Spatial frequency channels mediate a visual metric for human spatial perception

Shijia Zhang[1,#], Ce Mo[2,#,*], Lei Mo,[1,*], Fang Fang[3,4,5,*]

1 Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, China.

2 Department of Psychology, Sun-Yat-Sen University, Guangzhou, China

3 School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China.

4 IDG/McGovern Institute for Brain Research, Peking University, Beijing, China.

5 Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China.

#equal contribution

*Email: moce3@mail.sysu.edu.cn; molei@m.scnu.edu.cn; ffang@pku.edu.cn

Abstract

An intriguing and ubiquitous phenomenon in vision is the perception of spatial property information (e.g. separation, size, shape etc.) that departs from the raw physical inputs yet conforms to the visual context. While it has been postulated that visual system employs an internal adaptable metric by which the signals conveying the spatial property information are scaled, it remains unknown whether and how such putative visual metric is represented in the brain. Here, we hypothesized the putative visual metric is mediated by the joint responses of differently-tuned spatial frequency (SF) channels, and investigated this idea by combining the use of psychophysics, fMRI-based population receptive field (pRF) mapping and computation modeling. We found that reweighting of spatial frequency (SF) channels, either by adaption or presentation of the corresponding SF background, led to systematic distortions of the perceived spatial property information. Moreover, the pRFs in V1 were more concentrated towards the fovea under the high SF stimulation regardless of their positions in the visual field. Importantly, both the perceptual distortions and the global-scale pRF displacement were functionally coupled with the changes in the population SF response induced by SF channel reweighting. Our findings revealed, for the first time, the neural apparatus signaling the putative visual metric that constrains spatial vision by establishing the comprehensive three-way connection between SF channel reweighting, pRF

position displacement and perceptual distortions of spatial property information.

Keywords: visual metric, spatial property perception, spatial frequency channel, population receptive field

Introduction

The cornerstone of visual processing is the perception of fundamental spatial properties of visual objects (size, length and separation, etc.), a process known as spatial vision. However, our spatial vision is rarely a veridical Cartesian projection of the external physical world[1]. As such, signals encoding the spatial property information must be cognitively scaled in some ways to create the visual experience that departs from the raw retinal input, as reflected in a plethora of geometrical illusions. For example, the same visual object differs drastically in apparent size when surrounded by small versus large items (the Ebbinghaus illusion)[2–4]. On the other hand, objects that project different retinal images could nevertheless be perceived as having the same size if they are associated with cues of different viewing distances[5]. At the neural level, the scaling of spatial property signals is reflected in the spatial profile of neural activities in primary visual cortex (V1)[6]. Specifically, a perceptually larger object elicits a more widespread activation in V1 than an object of equal retinal size but perceived to be smaller[7,8]. Moreover, it was found that the activated area of V1 by a visual stimulus closely matches the perceptual, rather than retinal, stimulus size that is altered by size adaptation[9].

The ubiquity of spatial property information scaling has given rise to the elegant theory of an internal, adaptable metric that weaves the 'fabrics' of visual space[10–13]. Perception of spatial property information (e.g. size or length) changes with sensory inputs or with the putative metric relative to which such information is scaled. Specifically, narrowing and expanding the unit area of the visual metric increases and decreases the number of metric units spanned by the visual object, thereby leading to corresponding inflation and shrinkage in perception, respectively[14,15]. Interestingly, the idea of an adaptable visual metric provides an excellent empirical explanation of the classical Oppel-Kundt illusion in which a part of the visual field subdivided by an array of

dense visual elements appears larger than the undivided part of the same spatial extent, as the intervening elements might impose an extra metric with small unit area[16–18]. More recently, seminal psychophysical findings by Hisakata and colleagues have shown that the metric signal with large unit area could be rendered dominant by adaptation to dense random dot textures[11,19,20], thereby causing subsequent reduction in both apparent separation and apparent size of the test stimuli. While these findings suggested that the underlying neural signals of visual metric could be modulated by visual contexts, it remains unknown where the putative visual metric is represented in the brain and how the perception of visual space is shaped by the adaptive changes of such representations.

The neural apparatus responsible for signaling the visual metric is likely comprised of a vast array of sensory pathways tuned to different spatial frequencies (SF) in early visual cortex, i.e. SF channels. The joint response of these SF channels constrains visual system's sensitivity to information at different spatial scales (i.e. spatial resolution). Whereas the enhancement of high SF channels facilitates processing of the fine-grained information, biasing towards the low SF channels instead optimizes the processing of coarse-grained information. Moreover, the relative weighting of these SF channels is also intrinsically related to the spatial profile of neuronal receptive fields (RF). Concentration of neuronal RFs is accompanied by a shift of the population's tuning preference towards the high SF, as it sharpens the population's representation of slight variations in local contrast and magnifies the otherwise negligible difference in response patterns elicited by two nearby stimuli in space. In contrast, the sparse tiling of neuronal RFs interferes with precise positional coding and dampens the population's response to high SFs, as it weakens the population's ability of resolving the fine-grained offsets in the stimulus image (**Figure 1**). Consistent with this idea, both neuronal RF concentration and sensitivity to high SF peak at the fovea and gradually decline with eccentricity in parallel to the systematic variation of spatial resolution across the visual space[21,22]. Further support comes from the mounting evidence of attentional modulation effects on perceived SF, spatial resolution and neuronal RF profiles[23]. Exogenous attention inflexibly increases the apparent SF at the target location, causing perceptual benefits in the periphery and costs in the fovea where the original spatial resolution is too low and too high for the behavioral task, respectively[24,25]. The simultaneous changes in perceived SF and

spatial resolution are thought to be mediated by the attention-induced neuronal RF center shifts towards the target location[26–28], which occurs not only in the vicinity of the attentional focus but also globally across the entire visual field[29]. Based on these findings, we hypothesized that visual metric is signaled by the preferred SF of the population response, with the area representing one luminance variation cycle (i.e. inverse of the preferred SF) corresponding to the 'grain size' (i.e. unit area) of the metric. Physiologically, amplification of the high and the low SF channels is functionally coupled with a concentration and a diffusion of the neuronal RFs, leading to increase and decrease of the metric's granularity, respectively. Finally, perceived spatial properties of visual objects are read out as the number of metric units spanned by the physical stimulus as if a mental 'ruler' is in play.

Here, we examined this hypothesis by investigating the three-way connection between SF sensitivity, perception of spatial properties and the spatial profile of RFs in early visual areas by combining visual psychophysics, functional magnetic resonance imaging (fMRI) and computation modeling. We found that both perceived separation and perceived shape were shrank by adaptation to high SF stimuli, and were inflated by selective attenuation of the low SF channel response. Using the fMRI-based population receptive field (pRF) mapping technique, we found that enhancement of the response of the high and the low SF channels was functionally coupled with global concentration and dispersion of the pRFs in V1, respectively. More importantly, a computational model based on the dynamical reweighting of SF channels accurately captured both changes in the pRF profile and the magnitude of distortion of perceived length induced by selectively driving the high and the low SF channels. Our findings thus provided novel and converging evidence that the putative visual metric governing spatial vision is signaled by the SF channels.

Results

Selective inhibition of SF channel via adaptation alters spatial property perception

In the psychophysical adaptation experiment, we investigated whether spatial property perception could be influenced by selective inhibition of SF channels. Adaptation is a time-honored tool in psychophysics for modulating the excitability of specific neuronal populations. The adapting

stimuli were two sinusoidal gratings simultaneously presented on the horizontal meridian at equal eccentricities (**Figure 2a**). One of the gratings had a constant SF of 0.25 cycles/° (fixed adaptor) while the SF of the other grating varied from 0.25 to 1.5 cycles/° (matching adaptor) in different blocks. After a 30-s pre-adaptation and a 5-s topping-up adaptation, subjects were tested for the perception of two fundamental spatial properties, namely separation and shape, in two different tasks. The task stimuli were two pairs of dots for separation perception and two ellipses for shape perception, which were presented for 100ms at the same locations as the adaptors (**Figure 2b**). Notably, the task stimulus ipsilateral to the fixed adaptor served as the standard stimulus that remained unchanged throughout the experiment, while the task stimulus ipsilateral to the matching adaptor was varied from trial to trial (i.e. probe) using a staircase procedure. Subjects reported the pair of dots that appeared more separated (separation judgment task) or the ellipse that appeared more elongated (shape judgment task).

