained population responses to natural scenes. Few electrophysiology studies have focused on the representation of natural stimuli in primary visual cortex,[15,19,20] and how attention affects the decodability of natural scenes in V1 is still unclear. We quantified the stimulus-specificity of V1 neuronal population responses by measuring the differences in spiking patterns evoked by different stimuli compared to their variability across trials (population d', STAR Methods). We found a significant increase in population d' with attention, i.e., responses evoked by different stimuli were more differentiable when the stimuli had been cued (attention-in) than when they had not (attention-out), particularly in the time windows preceding the stimulus change (Figures 1C and 1D, Wilcoxon signed-rank test; [1700-1900 ms], p = 0.0012; [TC-200, TC], p = 0.00023; n = 18 sessions in 2 monkeys). These attentional benefits in natural scene discriminability were significant in individual animals and occurred both in the presence (monkey A) and absence (monkey I) of average firing rate changes (Figure S2). Importantly, a significant difference in stimulus discriminability was not only elicited by the cue, but also by intrinsic attention-related factors: when the trials from the *attention-in* condition were sorted based on the monkeys' reaction times (RTs), we found that trials with faster RTs had higher d' values compared to those with slower RTs (Figure 1F, Wilcoxon signed-rank test; [1700-1900 ms], p = 0.0043; [TC-200, TC], p = 0.0123; n = 18 sessions in 2 monkeys). The relationship between d' and RTs points to a graded improvement in stimulus decodability proportionate to the intensity of the attentional allocation, suggesting a link between the modulatory signals identified in V1 and behavior.

Differential effects of attention for natural images and synthetic controls

The response properties of neurons at higher levels of the processing hierarchy are known to capture the higher-order regularities of natural scenes.[16] Indeed, when gauging V2 activity by using natural texture images, specific removal of high-level statistics was shown to result in dampened responses of V2 neurons.[21] Given the reciprocally connected structure of the visual cortex, information extracted and encoded at higher levels of the processing hierarchy can constrain activity at lower levels of processing through feedback. Consequently, such contextual feedback modulation of V1 responses is likely more effective when natural high-level features are present in stimuli.[4] We hypothesized that if contextual modulation and attention share top-down pathways, then the presence of high-level features constitutes a prerequisite for attentional enhancements in stimulus encoding across neuronal populations in V1.

To test this hypothesis, we matched the natural scene stimuli in every recording session with an equal number of synthetic control images. The synthetic images were constructed in two ways. First, by filter-scrambling, to remove spatial correlations between low-level features (Figure 2A, details in STAR Methods), which was previously shown to reduce top-down influences on V1 responses 4. Second, by phase-scrambling, to remove high-level regularities from images while leaving the spectral content intact (Figure 2B), which was previously shown to reduce both the intensity and specificity of responses of V2 neurons.[21] By construction, the synthetic controls contained no high-level features, but retained basic image properties: either the low-level structure preferred by V1 cells or the spectral content of the original natural scenes.

The synthetic controls matched the natural scenes in luminance contrast and evoked, on average, similar firing rate responses across the recorded neuronal populations in V1 (Figure S3). Synthetic stimuli were characterized by a task-related shift in LFP power from low to high frequencies with attention, of similar magnitude to that observed for natural stimuli (Figure S4). LFP power in the gamma band was higher for natural scenes compared to synthetic images (Figure S4), supporting previous suggestions that visual stimuli with higher structural predictability result in stronger gamma oscillations.[20] Attentional modulation of LFP power in the gamma band was not consistent across animals and was therefore difficult to interpret (Figure S4). However, the shift in LFP power from slow to high frequencies with attention was strong, consistent across animals, and characterized the responses to both natural and synthetic stimuli, indicating attentional effects in V1 irrespective of the stimulus type (Figure S4). Similar to the natural image condition, reaction times across trials in the attention-in condition for synthetic stimuli were positively correlated with the LFP power in the theta and beta bands and negatively correlated with the LFP power in the gamma band (Spearman correlation; theta $r = 0.11$, $p = 3e-5$; beta $r = 0.07$, $p = 0.008$; gamma $r = -0.11$, $p = 3e-04$) and were indistinguishable from reaction times to natural stimuli (Wilcoxon signed-rank test; p = 0.3, n.s.; n = 18 sessions in 2 monkeys), suggesting a similar level of engagement. We found no differences in the frequency of saccadic eye movements between natural scenes and synthetic stimuli, or with attentional allocation (Figure S5). The task performance of the animals was also not significantly affected by the stimulus type (Wilcoxon signed-rank test; p = 0.42, n.s.; n = 14 sessions in 2 monkeys).

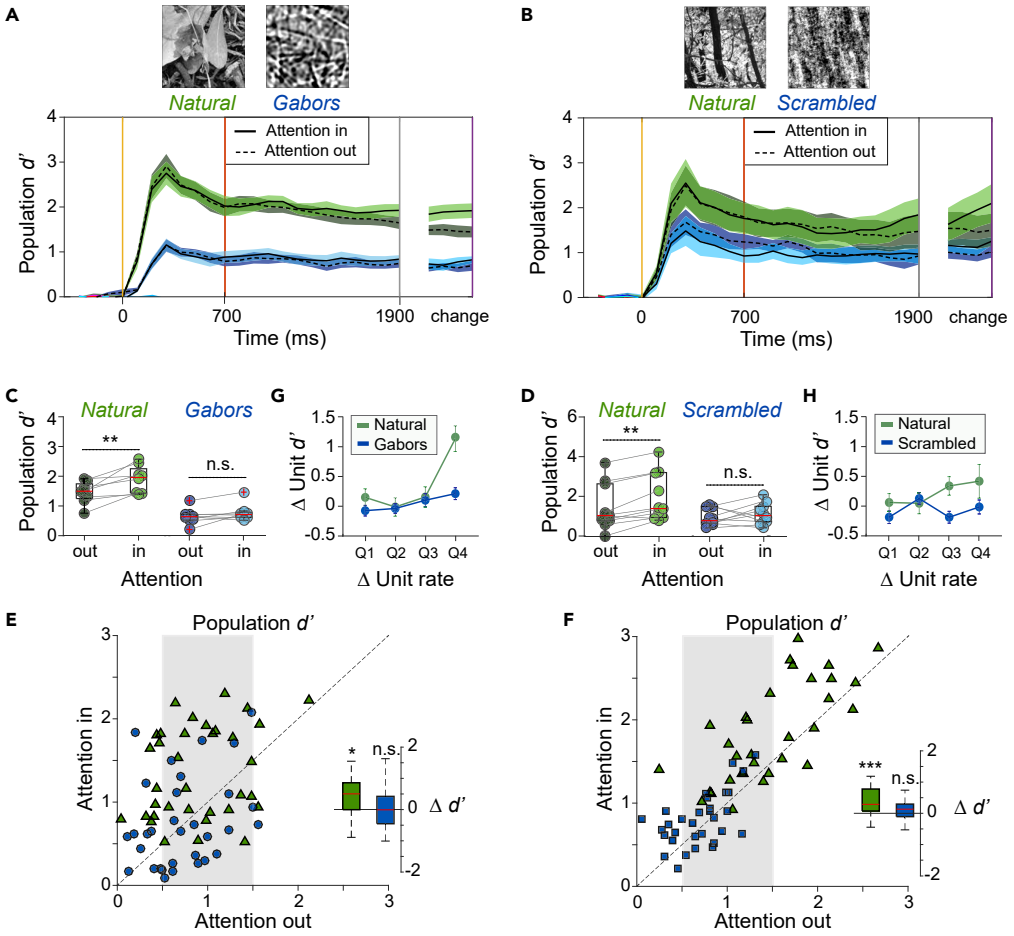


Figure 2. Attentional modulation depends on stimulus content

(A) Stimulus discriminability *d′* for natural scenes (green) and synthetic images composed of independent Gabor filters (blue), shown along the trial (200 ms sliding windows, baseline substracted). Shading indicates standard error of the mean.