In both tasks, because the data from the blocks in which the fixed adaptor was presented in the left and in the right visual fields showed a similar pattern, data from all blocks was pooled together and fitted with a cumulative Gaussian function. For each matching adaptor, we obtained the psychometric curve that described the percentage of trials in which the probe was chosen as a function of the corresponding physical feature (separation or aspect ratio) of the probe (**Figure 2c**). For the separation judgment task, when the matching adaptor shared the same SF as the fixed adaptor (0.25 cycles/°), the psychometric curve centered at the physical separation of the standard stimulus. However, as the SF of the matching adaptor became higher, the psychometric function showed a consistent horizontal rightward shift (**Figure 2c** left), indicating that the perceived separation became increasingly smaller when subjects adapted to increasingly higher SFs. A similar pattern was found in the results of the shape judgment task (**Figure 2c** right). It appeared as if both types of probes were ‘squeezed’ and ‘stretched’ when the high and the low SF channels were attenuated by adaptation, respectively, suggesting that perception of various spatial properties was commonly influenced by modulation of SF channels. To measure the SF adaptation effect, we first identified the critical point at which the probe and the standard stimulus were perceptually indistinguishable in terms of separation and the aspect ratio (the point of subjective equality, PSE). The magnitude of adaptation effect was calculated as the relative difference

between the PSE and the physical feature (separation or aspect ratio) of the standard stimulus in percentage units. As shown in **Figure 2d**, the magnitude of adaptation effect did not significantly differ from 0 when the SF of both adaptors was 0.25 cycles/° in either task (separation judgment task: mean ± S.E.M: -0.25 ± 1.05%, one-sample *t*-test: t_{11} = -0.25, p = 0.809; shape judgment task: mean ± S.E.M: -0.18 ± 0.83%, one-sample *t*-test: t_{11} = -0.22, p = 0.827). However, the magnitude of adaptation effect exhibited a monotonically climbing pattern with the increasing matching adaptor SF, which was consistently observed in both tasks (separation judgment: $F_{(3,36)} = 8.68$, $p < 0.001$; shape judgment: $F_{(3,36)} = 11.98$, $p < 10^{-4}$). These findings showed that adaptation to different SFs led to a clear discrepancy in both perceived separation and perceived shape in proportion to the SF difference between the two adaptors, which revealed an intrinsic causal connection from SF sensitivity changes to distortions of spatial property perception.

SF channel enhancement is coupled with a global-scale shift of V1 pRFs

Next, we investigated the relationship between neuronal RF profiles and SF channel modulation by examining the RF properties when the high and the low SF channels were selectively driven, respectively. To this end, we used fMRI-based population receptive field (pRF) mapping, a well-established technique that allows for effective characterization of neuronal RF profiles in human visual system[30,31]. This technique models the joint receptive field of the underlying neuronal population within a voxel as an isotropic 2-D Gaussian function. The location and the size of the pRF could be determined as the Gaussian function parameters that yield the best fit of the predicted response to the empirical BOLD signal. We focused on V1 as our primary cortical site of interest due to its well-established role in processing spatial property information, as well as its highest sensitivity to changes in RF profiles afforded by the smallest RF size[32–34]. For each voxel in V1, the pRF was mapped using two otherwise identical pRF mapping schemes (**Figure 3a**) that differed only in the stimulus SF (0.25 cycles/° and 1.5 cycles/°), which were designed to drive the high and the low SF channels, respectively. Unless otherwise specified, the reported findings were based on the same set of voxels for which the pRF model could explain at least 25% of the variance of the raw data in both SF mapping conditions (mean percentage of voxels included ± S.E.M.: 40.43 ± 2.12%). Notably, similar findings were obtained when we repeated the same analyses with two different criteria for inclusion of voxels (20% and 30%, see

supplementary information).

We first compared the mean pRF eccentricity and the mean pRF size of all the included voxels between the two SF mapping conditions (**Figure 3b**). We found that the average pRF eccentricity was significantly smaller in the high SF mapping condition in comparison to that in the low SF mapping condition (mean ± S.E.M.: high SF: 4.41° ± 0.16°; low SF: 4.88° ± 0.16°, paired *t*-test: t_{17} = -6.50, $p < 10^{-5}$). Moreover, the average inter-pRF distance was reduced in the high SF mapping condition (mean ± S.E.M.: high SF: 6.45° ± 0.15°; low SF: 6.96° ± 0.17°, paired *t*-test: t_{17} = -6.68, $p < 10^{-5}$). Interestingly, however, no significant difference in the average pRF size was found between the two SF mapping conditions (mean ± S.E.M.: high SF: 0.98° ± 0.04°; low SF: 1.00° ± 0.04°, paired *t*-test: t_{17} = -0.69, p = 0.502). One possibility is that the difference in pRF size between the two conditions varied with eccentricity, which was concealed to the averaged results. To test this possibility, we fitted a line relating pRF eccentricity with pRF size for each SF mapping condition for each individual subject. We found that pRF size increased with pRF eccentricity in a highly similar rate between the two SF mapping conditions (mean ± S.E.M.: high SF: slope: 0.11 ± 0.01, intercept: 0.62 ± 0.05; low SF: slope: 0.12 ± 0.01, intercept: 0.61 ± 0.06) with no significant difference (Wilcoxon signed-rank test: slope: p = 0.528; intercept: p = 0.811) (**Figure 3c**), suggesting that modulation of SF channel was not likely to have a major influence on neuronal RF size. Nevertheless, these observations combined to show that the pRFs clustered around the fovea under the high SF stimulation, leading to a dense distribution in which the pairwise pRF distance was reduced. In contrast, when the low SF channels dominated over the high SF channels, the pRFs instead became scattered, forming a sparser and wider-spread distribution. Importantly, these findings further raised the possibility of a global neuronal RF shift associated with the selective enhancement of different SF channels, where the pRFs appeared to be ‘dragged’ to the fovea by the high SF stimulation, and meanwhile were ‘pushed’ away from the fovea under the low SF stimulation. In this case, one would predict consistent pRF position changes from the high to the low SF mapping condition towards the fovea regardless of their locations, and consequently, altered representation of visual space spanned by the voxel-wise pRFs.

We then investigated the possibility of global-scale pRF shifts by explicitly examining these two key predictions in two complementary analyses. Because the pRF profiles in all four quadrants of the visual field showed a similar pattern, the data was collapsed across both horizontal and vertical meridians into a single quadrant to maximize sensitivity[32]. First, to investigate the consistency of pRF position changes, we partitioned the combined quadrant with a 6-by-6 grid and binned the voxel-wise pRFs in the each SF mapping condition with respect to the grid, which yielded two binning schemes. For each binning scheme, a positional change vector from the mean pRF center coordinate of the binned voxels in the low SF condition to that in the high SF condition was computed in each grid cell, which was weighted by the percentage of the binned voxels. Finally, the two weighted positional change vectors were averaged, and the averaged vector was plotted in each cell to characterize the local pRF position change from the low to the high SF mapping condition (**Figure 3d**). We found that the averaged positional change vector pointed at direction of the fovea in almost every grid cell, indicating remarkable consistency in local pRF position changes across different parts of the visual space. To statistically evaluate this observation, we calculated the angular deviation between the direction of the averaged positional change vector and the direction of the fovea in each grid cell (see Methods for details). As such, a deviation value of 0° indicates a change in the pRF position from the low to the high SF mapping condition directly towards the fovea. As shown in **Figure 3e**, the distribution of angular deviations across different grid cells centered around 0° (mean phase: -0.38°; *V*-test: $v = 25.20$, $p < 10^{-8}$). Second, to investigate whether and how the representation of visual space was altered by the enhancement of different SF channels, we reconstructed a map of pRF density by calculating the moving average of the map of the voxel-wise pRF locations for each SF mapping condition, and computed the difference between the two maps (Figure 3f). Interestingly, we identified a clear gradient from the peripheral to the foveal region in the difference map from which two oppositely signed clusters emerged (non-parametric cluster-based permutation test: $p < 0.05$, multiple comparisons corrected). While the positive cluster indicated concentration of pRFs in the vicinity of fovea under the high SF stimulation (cluster peak eccentricity: 0.59°), the negative cluster in the periphery instead signaled diffusion of pRFs at higher eccentricities (cluster peak eccentricity: 8.43°). Together, these observations were in excellent agreement with a convergent shift of pRFs with respect to the fovea when the high SF channels were driven and a divergent shift when the

low SF channels were driven, respectively.

Perceptual distortions linked with global pRF shifts via SF channel modulation

The preceding results have demonstrated that modulation of SF channels, either by adaptation or stimulation, caused both a systematic distortion in spatial property perception and a global-scale shift of population receptive fields in V1. However, it remained unknown whether and how the pRF changes were functionally linked with the perceptual distortions via modulation of SF channels. To address this question, we built a model of the population SF response based on the behavioral data from an out-scanner psychophysical experiment, and used the same model to predict the pRF changes measured from the fMRI data. In the psychophysical experiment, subjects were presented with two oblique bars on the background of two square-shaped checkerboards with different SFs. Subjects discriminated the lengths of the two bar stimuli, with the one superimposed on the low SF checkerboard (0.25 cycles/°) as the invariant standard stimulus and the other on the high SF checkerboard (1.5 cycles/°) as the probe that was varied by a staircase procedure (**Figure 4a** left). To simulate the enhancement of SF channels during pRF mapping, the checkerboard backgrounds were first solely presented for 1900ms and were then superimposed by the task stimuli for 100ms, such that the total duration of the checkerboards was identical to the duration of pRF mapping stimulus in each step. Two Gaussian white noise masks were presented for 2s subsequently after the task stimuli to restore the original population SF response (**Figure 4a** right). Similar as in the adaptation experiments, we identified the PSE of the standard stimulus from the fitted psychometric curve. As a result, the psychometric function was displaced to the left of the physical length of the standard stimulus, i.e. 3.5^{o} (**Figure 4b** left), indicating a clear perceptual distortion in which the bar on the high SF background appeared longer while the bar on the low SF background appeared shorter. The perceptual distortion effect was measured as the relative difference between the PSE and the standard stimulus in percentage units, which was significantly above zero (mean ± S.E.M.: 3.8 ± 0.79%, one-sample *t*-test: $t_{17} = 6.14$, $p < 10^{-4}$) (**Figure 4b** right). These results showed that SF channel modulation influenced length perception in the same way as the perception of other fundamental spatial properties, which further reinforced our findings from the psychophysical adaptation experiments and were strongly advocated the generalizability of our theory. Moreover, we correlated the magnitude of perceptual distortion and the difference in mean

pRF eccentricity and size between the two SF mapping conditions (low SF mapping - high SF mapping) (**Figure 4c**). While the perceptual distortion magnitude increased with the pRF eccentricity difference (r = 0.54, p = 0.021), no correlation was observed between the perceptual distortion magnitude and pRF size difference (r = -0.001, p = 0.998). These results thus demonstrated the correspondence between the psychophysical and neural observations. In addition, they also argued against any possible connection between perceptual distortions induced by SF channels reweighting and pRF size changes.

We modeled the population SF response elicited by the checkerboard backgrounds as the weighted sum of multiple differently-tuned SF channels characterized as logarithmic Gaussian functions peaked at different SFs (0.2 to 1.7 cycles/° in steps of 0.25 cycles/°). The channel weights encode the underlying contribution of each SF channel to the population response, hence the enhancement and the attenuation of a specific SF channel is coded by a change of its weight, respectively. In addition, the full width at half maxima (FWHM) of the logarithmic Gaussian functions encode the selectivity or tuning width of the SF channels (**Figure 4d**). We predicted subject's trial-by-trial response by comparing the perceived length of the standard stimulus and the probe in each trial. Specifically, the perceived length of each bar stimulus was determined as the number of metric units it spanned, which was calculated by multiplying its physical length by the preferred SF of the population response (i.e. the inverse of the metric's grain size) elicited by its background. Using a reverse correlation approach, we obtained the population SF response for each checkerboard background by finding the channel weights and widths that best predicted the trial-by-trial responses. To verify the biological plausibility of the model, we reconstructed and compared the population SF responses elicited by the two checkerboard backgrounds (**Figure 4e**). Under the high SF stimulation, the peak of the population SF response was located at around 0.83 cycles/°. In contrast, the peak of the population SF response in turn shifted to 0.78 cycles/° in the low SF condition. These findings showed that enhancing the high versus the low SF channels produced a marked difference in the population SF response (paired t-test: t_{17} = 4.98, p < 0.001), as we expected. We then examined the validity of the model by comparing the predicted response with the empirical response across different trials. As shown in **Figure 4f**, the cross-validated prediction accuracy significantly exceeded the chance-level performance (50%) by a large margin

for all subjects (mean accuracy ± S.E.M. = 76.07 ± 1.22%, permutation test: $p < 0.001$, Bonferroni corrected for multiple comparisons), indicating that our model was able to predict subjects' trial-by-trial responses with exceptional accuracy.

Having demonstrated both the plausibility and the validity of our model, we then examined how modulation of SF channels, as measured by the model, could be connected to changes in subjects' spatial property perception. To measure the magnitude of SF channel modulation effect, we computed the difference in the population's preferred SF between the high and the low SF background conditions (high SF background – low SF background). As it turned out, there was a significant positive correlation between the magnitude of perceptual distortion and the magnitude of SF channel modulation ($r = 0.55$, $p = 0.017$), which revealed a strong association between the perceived length and the reweighting of SF channels (**Figure 4g**). Finally, we examined the link between SF channel modulation and pRF dynamics by correlating the magnitude of SF channel enhancement effect with the difference in pRF properties between the two SF mapping conditions (Figure 4h). We found that the magnitude of SF channel modulation positively correlated with pRF eccentricity difference ($r = 0.57$, $p = 0.014$) but not with pRF size difference ($r = -0.29$, $p = 0.242$), indicating that selective modulation of SF channels was functionally coupled with pRF position shifts but had little or none influence on pRF sizes. As an additional note, we repeated the same analyses involving neural data using two different criteria (20% and 30%) for voxel inclusion and we replicated all the findings as reported above (see supplementary information for details). Taken together, these results established a comprehensive three-way connection between distortions of spatial property perception, global-scale pRF position shift and SF channel modulation.

Discussion

In the current study, we combined the use of psychophysics, fMRI and computational modeling to investigate the neural correlates of visual metric. We found that selective inhibition of the high and the low SF channels via psychophysical adaptation could cause shrinkage and inflation of various perceived spatial properties, respectively. Moreover, enhancement of the high and the low SF channels elicited a global-scale difference in V1 neuronal RF locations, where V1 pRFs in

different parts of the visual space consistently concentrated towards the fovea under the high SF stimulation yet became diffused under the low SF stimulation. Using computational modeling, we further found that both the perceptual distortion and the systematic pRF displacement could be predicted by the changes in the joint responses of different SF channels. Our results thus provided converging evidence that the SF channels in early visual areas serve as the fundamental neural apparatus for signaling the visual metric, which commonly governs the perception of different spatial properties.