(B) Stimulus discriminability *d′* for natural scenes (green) and phase-scrambled images (blue).

(C and D) Attentional effects on *d′* for natural (green) and synthetic images (blue) in the time window preceding the stimulus change (circles depict individual sessions).

(E and F) Scatterplot of *d′*-values across all natural and synthetic stimulus pairs, with and without attention. Smaller inset: attentional modulation of *d′*-values for the stimulus pairs marked by the shaded area, selected to increase overlap between stimulus types in the attention-out condition (* p-val < 0.05, *** p-val < 0.001, n.s. p-val > 0.05).

(G and H) Change in *d′* as a function of change in firing rate with attention. Analysis performed across all recorded units. Responses were sorted by amplitude of rate change and grouped into quartiles (marked on *x*-axis). Large increases in firing rate with attention result in larger positive changes in *d′* for natural (green), compared to synthetic stimuli (blue). Error bars indicate the standard error of the mean.

In agreement with our hypothesis, we found that the discriminability of natural stimuli, but not that of synthetic stimuli, was enhanced by attention (Figures 2A–2D, Wilcoxon signed-rank test; [TC-200, TC]; natural images p = 0.0078, Gabors p = 0.25 n.s.; n = 8 sessions in 2 monkeys; natural images p = 0.0039, scrambled p = 0.375 n.s.; n = 10 sessions in 2 monkeys). In spite of the fact that the overall response amplitudes to natural and synthetic stimuli were very similar (Figure S3), the spike-count vectors evoked by synthetic stimuli were less discriminable than those evoked by natural scenes (compare green and blue in Figures 2A and 2B). To limit these differences, we pulled together all pairs of natural and synthetic stimuli from all recording sessions and selected a subset of natural and synthetic stimulus pairs that were at similar levels (0.5–1.5 *d′* units above the baseline) in the attention-out condition (Figures 2E and 2F, shaded areas). Also in this case, attention increased the discriminability of neuronal responses to natural stimuli, but not to synthetic stimuli (insets in Figure 2E Wilcoxon rank-sum test; p = 0.03, natural pairs, p = 0.35 n.s., synthetic pairs; and F, p = 0.0001, natural pairs, p = 0.87 n.s., synthetic pairs). Finally, when responses of all recorded units were sorted by amplitude of rate change with attention and grouped into quartiles, we found that equally strong increases in firing rates resulted in *d′* enhancements for natural but not for synthetic stimuli (Figure 2E, 256 units; and F, 315 units), suggesting a dissociation between the strength of the attentional modulation of firing responses and the resulting gains in stimulus specificity (Figure S3).

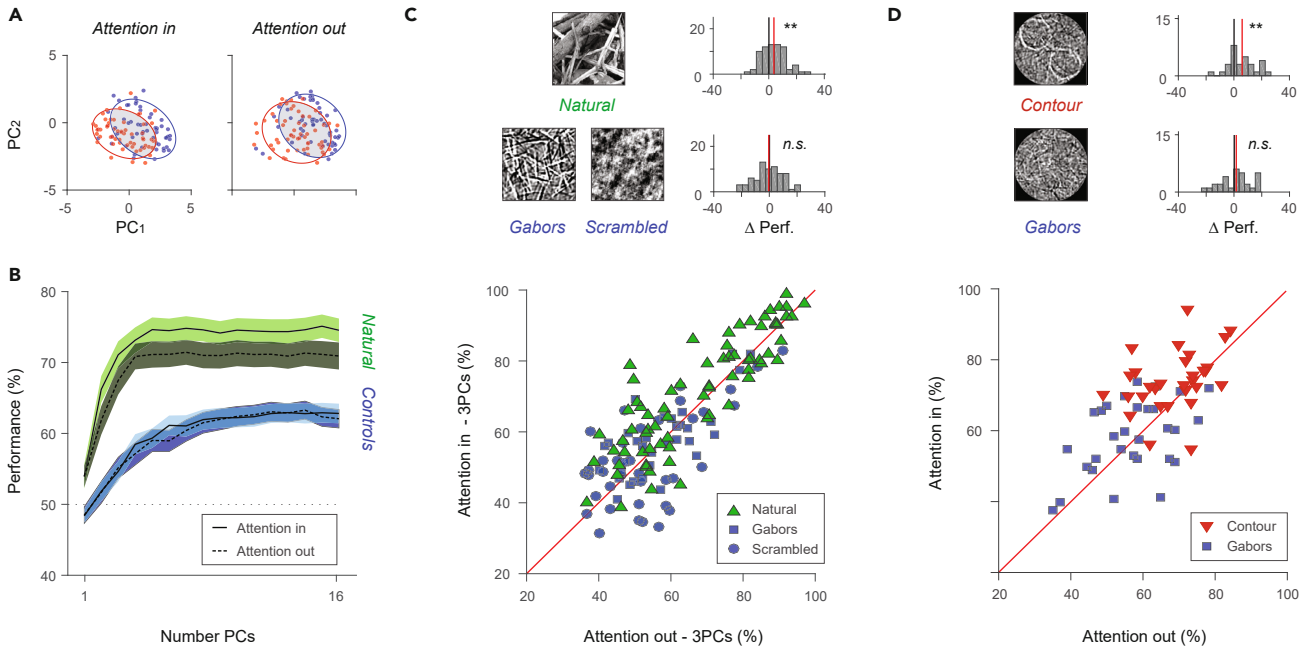


Figure 3. Effects of attention on stimulus encoding in principle component space

(A) Example session depicting population responses to two natural scene stimuli (red and blue) for the two attentional conditions in the space described by the first 2 principal components (200 ms spike-count vectors, 1700–1900 ms from stimulus onset; each point represents a trial).

(B) Stimulus decoding performance in principle component space for natural scenes (green) was higher than for synthetic stimuli (blue) and was modulated by attention (based on 1700–1900 ms spike count vectors; PCA was performed on pre-cue activity 500–700 ms window, separately for natural and synthetic stimuli). The shaded areas indicate the standard error of the mean. Attentional differences in stimulus decoding are apparent from a low number of PCs, suggesting an alignment between attentional and stimulus variance.

(C) Impact of attention on performance scores in low-dimensional projections (first 3 PCs) depends on stimulus type. Scatterplot shows performance scores for natural scenes (green) and synthetic controls (Gabors, blue squares; scrambled, blue circles; markers represent stimulus pairs n = 71; 18 recording sessions, 2 monkeys). Differences in performance with attention are significant for natural stimuli (top histogram, ** p-val < 0.01) but not controls (bottom histogram, n.s. p-val > 0.05).

(D) Contour stimuli are compared to synthetic controls (n = 30 stimulus pairs from 5 recording sessions in 1 monkey). Differences in performance with attention are significant for contour stimuli (p-val < 0.01), not for controls (p-val > 0.05).

Overlapping representational spaces for stimulus and attention

These findings raise the question of why responses to natural scenes profit more from attentional refinement than responses to manipulated stimuli, or, in other words, what is special about natural scenes? Previous research has shown that natural images are statistically redundant since light-intensities at neighboring locations are likely to be correlated and consequently, they can be efficiently compressed.[22,23] Structured compressible visual stimuli are well captured by neuronal population dynamics in low-dimensional manifolds,[20,24–26] but see.[27] How could such low-dimensional collective representations of natural images in V1 be further optimized by the allocation of top-down attention?