Our findings demonstrated that the adaptable visual metric is mediated by the flexible weighting of SF channels, which have important functional bearings on a plethora of visual phenomena. First, they provided a novel mechanistic account for various illusions in which the perceived stimulus size departed from the physical input. For example, in Ebbinghaus-like figures, the apparent size of the central object can be inflated, independently of other factors, by decreasing the size or the proximity of the surrounding contextual objects[2,35,36]. Moreover, in Ponzo-like illusions, visual objects located among aggregated local image elements (e.g. multiple converging lines that form the linear perspective) or embedded within dense texture patterns are typically perceived larger even when no other pictorial depth cues are available[37–39]. Common in all these scenarios is the enhancement of high SF channels corresponding to the image region containing the target object, which, according to our hypotheses, reduces the metric's grain size and recruits more metric units for the perceptual representation. In line with this idea, we found a causal relationship between reweighting of SF channels (either by adaptation or presentation of corresponding SF backgrounds) and distortions in perceived separation, length and shape. Our findings therefore extended the previous findings by disclosing a novel SF-based scaling mechanism beyond the size perception. Second, the visual metric based on SF channels is also reflected in the functional division of labor between global versus local processing. Our hypotheses suggested that local edges are optimally resolved by the fine-grained metric unit signaled by high SF channels, while the continuous, global-scale luminance variations are best captured by the coarse-grained metric unit signaled by low SF channels. Enhancing observer's sensitivity to low and high SFs therefore recalibrates the metric by reweighting the corresponding SF channels, which promotes effective processing of the global and local stimulus features, as indicated by a well-established line of psychophysical

evidence[40–43]. Third, the idea of visual metric based on SF channel also echoes the findings that density is estimated from the relative response of the high- and the low-SF channels. Tipping the balance between the high- and the low-SF channels alters the perception of density, as the high SF channels are crucial for estimating the number of objects, whereas the low SF channels relate more to the spatial extent occupied by the object ensemble[10,44]. Together, our findings combined to show that SF sensitivity might serve as a critical limiting factor in various visual tasks and are suggestive of a unified mechanism of cognitive scaling that is commonly operative in multiple levels of perceptual processing.

Notably, our findings are reminiscent of the attentional modulation of both neuronal RF profiles and appearance of spatial features. Pioneering neural findings have shown that attention attracts both individual neuronal RFs[28,45–47] and the population-level spatial tuning preference towards the attentional focus in multiple visual areas[29]. Moreover, the systematic RF shifts are thought to mediate the perceptual changes induced by attention in various psychophysical tasks[48–50]. While the relationship between changes in perceived spatial property information and neuronal RF profiles has been verified by linking the behavioral and neurophysiological effects of attention[23], it has not been systematically examined outside the domain of attention studies. The central challenge probably lies in the difficulty of linking the neural measures of RF change to the psychophysical measures of perception, as experimental stimuli and paradigm usually differ between behavioral and neural studies. Our study shed light on this issue by driving the SF channels in a similar manner between the fMRI and the psychophysical experiments. We managed to identify a global-scale displacement of population receptive fields in V1 under different SF stimulation conditions, which was paralleled by distortions of various perceived spatial properties induced by reweighting of SF channels. Moreover, benefitting from the computational model of population SF response, we showed that, for the first time, SF sensitivity difference could be quantitatively translated into the magnitude of both global pRF shifts and distortions in perceived spatial properties. These findings bridged the critical gap that precludes a comprehensive three-way connection between population SF responses, RF position profiles and spatial property perception. Interestingly, however, we did not observe any systematic changes in pRF size between different SF stimulation conditions, nor did we find any quantitative relationship between

pRF size changes and the perceptual distortions. Since the capacity of integrating sensory information of a visual neuron is determined by the size of its RF[34,51,52], a change in RF size might therefore reflect an adjustment in the spatial range over which sensory information is integrated. However, this is not required when no visual objects were displayed and subjects were required to maintain central fixation, as in our case. As a final note, concentration of neuronal RFs promotes an attention-centered spatial code in which the locations of all visual stimuli are represented relative to the attentional focus[53]. This provides a possible explanation for the RF shifts centered at the task-relevant object[15,54], as well as the RF shifts centered at fixation (i.e. the focus of overt attention) in the absence of visual objects, as observed in our study.

Although our findings are based on the pRF dynamics in V1, this is not to say, however, that the scaling of perceived spatial property information is solely mediated by the earliest cortical processing stage. While our findings have demonstrated systematic changes in neuronal RF profiles induce by modulation of SF sensitivity, it is unlikely that all forms of perceptual distortions in complex visual scenes commonly originate from reweighting of SF channels in early visual areas. First, in natural viewing environments, the retinal images projected by visual objects are transformed to retain object constancy in accordance with a multitude of contextual cues to depth information and 3-D surface configuration. These cues are extracted and processed in the cortical network comprised of higher-order extrastriate visual areas and parietal–occipital junctions[55–57]. Second, it was also found that the spacing between neighboring visual objects is influenced by the contextual motion, which is processed in hV5/MT+[58–60]. These findings combine to suggest that the source of the perceptual deviations in spatial property information in complex visual scenes could be located elsewhere beyond early visual areas. In this sense, we proposed a recurrent mechanism underlying the cognitive scaling of the perceived spatial property information that centers at V1. On one hand, processing of the incoming visuospatial signals is constrained by the visual metric's grain size signaled by the population SF response, which could be biased by feedback signals from higher-order brain areas. On the other hand, local enhancement of specific SF signals might facilitate the interpretation of visual contexts in a way that is more consistent with the visual metric signal, as reflected in the common technique of inducing depth perception on 2-D images by visual artists[61]. The terminal of this ongoing

recurrent, interactive process is located in V1, where the final perceptual outcome is directly read out from the spatial profile of visual neuron RFs.

In summary, we revealed, for the first time, the neural apparatus for signaling the visual metric by establishing a comprehensive three-way connection between SF channel modulation, RF position shifts and distortion in spatial property perception. These findings unraveled a unified cognitive scaling mechanism that commonly governs the perception of various spatial property information.

Methods

Subjects

A total of 37 students (22 females, 18-26 years old) were recruited as subjects from South China Normal University. 22 subjects participated in the psychophysical adaptation experiments (11 for each task) and there were 18 subjects in the fMRI and the out-scanner psychophysical experiments (three of them also participated in the psychophysical adaptation experiments). All subjects had normal or corrected-to-normal vision and reported no known neurological or psychiatric disorders. All subjects were naïve to the purpose of the study. Written informed consents were collected individually before the experiments. The experimental protocol was approved by the human subject review committee in South China Normal University.

Psychophysical adaptation experiments

Visual stimuli were generated using Psychophysics Toolbox 3 (http://psychtoolbox.org/) and were displayed on an LCD monitor (refresh rate: 60 Hz, resolution: 1920×1080 pixels, background luminance: 13.44 cd/m^2). The adaptors were two full-contrast, phase-randomized sinusoidal gratings subtending an area of 10°×10°, which were simultaneously presented in the two visual fields. The centers of the adaptors aligned with the fixation on the horizontal meridian at an eccentricity of 10°. The spatial frequency of one adaptor (the fixed adaptor) was fixed at 0.25 cycles/°, while the SF of the other adaptor (the matching adaptor) could be one of the four SF values (0.25, 0.75, 1 and 1.5 cycles/°). Both adaptors had the same orientation, and were drifting in synchrony at a constant speed of 2.5°/s with their directions reversed every 5 seconds. The task stimuli for separation judgment were comprised of two pairs of horizontally aligned dots (dot size:

0.4°). For shape judgment, the task stimuli were two ellipses with vertical major axes (length: 8.75°). The eccentricities of the task stimuli were the same as that of the adaptors. The task stimulus ipsilateral to the fixed adaptor always served as the standard stimulus (dot pair separation: 3.5°; ellipse aspect ratio: 5/2), while the task stimulus in the contralateral visual field was the probe that varied in separation or aspect ratio.