To answer this question, we contrasted the low-dimensional subspace accommodating the presented images with the space affected by the attentional signal. Given that, in the current task, the stimuli precede the attentional cue, we could directly inquire whether the variance added by attention was orthogonal to or belonged to the same dimensions as the variance produced by the stimulus. Specifically, in each recording session, we projected the activity from the time window preceding the stimulus change (1700–1900 ms) into the principal component space defined by activity recorded after the stimulus onset but before the presentation of the attentional cue (500–700 ms). We reasoned that if the attentional variance was largely aligned to the stimulus variance, the attentional differences in stimulus decodability would become apparent in the low-dimensional space described by the first components. This is indeed what we found (examples in Figures 3A and S6).

We quantified the attentional differences in natural-scene representations in principle component space by applying a decoding technique. Specifically, a cross-validated Bayesian decoder was trained to predict stimulus identity based on data projections (spike-count vectors over the 1700–1900 ms interval) into the pre-cue PCA space described by the first k principal components, and test performance was estimated in this same space based on unseen trials (Figure 3B, 5-fold validation, details in STAR Methods). We found that a small number of components captured the majority of variance produced by natural stimuli before the onset of the cue, allowing natural scenes to be well distinguished in principle component space (Figure 3B, green lines denote decoding performance, range 61.8 – 75.13% for $k \geq 2$, pre-cue PCA space constructed based on responses to natural scenes). Importantly, the same components captured a large fraction of the attentional effects, as reflected by the significant modulation of population responses in low-dimensional subspaces (Figure 3B, Wilcoxon signed-rank

test; $p < 0.015$ for $k \geq 2$; Holm-Bonferroni correction showed significance for all $k \geq 2$; n = 71 stimulus pairs, 18 sessions, 2 monkeys; scatterplot and histogram of attentional effects for $k = 3$ are shown in Figure 3C). In comparison, the synthetic images performed more poorly (Figure 3B, blue lines denote decoding performance, range 51.7 – 63.3% for $k \geq 2$, pre-cue PCA space constructed based on responses to synthetic stimuli) and showed no attentional modulation (Figures 3B and 3C, Wilcoxon signed-rank test $p > 0.05$ for all k; n = 71 stimulus pairs, 18 sessions, 2 monkeys), in spite of residing in principle component spaces with similar levels of overall variance (Figure S7B).

In a final set of experiments, we generated synthetic images that combined Gabor functions into simple contour-like patterns, thus introducing the kind of higher-level structure expected to elicit differential responses at higher processing stages. In these additional datasets, both main effects described previously were reproduced: the structured contour stimuli were well distinguished in principle component space, while the unstructured controls were not (decoding performance range contour stimuli 66.4 – 74.7% and synthetic Gabors 54.9 – 60.4% for $k \geq 2$, Figure S7) and the attentional effects were specific to the contour stimuli and captured already by a low number of components (contour stimuli Wilcoxon signed-rank test; $p < 0.05$ for all $k \geq 2$; n = 30 stimulus pairs, in 1 monkey; control images $p > 0.05$ for all k, Figure S7; scatterplot and histogram of attentional effects for $k = 3$ Figure 3C). Interestingly, trial-shuffling within stimulus condition reduced the attentional differences in decoding performance for the contour stimuli and the original natural scenes, suggesting that in these low-dimensional projections, decoders benefited from the intact correlation structure present in the data (Figure S7). Overall, by constructing synthetic images with controlled statistical structure, we confirmed that the attentional benefits in stimulus encoding across collective neuronal responses in V1 were specific to images containing higher level structural regularities. Such images are more likely to engage structured feedback from higher levels of processing.

Critically, we found that the dimensions that captured the majority of variance produced by natural stimuli also captured a large fraction of the attentional effects. To control for the specificity of this alignment, we constructed three alternative projections of the same spike-count vectors preceding stimulus change (1700-1900 ms) and assessed how these spaces captured both the natural stimulus and the attention signal. We considered a random orthogonal basis (Figure 4A, yellow), a PC space constructed based on pre-stimulus spontaneous activity (Figure 4A, orange), and a PC space based on early evoked activity (Figure 4A, purple). The minimum number of principal components required to reach 90 % of peak performance accuracy for the attention in condition was significantly higher for the three projections compared to the original pre-cue projection (Figure 4B, compare yellow, orange, purple to green; Wilcoxon signed-rank test, $p < 0.001$). A direct examination of the first principal components describing the natural scene space revealed highly distributed representations across units, making a simple selectivity-based explanation unlikely (example Figure 4C). In addition, we found that all three alternative projections yielded weaker attentional differences in decoding performance for both the first three and five components (Figure 4D; Wilcoxon signed-rank test, $p < 0.01$). Most interesting was the difference between the projections constructed from the early and late evoked responses, which were compared directly in Figure 4E. Session-by-session comparison of attention-in and attention-out decodability in these two projection spaces, revealed a dual effect: 1, Decoding of stimulus was more efficient in the basis constructed based on late responses (green) than that based on early responses (purple), irrespective of the attentional state (Wilcoxon signed-rank test, $p < 0.001$ attention-in; $p < 0.001$ attention-out); 2, Attention made a significant contribution to decoding in the basis constructed using late responses (Figure 4E, green histogram; Wilcoxon signed-rank test, $p = 0.000014$). Note that neither of these two projections had information about which stimulus to attend, since both are pre-cue; therefore, it is striking that a precise alignment developed later in the trial. This result indicates that during stimulus presentation, the population response is transformed into the space in which attention can be effectively deployed.

DISCUSSION

In this study, we found that attention improves the representation of natural scenes across neuronal population vectors in area V1. By constructing synthetic stimuli with controlled statistical structure, we could link the attentional benefits of stimulus encoding to the presence of higher-order regularities that are known to be abundant in natural images and are primarily represented at higher-level areas of the ventral stream. Population analysis revealed that the attentional signal was aligned with the compact subspace carrying information about stimulus identity. Temporal evolution of the stimulus-representing subspace revealed that alignment was not present in early responses to natural stimuli, but it emerged later, still preceding the delivery of the attentional modulation. Taken together, attentional enhancement of V1 representation of natural stimuli harnesses high-level statistical structure represented in higher visual cortical areas and is carefully and specifically aligned with the activity carrying information about natural structure.

In our experiments, animals were trained to respond swiftly to the rotation of the cued stimulus, not to recognize particular features in the image. Thus, our task was a classical spatial attention task. In agreement with previous studies allocating spatial attention caused moderate increases in discharge rate, reduced power of low-frequency oscillations, and enhanced power in the high-frequency bands of local field potentials. As expected, these effects did not depend on the stimulus structure. In agreement with psychophysical and previous electrophysiological investigations, this indicates that spatial attention enhances the salience of responses.[28] However, allocating spatial attention had the additional effect of selectively enhancing the decodability of population responses of V1 to the attended stimulus, provided that the stimulus contained higher order statistical regularities characteristic of natural scenes.

Natural scenes are evaluated by comparing sensory evidence with previously acquired priors about the statistical structure of natural environments.[1] These internal priors are stored in the functional architecture of cortical circuits at all levels of the visual processing hierarchy and some of these circuits get refined by experience to capture characteristic properties of the visual environment.[2,3,29] Recombination of feedforward connections renders neurons selective for increasingly complex constellations of features,[30–32] and the abundant horizontal intra-areal and feedback connections between processing levels allow for contextual modulation of these feature selective responses.[10,33]

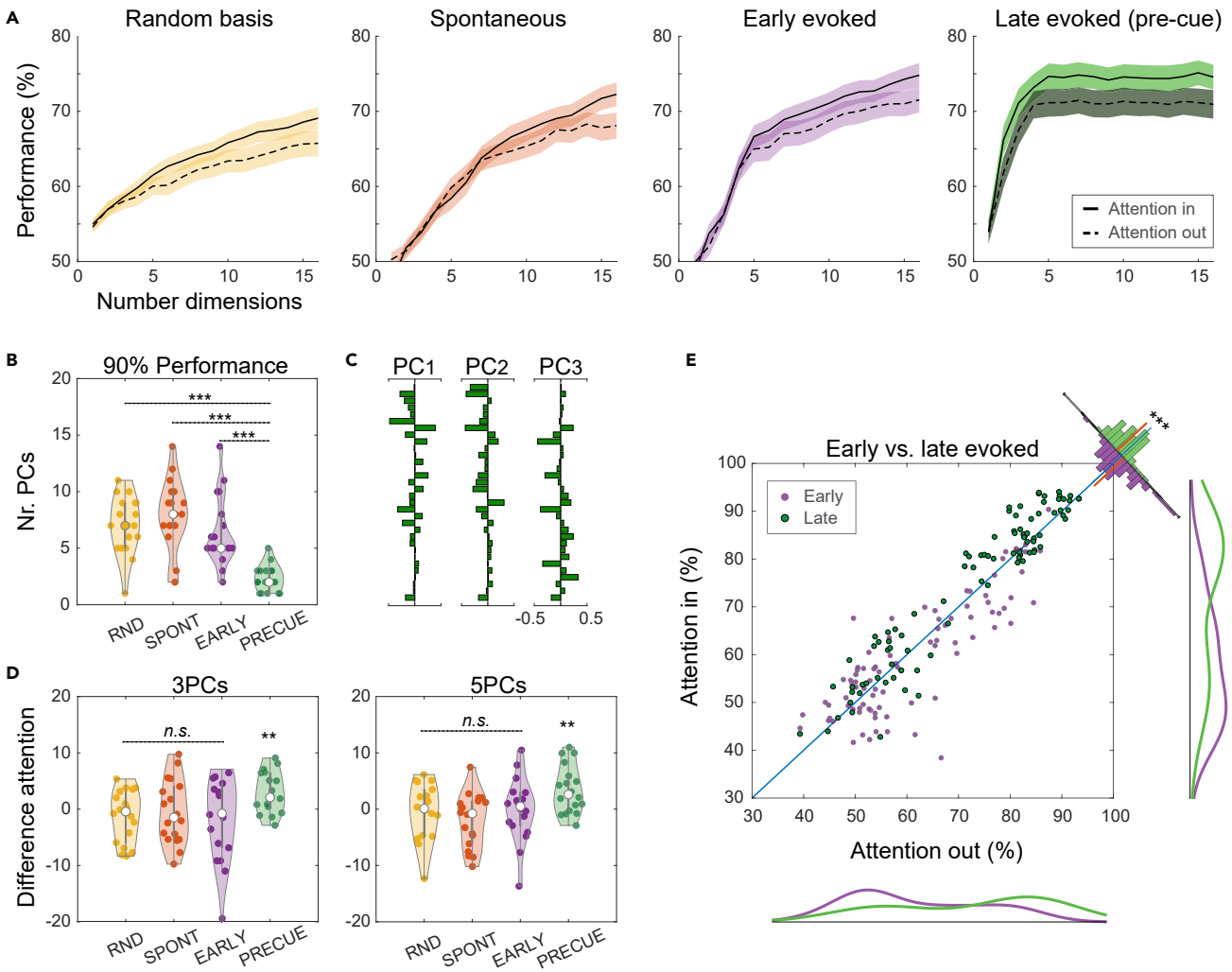


Figure 4. Geometry of population activity in response to natural scenes

(A) Spike-count vectors from the post-cue interval (1700–1900 ms) are projected in four different bases: a random projection space (yellow), a PC-space based on spontaneous activity (orange, interval -200-0 ms), a PC-space based on early evoked activity (purple, 100–300 ms window), and a PC-space based on late evoked activity (green, pre-cue, 500–700 ms window, same as Figure 3B). When using PC decompositions, the dimensions correspond to the first PCs of the PCA basis. Average performance of the stimulus decoder, across all stimulus pairs, from all recording sessions, shown as a function of the number of dimensions used to reconstruct population activity. Attention-in (solid line) and attention-out trials (dashed line) are decoded separately. Shading indicates the standard error of the mean.

(B) Number of dimensions necessary to reach 90 % of peak decoding accuracy, shown across recording sessions. The late evoked projection space (green), requires fewer than five components. Interestingly, the early evoked projection space (purple) requires up to 15 components to reach the same decoding accuracy, suggesting a higher dimensional encoding space. (*** p-val < 0.001).

(C) Composition of first principal components for the space describing responses to natural scene stimuli in an example session.

(D) Attentional difference in decoding accuracy (attention in - attention out), using a three or five-dimensional decoding space, shown across sessions. Differences are significant in the projection space defined based on late evoked activity (** p-val < 0.01), but not the alternative projection spaces (n.s. p-val > 0.05).

(E) Direct comparison of post-cue activity in the early (purple) and late evoked (green) projection spaces. A projection of the post-cue activity is used to decode stimulus identity from attention-in and attention-out trials. Dots indicate individual sessions, data pulled for the first 5 components. Marginals of decoding performances along the axes (solid lines) show overall decoding performance differences for early vs. late responses, indicating differences in the dimensionality of the population responses. Difference of attention-in and attention-out decoding performance (histogram perpendicular to the identity line) indicates differences in the alignment of the attentional modulation with the dominant activities in early and late responses (*** p-val < 0.001). Both stimulus decoding and the attentional benefits are stronger in the PC space constructed on late evoked activity.

These modulations impact stimulus saliency[34] and perceived brightness,[35,36] support perceptual grouping,[37] and figure-ground segregation.[38–41] The electrophysiological correlates of these interactions consist of changes in discharge rate and/or synchrony, and these effects tend to have longer latencies than the initial phasic responses. Therefore, context-sensitive processes are thought to be mediated by

recurrent interactions within cortical areas and top-down signaling across processing stages. Here, we argued that the overlap between contextual and attentional signals for the late responses to natural stimuli makes the inter-areal top-down account plausible. Our interpretation gained further support from the systematic changes we observed with the removal and reintroduction of higher-order statistical structure in synthetic stimuli.

In the following, we discuss how the allocation of spatial attention, which is also supposed to be mediated by top-down connections, interacts with these feature-sensitive mechanisms.