The psychophysical adaptation experiments were carried out in a behavioral test room. Subjects viewed the stimuli at a distance of 80cm with their heads rested on a chin rest. Each subject completed a total of 12 blocks for each of the four matching adaptors. In half of the blocks, the fixed adaptor was presented in the left visual field. Each block began with a 30-sec pre-adaptation. In a trial, the task stimuli were presented for 100ms immediately after a 5-sec topping-up adaptation. Subjects performed a two-alternative forced choice (AFC) task in which they indicated which of the two pairs of dot appeared more separated (separation judgment) or which of the two ellipses appeared more elongated (shape judgment). A staircase procedure was employed in which the separation or the aspect ratio of the probe was continuously adjusted on a trial-by-trial basis. For each observer, the behavioral data was fitted with a cumulative Gaussian function weighted by the number of trials for each dot separation or aspect ratio using PSignifit 3.0[62], which yielded the psychometric curves describing the percentage of the trials in which the probe was chosen over the standard stimulus. The point of subjective equality (PSE) of the standard stimulus was obtained as the point corresponding to 50% probability from the psychometric curve. For each matching adaptor, the magnitude of the adaptation effect was measured as the relative difference between the perceived and the physical property (dot separation or aspect ratio) of the standard stimuli in percentage units:

$$\text{Adaptation effect} = \frac{\text{PSE} - \text{Standard stimulus}}{\text{Standard stimulus}} \times 100\%$$

pRF mapping

MRI data was acquired with a 3T Siemens MAGNETOM Prisma scanner at the MRI center in South China Normal University with a 20-channel phase-array head coil. Visual stimuli were back-projected via a video projector (refresh rate: 60 Hz, resolution: 1024×768) onto a screen positioned inside the scanning bore. Subjects viewed the stimuli through a mirror mounted on the

head coil. The viewing distance was 80 cm. BOLD signals were measured using an echo planar imaging sequence (TE: 33 ms, TR: 2000 ms, FOV: 192×192 mm^2, flip angle: 90°). 25 coronal slices orthogonal to the calcarine sulcus were collected in each functional volume, which encompassed the occipital lobe and a small portion of the cerebellum (slice thickness: 2 mm, gap: 0 mm). A T1-weighted high-resolution 3D structural image was acquired at the beginning of the scanning using the 3D-MPRAGE sequence.

The functional scanning was comprised of 4 separate runs for retinotopic mapping and 8 runs for pRF model estimation. In the retinotopic mapping runs, subjects were presented with a rotating wedge and an expanding/contracting ring that created traveling waves of neural activity in visual cortex. The stimuli for pRF mapping had two SFs (high SF: 1.5 cycles/°; low SF: 0.25 cycles/°). For each SF condition, the hemodynamic response function (HRF) in V1 was first measured for each subject in a single run containing 12 trials. In each trial, a full-contrast flickering checkerboard stimulus subtending an area of 16°-by-16° centered at fixation point was presented for 2s, which was then followed by a 30-s blank interval. The HRF was obtained by fitting the convolution of a 6-parameter double-gamma function with a 2-s boxcar function to the BOLD response elicited by the checkerboard. PRFs were mapped with a full-contrast flickering checkered bar (width: 2°) oriented either horizontally or vertically, which moved in two opposite directions (up-down, left-right). The bar swept through the visual field in 32s (step size: 0.875°), with every two sweeps followed by a 12-s blank screen. Each pRF mapping run lasted for 152s and there were three pRF mapping runs for each SF condition. Throughout the entire scanning session, subjects performed a color change detection task on the central fixation point to maintain fixation and to control attention.

fMRI data was preprocessed using SPM 12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) and custom scripts written in MATLAB 2019 (The MathWorks). The first 6s of the BOLD signals were discarded to minimize transient magnetic saturation effects. The functional volumes were corrected for head motion and were co-registered with the anatomical scan. No subjects exhibited excessive head movement (>1 mm in translation or >1° in rotation). The voxel time series were band-pass filtered between 0.015 and 0.1 Hz to remove low-frequency noise and to improve pRF

model fitting efficiency. Retinotopic area V1 was delineated on the reconstructed cortical surface using FreeSurfer (https://surfer.nmr.mgh.harvard.edu/) using the standard phase-encoded method[63]. For both the high and the low SF pRF mapping stimuli, voxel-wise pRF model parameters were estimated using the same coarse-to-fine search method described in Dumoulin and Wandell (2008)[30]. For each voxel, the effective stimulation at each time sample was calculated as the spatial overlap between the mapping stimulus and the population receptive field modeled as an isotropic 2-D Gaussian function. The model has three parameters x_0, y_0 and σ, which describe the center coordinate and the size of the population receptive field, respectively. The predicted BOLD signal time series was then obtained by convolving the HRF with the effective stimulation. Model parameters were adjusted to obtain the best possible fit of the empirical BOLD signal. Only the voxels for which the pRF model could explain at least 25% of the variance of the raw data in both SF mapping conditions were included for further analyses.

pRF model analyses

We analyzed the basic pRF properties corresponding to the two SF mapping conditions by computing the mean pRF eccentricity and the mean pRF size of all voxels for each subject. The average distance between each pair of pRF centers (inter-pRF distance) was also calculated. Each of the three indices was then statistically compared across all subjects between the two conditions. To further resolve the potential difference in pRF size between the two SF mapping conditions, we quantified the relationship between the pRF size and the pRF eccentricity. For each subject, we binned the voxel-wise pRFs in eccentricity steps of 0.2°, and calculated the average pRF size in each eccentricity bin. Comparison of the slope and the intercept of the fitted lines between the two conditions was done using the non-parametric Wilcoxon signed rank test. To obtain the confidence interval of the size-eccentricity relationship at the group level, a bootstrapping method was employed in which we iteratively sampled the subject's data with replacement, and calculated the average of the slope and intercept in each iteration[64]. This procedure was repeated for 1000 times, yielding a bootstrapped distribution of group-level slope and intercept, respectively. The 95% confidence interval (C.I.) was determined from the slopes and the intercepts corresponding to the 97.5 and the 2.5 percentiles, respectively.

We conducted two analyses to investigate the pRF changes pertaining to different parts of the visual field, and the influence of modulation of SF channels on the spatial profile of pRFs. In both analyses, the voxel-wise pRF data from all subjects were pooled together and were merged into one quadrant (i.e. the first quandrant) in prior. To collapse the data from different quadrants, we converted the pRF centers located in the second, the third and the fourth quadrants to their symmetrical points about the vertical meridian, the fovea and the horizontal meridian, respectively. In the first analysis, we partitioned the merged quadrant into a 6-by-6 grid and binned the voxel-wise pRFs in each SF mapping condition, respectively. In each binning scheme, we identified the voxels whose pRF centers were located within each grid cell, and computed a vector from the average pRF coordinate of these voxels in the low SF mapping condition to that in the high SF mapping condition. This vector was weighted by the percentage of the voxels corresponding to that cell to account for voxel number differences between grid cells. Finally, the two weighted vectors obtained from the two binning schemes were averaged to obtain the positional change vector as a measurement of the local pRF positional difference between the two SF mapping conditions pertaining to each cell. To quantify the direction of local pRF position changes in each grid cell, we defined the empirical cell center as the average pRF coordinate in the low SF mapping condition across the two voxel bins, and calculated the angular deviation between the positional change vector and the direction from the empirical cell center to the fovea. Statistical evaluation of the concentration of angular deviations was conducted using the circular *V* test[65]. In the second analysis, we first reconstructed a map in which each location was indexed by the number of the voxel-wise pRFs whose centers coincided with that location for each SF mapping condition. We then obtained a continuous representation of the pRF density was then obtained by calculating the moving average of the map using a 3.5°-by-3.5° aperture. Finally, we computed the difference in the continuous pRF density representation between the two SF mapping conditions. Statistical assessment of the pRF density difference was conducted using the non-parametric cluster-based permutation analysis[66]. To define the supra-threshold clusters, we obtained the point-by-point cluster-defining threshold for each location using a randomization procedure[67]. In each iteration, we exchanged the pRF data between the two SF mapping conditions in a random subset of subjects and performed the same analysis procedure as described above. The randomization process was repeated for 1000 times to generate a null distribution of the pRF

density difference at each location and we chose the 95th percentile of the null distribution as the cluster-defining threshold for both the empirical and the surrogate data. We identified the supra-threshold cluster with the maximal absolute value in the surrogate pRF density representation in each iteration, which yield the null distribution of cluster-based values from which the statistical threshold ($p < 0.05$, corrected for multiple comparisons) for clusters was obtained.

Out-scanner psychophysical experiment

Outside the scanner, subjects performed an additional psychophysical experiment to measure the distortion in perceived spatial property induced by enhanced SF channel response via presentation of high and low SF backgrounds. The experimental setup was the same as that in the psychophysical adaptation experiments. The task stimuli were two oblique bars (bar width: 0.35°), which were presented on the background of two full contrast, phase-randomized checkerboards of different SFs (1.5 and 0.25 cycles/°). Both checkerboard backgrounds subtended an area of 6°×6°, which were located at an eccentricity of 5° on the horizontal meridian. The bar superimposed on the low SF checkerboard served as the standard stimulus (bar length: 3.5°), while the bar on the high SF background was the probe with variable length. The standard stimulus along with the low SF checkerboard background could appear in either visual field with equal probability. To mimic the SF channel enhancement during the pRF mapping, the checkerboard backgrounds were presented for the same duration of 2000ms as the pRF mapping stimulus in each step, during which the two checkerboards were solely presented for 1900ms before the presentation of the bar stimuli on top of them for 100ms. To minimize potential adaptation or temporal expectation effects, the polarity of the checkerboard stimuli was reversed with 50% probability every 166ms. Subjects were required to indicate which of the two bars appeared longer (length discrimination). Next, two Gaussian white noise patches were presented for 2s after a 1.025-s blank screen to restore the default, unmodulated SF channel responses. The white noise patches were matched with the background stimuli in mean luminance and size. Similarly as in the adaptation experiments, a staircase procedure was employed to find the PSE corresponding to the length of the standard stimulus by adjusting the length of the probe on the high SF background. Each subject completed a total of 8 staircase blocks. We quantified the perceptual distortion effect as the

relative difference between the PSE and the physical length of the standard stimulus in percentage units.

$$\text{Perceptual distortion effect} = \frac{\text{Standard stimulus} - \text{PSE}}{\text{Standard stimulus}} \times 100\%$$

Computational modeling

We characterized the population SF response as the weighted sum of a series of SF channels that are selectively tuned to different SFs, which was parameterized by three vectors comprised of the weight, the preferred SF and the tuning width of each SF channel, respectively:

$$f\left(x|\vec{\boldsymbol{w}}, \vec{\boldsymbol{\mu}}, \vec{\boldsymbol{F}}\right) = \sum_k w_k\, t_k(x|\mu_k, F_k)$$

where x denotes to the stimulus SF. w_k represents the contribution of the k-th SF channel to the population response (i.e. channel weight), whose changes reflect the modulation effect of the corresponding channel. The idealized tuning function of the k-th SF channel was modeled as a logarithmic Gaussian function $t_k(x)$, with the preferred SF and the full width at half maxima (FWHM) denoted as μ_k and F_k, respectively. We considered seven SF channels centered at 0.2 to 1.7 cycles/° in steps of 0.25 cycles/°, hence the idealized tuning function is given as:

$$t_k(x|\mu_k, F_k) = e^{-\frac{[\log(x) - \log(\mu_k)]^2}{2\sigma_k^2}}, \mu_k = 0.2, 0.45, \ldots, 1.7 \text{ cycles}/°$$

where σ_k is computed from μ_k and F_k as:

$$\sigma_k = \frac{1}{\sqrt{2\log 2}} \log \frac{\left(\frac{F_k}{\mu_k}\right) + \sqrt{\left(\frac{F_k}{\mu_k}\right)^2 + 4}}{2}$$

A reverse correlation approach was used to find the vector of channel weight and the vector of tuning width (i.e. FWHM) corresponding to each checkerboard background condition. First, we obtained the population SF response by fitting another logarithmic Gaussian function to the weighted sum of SF channel tuning functions for each checkerboard background condition. Second, we computed the perceived length of each bar by multiplying its physical length by the preferred SF of the corresponding population response. Third, a prediction of subject's response

was derived in each trial by comparing the perceived length of the two bars. Finally, we obtained the channel parameters for each channel that yielded the best fit to the behavioral data by minimizing the error between the predicted and the empirical trial-by-trial responses.

Author contributions

S.Z., C.M., and F.F. designed the experiment. S.Z. performed the experiment. L.M. supervised the study. C.M., S.Z., and F.F. analyzed the data. S.Z., C.M. and F.F. wrote the paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (32471104) to C.M..

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. Illustration of the neural apparatus signaling the visual metric. **Left)** Enhancement of the high SF channels decreased the 'grain size' (i.e. unit area) of the metric (bold line segments), which is coupled with a concentration of neuronal RFs (black circles). The reduced unit area results in the recruitment of more metric units spanned by the physical stimulus, which inflates the perceptual readout of its spatial properties (e.g. length). **Right)** Enhancement of the low SF channels instead causes an exactly opposite effect, where the length perception of the same stimulus is shrunk by the recruitment of fewer metric units (with larger area).

Figure 2. Psychophysical adaptation experiments. ***a.*** Experimental stimuli. **Left)** Illustration of the adapting stimuli. Both the fixed adaptor (indicated by the blue arrowhead) and the matching adaptor (red arrowhead) could appear in either visual field. **Right)** Illustration of the task stimuli. In this example, the standard stimulus of the separation judgment task (upper) is shown in the left visual field, while that of the shape judgment is shown in the right visual field. The standard stimuli were always preceded by the fixed adaptor. ***b.*** Experimental procedure of a staircase block. Subjects indicated the perceptually more separated pair of dots (separation judgment) or the perceptually more elongated ellipse (shape judgment). ***c.*** Psychometric curves corresponding to the four matching adaptors (coded in different colors) of the separation judgment and the shape judgment task, which were obtained by fitting the cumulative Gaussian function to the data pooled from all participants in each condition. The abscissa refers to the paired-dots separation (separation judgment task) or the aspect ratio (shape judgment task) of the probe. The ordinate refers to the percentage of trials in which the probe stimulus was perceived more separated or more elongated. Dotted lines indicate the physical separation or aspect ratio corresponding to the

50% probability that subjects chose the probe over the standard stimulus (i.e. point of subjective equality of the standard stimulus, PSE). Data points are binned for better visualization, with the size of the dot proportional to the number of trial in that bin. ***d.*** Magnitude of adaptation effect as a function of the matching adaptor SF in both tasks. Each point denotes the data from one subject in one matching adaptor condition. Error bars denote one standard error of the mean (S.E.M.) across all subjects. *: $p < 0.05$.

Figure 3. pRF mapping experiment. ***a.*** Stimuli of the high (left) and the low (right) SF mapping conditions. used for pRF mapping. White arrows indicate the possible sweeping directions of the bar stimuli. ***b.*** Comparison of the pRF eccentricity and the pRF size of all V1 voxels between the two SF mapping conditions. Error bars denote one S.E.M. across all subjects. *: $p < 0.05$, n.s.: not significant. ***c.*** Relationship between pRF size and pRF eccentricity in the two SF mapping conditions. The solid lines denote the linear fits across subjects. The shaded bands around the lines indicate the 95% confidence interval estimated using a bootstrapping method. ***d.*** Positional change vectors corresponding to different parts of the visual space. Grey dashed lines denoted the boundaries of the grid cells partitioning the merged quadrant. The blue circles indicate the averaged voxel-wise pRF positions corresponding to the low SF mapping condition, i.e. the empirical cell centers. The red circles denote the average pRF positions corresponding to the high SF mapping condition. The norms of the vectors are scaled for better visualization. ***e.*** Distribution of the angular deviation between the positional change vector and the direction from the empirical cell center to the fovea. Specifically, a value of 0° indicates that the position change oriented directed at the fovea. The thick solid bar indicates the mean of the angular deviations. ***f.*** Spatial profile of the difference in pRF density between the two conditions. White curves indicate the boundary of the significant clusters identified using the non-parametric cluster-based analysis ($p < 0.05$, multiple comparisons corrected).