Attentional influences on visual processing have traditionally been divided into spatial[42–44] and object/feature-based attention[45,46] and it has been proposed that both contribute in complementary ways to the parsing of image content.[47,48] Our results support this notion and provide some indications as to the mechanisms underlying these complex interactions, in the context of natural stimulation. In the present task, the allocation of spatial attention contributed additional variance in the first principal components of responses to natural but not to manipulated images, and thereby enhancing the decodability of the former. The finding that spatial attention had no effect on the decodability of manipulated stimuli indicates that spatial attention has per se no refining effect on distributed stimulus representations in V1, but selectively improves representations of stimuli characterized by the higher order regularities of natural scenes. Abundant evidence indicates[49] that these higher order regularities are evaluated by downstream areas of the visual processing hierarchy. Therefore, the enhanced decodabilty of responses to natural images is likely to have been mediated by top-down signals from these areas. This raises the question, why these high-level processes were more involved when spatial attention was allocated to the stimulus. One possibility is that higher level processes do not engage by default even when stimuli match high-order priors but get involved only for stimuli to which spatial attention is allocated. In this case, spatial attention would be a prerequisite for the engagement of mechanisms that provide top-down signals commonly attributed to feature or object-specific attention, suggesting some hierarchy in the interactions between spatial and object centered or feature-specific attention. An alternative possibility is that higher level processes engage by default when stimuli match high-order priors and work cooperatively alongside spatial attention. In this case, the visual system performs a search that attempts to infer task-relevant features of an image based on both the spatial aspects of the visual scene and the low-level and high-level structural regularities. Thus, spatial and object-based attention act in unison and share an internal representation of features, with the inference slowly unfolding over reciprocal interactions across multiple hierarchical levels of processing.

The latter interpretation is supported by the observation that the natural scenes could be discriminated against surprisingly well, given the relatively low number of units and their location in area V1 (Figures 2 and 3), regardless of the attentional cue. The fact that the discriminability of natural stimuli was high even when attention was directed away from the stimulus implies that the natural scenes were efficiently encoded, irrespective of the attentional state. Previous studies found that an efficient encoding of global scene statistics remained possible in situations associated with reduced visual attention.[50] In such cases, the visual cortex is thought to extract a compressed "summary" code that does not capture the full distribution of local details yet provides a good representation of group features. Since natural scenes are structured, redundant, low-dimensional images, they are compressible. In comparison, the low-level synthetic images are difficult to compress and must be represented exhaustively without the help of internally generated or previously acquired priors on summary statistics. Thus, the resulting neuronal activity vectors for synthetic stimuli are likely to inhabit higher-dimensional or more variable substates, potentially accounting for the overall poorer performance of the classifier.

Principal component analysis was used to capture the efficient encoding of natural scenes across neuronal populations in V1. In principle component space, the late response vectors to natural stimuli reflected their low-dimensionality and could be well described by a few principal components. The higher-order stimulus structure, characteristic of natural scenes, was thus well-separated by low-dimensional subspaces. In a sense, these subspaces reflect some of the higher-order selectivity normally associated with the responses of individual neurons at higher levels of processing.[51,52] The early responses dominated by feedforward information were shown to differ from the late responses dominated by top-down information in parallel recordings of V1, V2, and V4 from macaque monkeys viewing grating stimuli.[53] This finding appears congruent with the transformation in response geometry between early and late responses to natural scenes observed here in V1. Moreover, we found that spatial attention enhanced the encoding of natural scenes along the dominant representational dimensions. Earlier, behavioral studies hinted at interactions between attentional effects and regularities beyond the simple features represented in lower visual areas.[54] Here, the match between representational and attentional signals in V1 shows a remarkable alignment, likely advantageous for efficient processing. This match appears compatible with earlier results highlighting similarities between the effects of representational learning and attention in downstream area V4.[55] Interestingly, previous theoretical arguments suggested that attentional effects are a consequence of resource optimization and result in task-dependent forms of modulation.[56] Our results provide an additional critical insight: since attentional effects were absent for synthetic stimuli, resource optimization appears to be most effective for the types of stimuli that the internal circuitry was optimized to process.

Recently, it has been shown that the receptive fields of V1 neurons are more complex than earlier described.[57] Thus, it is possible that the receptive fields of the recorded neurons matched the natural scenes better than the synthetic stimuli, which could also explain the higher discriminability of natural scenes. However, the low-dimensional neuronal representation for natural stimuli and its alignment with the attentional signals were not present for the early evoked responses, suggesting that processing time was required. The late emergence of low-dimensional neuronal representation for natural stimuli may arise through top-down feedback, as considered by our hypothesis, but alternatively through intra-areal recurrent interactions. Our V1 recordings cannot decisively determine the interactions taking place in this attention task. However, we designed synthetic stimuli based on evidence that the statistics we manipulated were characteristic of the secondary visual cortex, and therefore such manipulations are expected to affect the information that is fed back to V1, even when the

targeted responses are complex. For example, in a psychophysics study, it has been shown that the spatial scale on which phase scrambling disrupts perceptual processes is more compatible with the involvement of V2 than V1.[16] Further, phase scrambling had substantial effects on V2 activity.[21] Synthetic image manipulation was shown to contribute to population-level phenomena in V1,[15] and was confirmed in these data by the wide participation of the recorded population in the dominant PCA components. Thus, network-level changes are responsible for the enhanced representation of natural scenes across V1 responses. A parsimonious interpretation of the stimulus statistics-specificity of attentional effects and the tight overlap between subspaces is that a shared mechanism underlies the shaping of both stimulus-related responses and attention.

Spatial attention modulated the dynamics of responses, as reflected by changes in the frequency distribution of LFP power. A shift in LFP power from low to high frequencies built up gradually from the onset of the cue to the temporal window preceding the stimulus change, for both natural and synthetic stimuli (Figures 1 and S4). This kinetics resembled a hazard function reflecting the increasing probability of having to execute a response, suggesting that the reduced power of the low frequency oscillations was probably related to the increased readiness to act. In agreement with this interpretation, the oscillatory power in the theta and beta bands of responses to the target stimulus was positively correlated with reaction times in the attention-in condition. These results are consistent with previous reports. Beta oscillations have been shown to decrease during the preparation of a motor response[58] and theta band power has been shown to decrease with attention.[59] In agreement with previous work is also the finding that attention enhances the power of the broad-band high-frequency activity, which likely reflects increased spiking and synaptic activity.[60] In addition to these attention-dependent effects on dynamics, we observed a build up of gamma oscillations for natural but not for synthetic stimuli, over the course of the trial (Figure S4). The synchronization of discharges in the gamma frequency range increases for responses to features that are well predicted by the embedding context.[20,61,62] Here, natural scenes induce more gamma oscillations than the synthetic stimuli, likely because they contain more compressible features and better match the priors resident in the synaptic weight distribution of cortical networks. However, previous studies on the attentional modulation of gamma oscillations in V1 have reported mixed results[63] and this heterogeneity was also reflected here, across the two monkeys.

One puzzling aspect of our findings is the confinement of the attentional benefits of stimulus encoding to the temporal interval preceding a change in the cued stimulus. Since the task requires only a suppression of reflexive responses to distractor change, no enhancements in encoding for the natural scenes are necessary or even expected. Yet these enhancements in stimulus discriminability occurred close to the anticipated stimulus change and were more pronounced in trials with short reaction times (Figure S1B). It is conceivable that these states are particularly favorable to permit refinement of stimulus representations by structured top-down signals, suggesting that in a different task context they may carry behavioral relevance.

In summary, we showed that the spatial allocation of attention toward a natural stimulus can engage mechanisms that exploit the higher order statistical regularities of natural images, resulting in enhanced decodability of neuronal population responses in area V1. The alignment between the attentional and natural stimulus variance in low-dimensional projections of V1 activity vectors, which was absent for synthetic low-level stimuli, suggests that attention can involve mechanisms optimized for the processing of natural images in order to refine stimulus representation in V1. These results highlight the importance of using natural stimuli when studying sensory processing and provide important insights into how such factors as natural image statistics and the animals' internal models of the visual world are central to visual processing even at early levels.