Figure 4. Results of computational modeling. ***a.*** Design of the out-scanner psychophysical experiment. Models of population SF response were fitted based on the data from this experiment. **Left)** Illustration of task stimuli (upper) and Gaussian white noise masks (lower). The task stimuli are two oblique bars (the standard stimulus and the probe) superimposed on two different SF backgrounds (0.25 and 1.5 cycles/°). The standard stimulus was always superimposed on the low SF background (presented in the right visual field in this example). **Right)** Single trial procedure. Subjects indicated the perceptually longer bar. ***b.*** Psychophysical results. **Left)** the psychometric function fit to the data from all subjects. The dashed line indicates the PSE of the standard stimulus. Data points are binned for better visualization, with the size of the dot proportional to the number of trial in that bin. **Right)** Magnitude of the perceptual distortion effect. Error bar denotes one S.E.M. across subjects. ***c.*** Correlation between the difference in pRF eccentricity and pRF size between the two SF mapping conditions (low SF mapping – high SF mapping) and the magnitude of perceptual distortion measured in the out-scanner psychophysical experiment. ***d.*** Illustration of the population SF response (the black dashed curve) modeled as the logarithmic Gaussian function fitted to the weighted sum of different SF channels (the gray dashed curve). Gray solid curves denote the idealized tuning functions of the SF channels. ***e.*** Averaged population SF response corresponding to the high (red) and the low (blue) SF backgrounds across all subjects. Shaded areas denote one S.E.M. Models were fitted based on the data from the out-scanner

psychophysical experiment. ***f.*** Eight-fold cross-validated accuracy of the model-based prediction of the trial-by-trial behavioral response for each individual subject. Error bars denoted one S.E.M. across the cross-validation folds. The black dotted line indicates chance-level performance. ***g.*** Correlation between the magnitude of SF channel modulation effect (high SF background - low SF background) and the magnitude of perceptual distortion measured in the out-scanner psychophysical experiment. ***h.*** Correlation between the magnitude of SF channel modulation effect and the difference in pRF properties between the two SF mapping conditions. *: $p < 0.05$; n.s.: not significant.

Reference

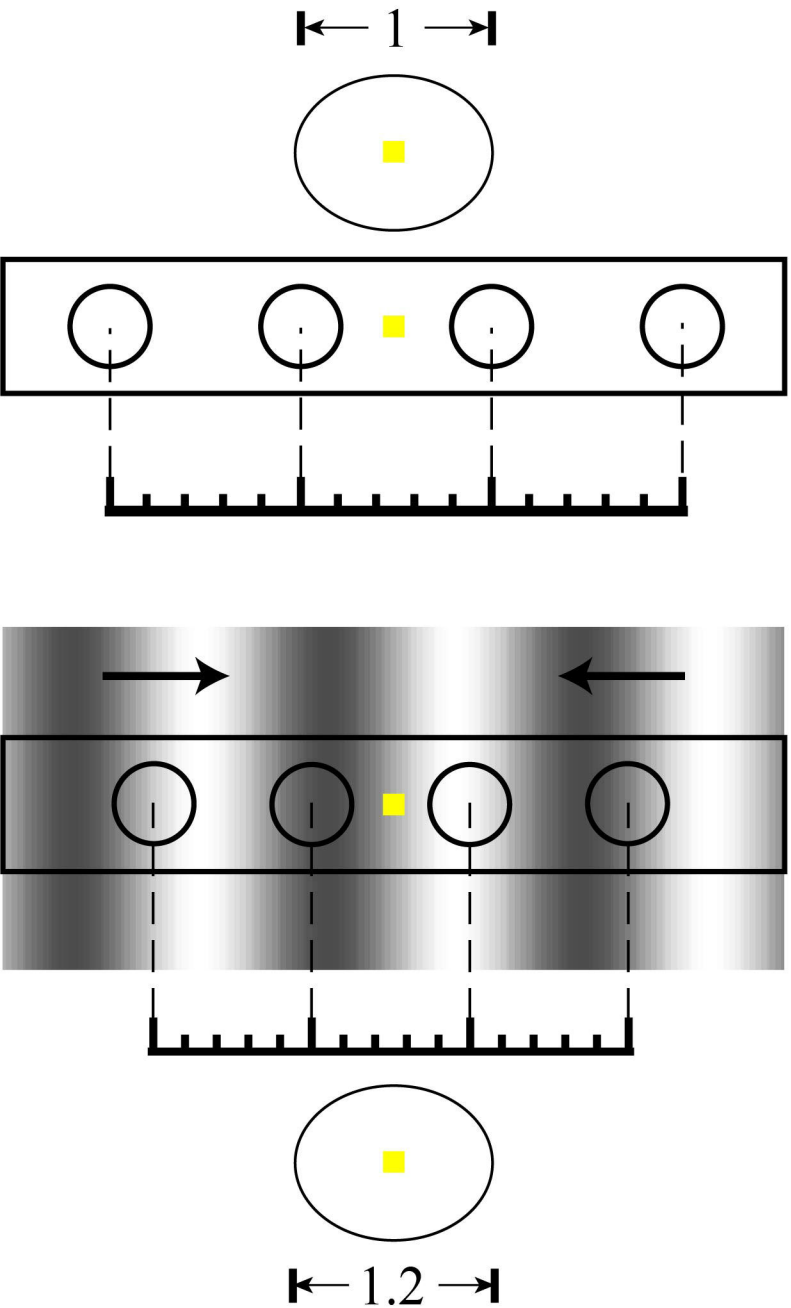
1. De Valois, R. L. & De Valois, K. K. *Spatial Vision.* viii, 381 (Oxford University Press, New York, NY, US, 1988).
2. Kirsch, W. & Kunde, W. On the origin of the Ebbinghaus illusion: The role of figural extent and spatial frequency of stimuli. *Vision Res.* **188**, 193–201 (2021).
3. Takao, S., Watanabe, K. & Cavanagh, P. Dynamic presentation boosts the Ebbinghaus illusion but reduces the Muller-Lyer and orientation contrast illusions. *J. Vis.* **21**, 4 (2021).
4. Urale, P. W. B. & Schwarzkopf, D. S. Effects of cortical distance on the Ebbinghaus and Delboeuf illusions. *PERCEPTION* **52**, 459–483 (2023).
5. Sperandio, I. & Chouinard, P. A. The Mechanisms of Size Constancy. *Multisensory Res.* **28**, 253–283 (2015).
6. Sperandio, I., Chouinard, P. A. & Goodale, M. A. Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nat. Neurosci.* **15**, 540–542 (2012).
7. Fang, F., Boyaci, H., Kersten, D. & Murray, S. O. Attention-Dependent Representation of a Size Illusion in Human V1. *Curr. Biol.* **18**, 1707–1712 (2008).
8. Murray, S. O., Boyaci, H. & Kersten, D. The representation of perceived angular size in human primary visual cortex. *Nat. Neurosci.* **9**, 429–434 (2006).
9. Pooresmaeili, A., Arrighi, R., Biagi, L. & Morrone, M. C. Blood Oxygen Level-Dependent Activation of the Primary Visual Cortex Predicts Size Adaptation Illusion. *J. Neurosci.* **33**, 15999–16008 (2013).
10. Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A. & Morgan, M. J. A common visual metric for approximate number and density. *Proc. Natl. Acad. Sci.* **108**, 19552–19557 (2011).
11. Hisakata, R., Nishida, S. & Johnston, A. An Adaptable Metric Shapes Perceptual Space. *Curr. Biol.* **26**, 1911–1915 (2016).
12. Parkinson, C., Liu, S. & Wheatley, T. A common cortical metric for spatial, temporal, and social distance. *J. Neurosci. Off. J. Soc. Neurosci.* **34**, 1979–1987 (2014).
13. Walsh, V. A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* **7**, 483–488 (2003).
14. MacEvoy, S. P. & Fitzpatrick, D. Visual Physiology: Perceived Size Looms Large. *Curr. Biol.* **16**, R330–R332 (2006).
15. Ni, A. M., Murray, S. O. & Horwitz, G. D. Object-Centered Shifts of Receptive Field Positions in Monkey Primary Visual Cortex. *Curr. Biol.* **24**, 1653–1658 (2014).
16. Kundt, A. Untersuchungen über Augenmaass und optische Täuschungen. *Ann. Phys.* **196**,