Limitations of the study

Here we provided evidence that attention enhances the specificity of neuronal population responses to natural stimuli in area V1 by modulating the distributed responses to higher-order features present in natural images. Our work, however, has several limitations. First, the attention task employed here is a classical spatial attention task. Extensions of this work should consider training animals to respond to finer changes in image content. In such a setup, enhancements in stimulus representation could result in stronger behavioral consequences. Other potential extensions could target manipulations of rewards. Second, the observed enhancement in natural scene encoding was restricted to a small temporal window preceding the stimulus rotation. A variation in the expected timing interval of the change would allow for a more in-depth exploration of the alignment between contextual and attentional feedback. Thirdly, a larger image set and a thorough characterization of its properties in terms of structural predictability and compressibility would allow for additional insights into the alignment of attentional signals and contextual signals arising from inter- and intra-areal recurrent interactions. Finally, our results were based on recordings of neural activity solely in V1. A more thorough assessment of top-down computations will require multi-site recordings of task-engaged animals.

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 STUDY PARTICIPANT DETAILS
  - Electrophysiological recordings

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.108816.

ACKNOWLEDGMENTS

We thank Gareth Bland for providing data management support and technical support in the laboratory. Many thanks to Patrick Jendritza for insightful discussions in the early days of the project. This work was supported by the Human Frontier Science Program (RGP0044/2018, A.L., G.O., and W.S.), the Deutsche Forschungsgemeinschaft (DFG Reinhart Koselleck Project 325248489, W.S.) and the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory (G.O.).

AUTHOR CONTRIBUTIONS

A.L., G.O., and W.S. conceived and designed the project. A.L., L.K., and J.K.-L. wrote the stimulation protocol and recorded the data. A.L., M.B. and G.O. generated synthetic stimuli. A.L. analyzed the data. A.L., G.O, and W.S wrote the article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 9, 2023
Revised: July 18, 2023
Accepted: January 2, 2024
Published: January 6, 2024

REFERENCES

1. Helmholtz, H.V. (1867). Handbuch der physiologischen Optik. Voss.
2. Singer, W., and Tretter, F. (1976). Unusually large receptive fields in cats with restricted visual experience. Exp. Brain Res. *26*, 171–184.
3. Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. Science *331*, 83–87.
4. Bányai, M., Lazar, A., Klein, L., Klon-Lipok, J., Stippinger, M., Singer, W., and Orbán, G. (2019a). Stimulus complexity shapes response correlations in primary visual cortex. Proc. Natl. Acad. Sci. USA *116*, 2723–2732.
5. Wurtz, R.H., and Mohler, C.W. (1976). Enhancement of visual responses in monkey striate cortex and frontal eye fields. J. Neurophysiol. *39*, 766–772.
6. Haenny, P.E., and Schiller, P.H. (1988). State dependent activity in monkey visual cortex. Exp. Brain Res. *69*, 225–244.
7. Duncan, J. (1984). Selective attention and the organization of visual information. J. Exp. Psychol. Gen. *113*, 501–517.
8. Roelfsema, P.R., Lamme, V.A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. Nature *395*, 376–381.
9. Ferro, D., van Kempen, J., Boyd, M., Panzeri, S., and Thiele, A. (2021). Directed information exchange between cortical layers in macaque v1 and v4 and its modulation by selective attention. Proc. Natl. Acad. Sci. USA *118*, e2022097118.
10. Chen, M., Yan, Y., Gong, X., Gilbert, C.D., Liang, H., and Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. Neuron *82*, 682–694.
11. Lee, T.S., and Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. Proc. Natl. Acad. Sci. USA *98*, 1907–1911.
12. Vinje, W.E., and Gallant, J.L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. Science *287*, 1273–1276.
13. Haider, B., Krause, M.R., Duque, A., Yu, Y., Touryan, J., Mazer, J.A., and McCormick, D.A. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. Neuron *65*, 107–121.
14. Baudot, P., Levy, M., Marre, O., Monier, C., Pananceau, M., and Frégnac, Y. (2013). Animation of natural scene by virtual eye-movements evokes high precision and low noise in v1 neurons. Front. Neural Circuits *7*, 206.
15. Froudarakis, E., Berens, P., Ecker, A.S., Cotton, R.J., Sinz, F.H., Yatsenko, D., Saggau, P., Bethge, M., and Tolias, A.S. (2014). Population code in mouse v1 facilitates readout of natural scenes through increased sparseness. Nat. Neurosci. *17*, 851–857.
16. Freeman, J., and Simoncelli, E.P. (2011). Metamers of the ventral stream. Nat. Neurosci. *14*, 1195–1201.
17. Angelucci, A., and Bressloff, P.C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate v1 neurons. Prog. Brain Res. *154*, 93–120.
18. Bányai, M., Nagy, D.G., and Orbán, G. (2019). Hierarchical semantic compression predicts texture selectivity in early vision. In Proceedings of the Conference on Cognitive Computational Neuroscience.
19. Kayser, C., Salazar, R.F., and Konig, P. (2003). Responses to natural scenes in cat v1. J. Neurophysiol. *90*, 1910–1920.
20. Uran, C., Peter, A., Lazar, A., Barnes, W., Klon-Lipok, J., Shapcott, K.A., Roese, R., Fries, P., Singer, W., and Vinck, M. (2022). Predictive coding of natural images by v1 firing rates and rhythmic synchronization. Neuron *110*, 1240–1257.e8.
21. Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., and Movshon, J.A. (2013). A functional and perceptual signature of the second visual area in primates. Nat. Neurosci. *16*, 974–981.
22. Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., and Goldstein, T. (2021). The intrinsic dimension of images and its impact on

learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.2104.08894.
23. Simoncelli, E.P., and Olshausen, B.A. (2001). Natural image statistics and neural representation. Annu. Rev. Neurosci. *24*, 1193–1216.
24. Hegdé, J., and Van Essen, D.C. (2004). Temporal dynamics of shape analysis in macaque visual area v2. J. Neurophysiol. *92*, 3030–3042.
25. Matsumoto, N., Okada, M., Sugase-Miyamoto, Y., Yamane, S., and Kawano, K. (2005). Population dynamics of face-responsive neurons in the inferior temporal cortex. Cereb. Cortex *15*, 1103–1112.
26. Lazar, A., Lewis, C., Fries, P., Singer, W., and Nikolic, D. (2021). Visual exposure enhances stimulus encoding and persistence in primary cortex. Proc. Natl. Acad. Sci. USA *118*, e2105276118.
27. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K.D. (2019). High-dimensional geometry of population responses in visual cortex. Nature *571*, 361–365.
28. Treue, S. (2003). Visual attention: the where, what, how and why of saliency. Curr. Opin. Neurobiol. *13*, 428–432.
29. Löwel, S., and Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. Science *255*, 209–212.
30. 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.
31. Hirabayashi, T., Takeuchi, D., Tamura, K., and Miyashita, Y. (2013). Microcircuits for hierarchical elaboration of object coding across primate temporal areas. Science *341*, 191–195.
32. Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.
33. Liang, H., Gong, X., Chen, M., Yan, Y., Li, W., and Gilbert, C.D. (2017). Interactions between feedback and lateral connections in the primary visual cortex. Proc. Natl. Acad. Sci. USA *114*, 8637–8642.
34. Knierim, J.J., and Van Essen, D.C. (1992). Neuronal responses to static texture patterns in area v1 of the alert macaque monkey. J. Neurophysiol. *67*, 961–980.
35. Rossi, A.F., Rittenhouse, C.D., and Paradiso, M.A. (1996). The representation of brightness in primary visual cortex. Science *273*, 1104–1107.
36. Biederlack, J., Castelo-Branco, M., Neuenschwander, S., Wheeler, D.W., Singer, W., and Nikolić, D. (2006). Brightness induction: Rate enhancement and neuronal synchronization as complementary codes. Neuron *52*, 1073–1083.
37. Kapadia, M.K., Ito, M., Gilbert, C.D., and Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in v1 of alert monkeys. Neuron *15*, 843–856.
38. Payne, B.R., Lomber, S.G., Villa, A.E., and Bullier, J. (1996). Reversible deactivation of cerebral network components. Trends Neurosci. *19*, 535–542.
39. Lamme, V.A., Supèr, H., and Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. Curr. Opin. Neurobiol. *8*, 529–535.
40. Lamme, V.A., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. *23*, 571–579.
41. Hembrook-Short, J.R., Mock, V.L., and Briggs, F. (2017). Attentional modulation of neuronal activity depends on neuronal feature selectivity. Curr. Biol. *27*, 1878–1887.e5.
42. Duncan, J., Humphreys, G., and Ward, R. (1997). Competitive brain activity in visual attention. Curr. Opin. Neurobiol. *7*, 255–261.
43. Cohen, M.R., and Maunsell, J.H.R. (2011). Using neuronal populations to study the mechanisms underlying spatial and feature attention. Neuron *70*, 1192–1204.
44. Tootell, R.B., Hadjikhani, N., Hall, E.K., Marrett, S., Vanduffel, W., Vaughan, J.T., and Dale, A.M. (1998). The retinotopy of visual spatial attention. Neuron *21*, 1409–1422.
45. Bichot, N.P., Rossi, A.F., and Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area v4. Science *308*, 529–534.
46. Treisman, A. (1998). The perception of features and objects. Visual Attention *8*, 26–54.
47. Mazer, J.A. (2011). Spatial attention, feature-based attention, and saccades: three sides of one coin? Biol. Psychiatry *69*, 1147–1152.
48. Goddard, E., Carlson, T.A., and Woolgar, A. (2022). Spatial and feature-selective attention have distinct, interacting effects on population-level tuning. J. Cogn. Neurosci. *34*, 290–312.
49. Orban, G.A. (2008). Higher order visual processing in macaque extrastriate cortex. Physiol. Rev. *88*, 59–89.
50. Alvarez, G.A., and Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. Proc. Natl. Acad. Sci. USA *106*, 7345–7350.
51. Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science *229*, 782–784.
52. Sheinberg, D.L., and Logothetis, N.K. (1997). The role of temporal cortical areas in perceptual organization. Proc. Natl. Acad. Sci. USA *94*, 3408–3413.
53. Semedo, J.D., Jasper, A.I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C.K., Kohn, A., and Yu, B.M. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. Nat. Commun. *13*, 1099.
54. Freeman, E., Driver, J., Sagi, D., and Zhaoping, L. (2003). Top-down modulation of lateral interactions in early vision: does attention affect integration of the whole or just perception of the parts? Curr. Biol. *13*, 985–989.
55. Ni, A.M., Ruff, D.A., Alberts, J.J., Symmonds, J., and Cohen, M.R. (2018). Learning and attention reveal a general relationship between population activity and behavior. Science *359*, 463–465.
56. Młynarski, W., and Tkačik, G. (2022). Efficient coding theory of dynamic attentional modulation. PLoS Biol. *20*, e3001889.
57. Tang, S., Lee, T.S., Li, M., Zhang, Y., Xu, Y., Liu, F., Teo, B., and Jiang, H. (2018). Complex pattern selectivity in macaque primary visual cortex revealed by large-scale two-photon imaging. Curr. Biol. *28*, 38–48.e3.
58. Engel, A.K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? Curr. Opin. Neurobiol. *20*, 156–165.
59. Spyropoulos, G., Bosman, C.A., and Fries, P. (2018). A theta rhythm in macaque visual cortex and its attentional modulation. Proc. Natl. Acad. Sci. USA *115*, E5614–E5623.
60. Ferro, D., van Kempen, J., Boyd, M., Panzeri, S., and Thiele, A. (2021b). Directed information exchange between cortical layers in macaque v1 and v4 and its modulation by selective attention. Proc. Natl. Acad. Sci. USA *118*, e2022097118.
61. Peter, A., Uran, C., Klon-Lipok, J., Roese, R., van Stijn, S., Barnes, W., Dowdall, J.R., Singer, W., Fries, P., and Vinck, M. (2019). Surface color and predictability determine contextual modulation of v1 firing and gamma oscillations. Elife *8*, e42101.
62. Lewis, C.M., Ni, J., Wunderle, T., Jendritza, P., Lazar, A., Diester, I., and Fries, P. (2021). Cortical gamma-band resonance preferentially transmits coherent input. Cell Rep. *35*, 109083.
63. Chalk, M., Herrero, J.L., Gieselmann, M.A., Delicato, L.S., Gotthardt, S., and Thiele, A. (2010). Attention reduces stimulus-driven gamma frequency oscillations and spike field coherence in v1. Neuron *66*, 114–125.
64. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. (2011). Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 1–9.
65. Markowitz, D.A., Wong, Y.T., Gray, C.M., and Pesaran, B. (2011). Optimizing the decoding of movement goals from local field potentials in macaque cortex. J. Neurosci. *31*, 18412–18422.
66. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Rhesus macaque (Macaca mulatta)	Medical Research Council Center for Macaques, Porton Down, Salisbury, SP4 0JQ	NA
Software and algorithms		
MATLAB	Mathworks	www.mathworks.com
FieldTrip	FieldTrip[64]	www.fieldtriptoolbox.org
Psychophysics Toolbox	Psychophysics Toolbox	www.psychtoolbox.org
Other		
Microdrive 32 electrodes	Gray Matter Research[65]	www.graymatter-research.com
Pre-amplifier (PZ2)	Tucker-Davis Technologies	www.tdt.com
Eye tracker (ET-49)	Thomas Recording	www.thomasrecording.com

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Wolf Singer (wolf.singer@esi-frankfurt.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact (W.S.) upon request.
- Software and packages used for the analysis include FieldTrip.[64] This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The data was obtained from two adult rhesus macaque monkeys, one male (monkey A) and one female (monkey I), aged 8 and 12 years, respectively, during the time of the study. All experimental procedures were approved by the local authorities (Regierungspräsidium, Hessen, Darmstadt, Germany) and were in accordance with the animal welfare guidelines of the European Union's Directive 2010/63/EU. Animals were housed in rooms with outdoor access to a play area and had regular veterinary care and balanced nutrition. The recording chamber was implanted under general anesthesia over the primary visual cortex, the exact location was determined based on stereotactic coordinates derived from MRI and CT scans.

Electrophysiological recordings

Signals were recorded using a chronically implanted microdrive containing 32 independently movable glass-coated tungsten electrodes with impedance between 0.7 and 1.5 M Ω and 1.5 mm inter-electrode distance (SC32; Gray Matter Research[65]), amplified (Tucker-Davis Technologies, PZ2 pre-amplifier) and digitized at a rate of 24.4 kHz. The signals were filtered between 300 and 4,000 Hz and a threshold was set at 4SD above noise level to extract multi-unit activity. LFP signals were obtained by low-pass filtering at 300 Hz and downsampling to 1.5 kHz.

Behavioral paradigm

Animals were seated in a custom primate chair at a distance of 64 cm in front of a 477 × 298 mm monitor (Samsung SyncMaster 2233RZ; 120 Hz refresh rate; gamma-corrected). Stimulation protocols were written using MATLAB (MathWorks) and Psychophysics Toolbox. At the start of each recording week, the receptive fields and orientation preferences of the recorded units were mapped with a moving light bar drifting in a randomized sequence in eight different directions.

The two monkeys performed an attention-modulated change detection task. During the task, eye tracking was performed using an infrared-camera eye-control system (ET-49; Thomas Recording). To initiate a trial, the monkey maintained fixation on a white spot (0.1° visual angle) presented in the center of a black screen and pressed a lever. After 500 ms, two identical visual stimuli appeared in an aperture of 2.8–5.1° at a distance of 2.3–3.2° from the fixation point. One of the stimuli covered the receptive fields of the recorded units, which were situated in the right hemifield, the other stimulus was placed at the mirror symmetric site in the left hemifield. After an additional 700 ms, the fixation spot changed color from white to either red or blue, cuing the monkey to covertly direct its attention to one of the two stimuli (red, right hemifield; blue, left hemifield). When the cued image was rotated (20°), the monkey released a lever in a fixed time window (600 ms for monkey A; 900 ms for monkey I) to receive a reward. A break in fixation (fixation window, 1.5° diameter) or an early lever release resulted in the abortion of the trial, which was announced by a tone signal.

Eye movements

The percentage of saccadic eye movements that exceeded the fixation window (1.5° diameter) over the course of the trial is shown in Figure S5A. These breaks in fixation occurred less frequently in the time interval preceding a stimulus change (1700-1900 ms, marked in gray) compared to the baseline (-200-0ms, marked in gray). This difference was highly significant (Figure S5B, Wilcoxon signed-rank test, $p = 0.00018$, $n = 18$ sessions in 2 monkeys), suggesting that the eyes were very stable in the time interval of interest. Additionally, we found no differences in the frequency of saccadic eye movements between natural and synthetic stimuli (Figure S5C). Finally, we found no differences in the frequency of saccadic eye movements with attention for neither natural scenes (Figure S5D) nor synthetic stimuli (Figure S5E). Taken together, these results suggest that eye movements are unlikely to be strong contributors to the differences observed in our data.

Visual stimulus design

Stimuli were static, black and white images, presented in a square or circular aperture. Within each recording session, all natural scenes and control images had equal luminance and contrast.

Control images were generated in two ways:

(1) Filter-scrambling: synthetic stimuli generated from an image model. Filter scrambling was realized by permuting Gabor filter activations elicited by a natural image across the elements of the complete filter bank. The filter bank was composed of a large set of Gabor functions, fitted to the receptive field (RF) characteristics of the recorded neurons. The positions and orientations of the Gabor functions covered the image uniformly, while their size was matched to the RFs of visual cortical neurons recorded at the same eccentricity. An activation variable determined the level of contribution of each particular Gabor function to the image. The synthetic images were generated by linearly combining the activation-scaled Gabor functions. For each synthetic image, the activations of 500–3,000 Gabor functions were sampled from the empirical distribution of Gabor filter responses to a particular natural image. The resulting control images lacked the higher-order structure of natural scenes but matched their low-level statistical properties.

(2) Phase-scrambled images. The 2D fast Fourier transform (FFT) of each natural image was computed to obtain a complex magnitude-phase map. The phase values were scrambled by assigning a random value to each element taken from a uniform distribution across the range $(-\pi, \pi)$. An inverse FFT was then applied to the resulting magnitude-phase maps to produce scrambled versions of the original natural images. These control images lacked the higher-order structure of natural scenes but matched their frequency spectrum.

In a second experiment, recorded in one monkey (monkey I), we contrasted synthetic stimuli with and without higher-order structure. These synthetic images were generated similarly to the control images described in (1) and matched the low-level statistical properties of natural scenes. To add higher-order structure, a subset of Gabor functions were arranged in a manner that produced simple contour-like patterns (example of contour synthetic image in Figure 3D).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

Data analysis was performed using custom code in MATLAB (MathWorks) and the Fieldtrip toolbox.[64] Data analysis was based on completed correct trials (mean number per session 967 trials for monkey A; 720 trials for monkey I).

We applied non-parametric statistical tests to avoid assumptions about the distributions of the empirical data. Information about sample variables and size is reported in the results section. Critical results and statistics are reported separately for individual animals in the supplementary materials.

Discriminability index

The *unit d'*, also known as Cohen's effect size,[66] for a pair of stimuli, was calculated as:

$$d'_i = (m_1 - m_2) / \sigma \quad \text{(Equation 1)}$$

where m_1 and m_2 are the mean spike-counts across trials of unit i to the two stimuli and $\sigma = (\sigma_1 + \sigma_2)/2$ is the mean standard deviation. This measure was used in Figures 2G, 2H, S3C and S3D.

The *population d'* was calculated similarly, except in this case m_1, m_2 and σ are *n*-dimensional vectors of spike-counts, where *n* is the number of simultaneously recorded units in a session. Distances in vector space were calculated using the Euclidean distance. The *population d'* was used in Figures 1, 2, S1, and S2.

Principal component analysis and stimulus classification

PCA was applied on population spike-count vectors calculated over a 200 ms time window in the trial, 500–700 ms after stimulus onset (pre-cue) and included trials from both attentional conditions. The control stimuli and contour stimuli were analyzed separately, in a similar manner. Thus, the projection space obtained via PCA was different for natural scenes and control stimuli. A comparison of the percentage of variance explained by an increasing number of principal components, for natural scenes and controls recorded in the same sessions, can be found in Figure S7.

To test alignement between the stimulus and the attentional variance (Figure 3), spike-counts over the 1700–1900 ms interval (post-cue) were projected into the PCA space constructed pre-cue. Stimulus classification was performed based on projected data, separately, for trials belonging to the attention in/out conditions. For each attentional condition, Naive Bayes classifiers were trained to decode the stimulus identity based on data points mapped in the space described by the first *n* principal components, with $n = 1$ to 16. Cross-validation was performed by randomly subsampling the data ($k - 1$ data partitions used for training, 1 used for test, *k* repetitions; $k = 5$). This meant that, for each pair of stimuli and each *n*, we ran classifiers $k = 5$ times, on each iteration randomly sampling the population response to the two stimuli. The unseen trials were then used to assess test performance. The performance values reported in Figure 3B are mean validation scores pulled across all stimulus pairs and all recording sessions, with shaded areas representing the standard error of the mean. Chance level was 50 %.

To enquire whether stimulus decoders benefit from knowledge on correlated variability across trials, stimulus classification for variable numbers of PCs was compared for shuffled and unshuffled data in Figure S7. In this case, shuffling was performed across trials, within each stimulus condition (i.e., signal correlations were not affected), after the construction of the PCA projection space, but before the training of the Bayesian classifier. Test data was unshuffled, so that the distribution of original spike-counts was not affected.

Three additional projection spaces were considered for the analysis presented in Figure 4. First, a random orthogonal basis space was generated with the same number of dimensions as the original population space. Second, PCA was computed either on spontaneous activity (spike-count vectors over the -200-0 ms window) or early evoked activity (spike-count vectors over the 100–300 ms window). These projections were performed in individual sessions and averaged over sessions.

LFP analysis

Power spectra were computed using a frequency-dependent window length (5 cycles per time window). This approach decreases the temporal smoothing at higher frequencies and increases the sensitivity to brief effects. The time-windows were moved in steps of 10 ms and Hann-tapered to avoid spectral leakage.

LFP power differences between trials in the attention-in and attention-out conditions are captured in Figure 1D (18 sessions, precue baseline substracted). A more extensive analysis of attentional differences and stimulus-type differences was performed in individual animals (Figure S4).

Stimulus classification (Naive Bayes) based on LFP power in various frequency bands, was applied in individual animals, following a similar cross-validation procedure as the one described above for the PCA projections (Figure S8).