118–158 (1863).
17. Landwehr, K. The Oppel-Kundt Illusion and Its Relation to Horizontal-Vertical and Oblique Effects. *PERCEPTION* **50**, 470–478 (2021).
18. Wade, N. J., Todorović, D., Phillips, D. & Lingelbach, B. Johann Joseph Oppel (1855) on Geometrical–Optical Illusions: A Translation and Commentary. *-Percept.* **8**, 2041669517712724 (2017).
19. Hisakata, R. & Kaneko, H. Temporal enhancement of cross-adaptation between density and size perception based on the theory of magnitude. *J. Vis.* **21**, 11 (2021).
20. Petrizzo, I., Anobile, G. & Arrighi, R. Motor adaptation distorts visual space. *Vision Res.* **171**, 31–35 (2020).
21. Rosenholtz, R. Capabilities and Limitations of Peripheral Vision. in *ANNUAL REVIEW OF VISION SCIENCE, VOL 2* (eds. Movshon, J. A. & Wandell, B. A.) vol. 2 437–457 (Annual Reviews, Palo Alto, 2016).
22. Stewart, E. E. M., Valsecchi, M. & Schütz, A. C. A review of interactions between peripheral and foveal vision. *J. Vis.* **20**, 2 (2020).
23. Anton-Erxleben, K. & Carrasco, M. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nat. Rev. Neurosci.* **14**, 188–200 (2013).
24. Fernández, A., Okun, S. & Carrasco, M. Differential Effects of Endogenous and Exogenous Attention on Sensory Tuning. *J. Neurosci.* **42**, 1316–1327 (2022).
25. Jigo, M., Heeger, D. J. & Carrasco, M. An image-computable model of how endogenous and exogenous attention differentially alter visual perception. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2106436118 (2021).
26. Anton-Erxleben, K., Stephan, V. M. & Treue, S. Attention reshapes center-surround receptive field structure in macaque cortical area MT. *Cereb. Cortex N. Y. N 1991* **19**, 2466–2478 (2009).
27. Conner, C. E., Preddie, D. C., Gallant, J. L. & VanEssen, D. C. Spatial attention effects in macaque area V4. *J. Neurosci.* **17**, 3201–3214 (1997).
28. Womelsdorf, T., Anton-Erxleben, K., Pieper, F. & Treue, S. Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat. Neurosci.* **9**, 1156–1160 (2006).
29. Klein, B. P., Harvey, B. M. & Dumoulin, S. O. Attraction of Position Preference by Spatial Attention throughout Human Visual Cortex. *Neuron* **84**, 227–237 (2014).
30. Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *NeuroImage* **39**, 647–660 (2008).
31. Wandell, B. A. & Winawer, J. Computational neuroimaging and population receptive fields. *Trends Cogn. Sci.* **19**, 349–357 (2015).
32. Ho, M.-L. & Schwarzkopf, D. S. The human primary visual cortex (V1) encodes the perceived position of static but not moving objects. *Commun. Biol.* **5**, 181 (2022).
33. Schwarzkopf, D. S., Song, C. & Rees, G. The surface area of human V1 predicts the subjective experience of object size. *Nat. Neurosci.* **14**, 28–30 (2011).
34. Song, C., Schwarzkopf, D. S., Kanai, R. & Rees, G. Neural population tuning links visual cortical anatomy to human visual perception. *Neuron* **85**, 641–656 (2015).
35. Massaro, D. W. & Anderson, N. H. Judgmental model of the Ebbinghaus illusion. *J. Exp. Psychol.* **89**, 147–151 (1971).
36. Roberts, B., Harris, M. G. & Yates, T. A. The Roles of Inducer Size and Distance in the

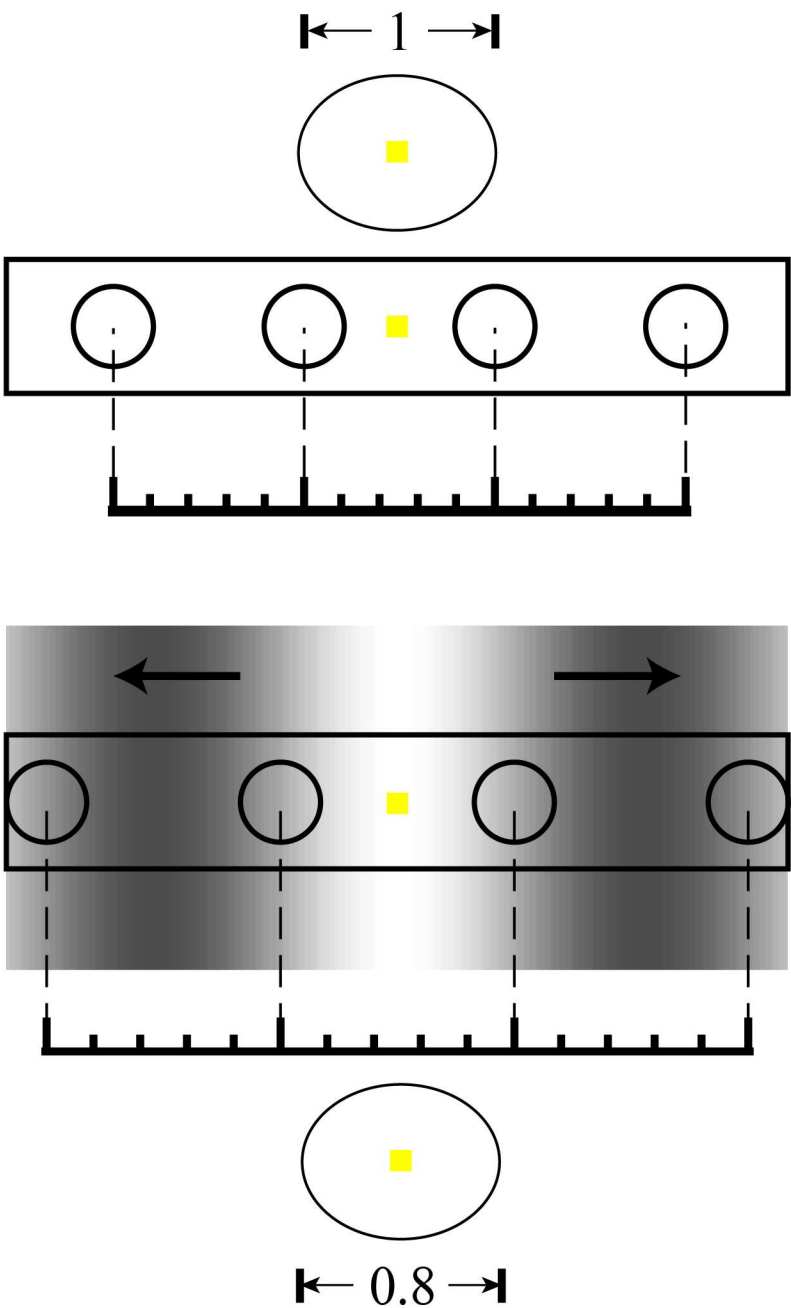
Ebbinghaus Illusion (Titchener Circles). *Perception* **34**, 847–856 (2005).
37. Yildiz, G. Y., Sperandio, I., Kettle, C. & Chouinard, P. A. The contribution of linear perspective cues and texture gradients in the perceptual rescaling of stimuli inside a Ponzo illusion corridor. *PLOS ONE* **14**, e0223583 (2019).
38. Yildiz, G. Y., Sperandio, I., Kettle, C. & Chouinard, P. A. The conceptual understanding of depth rather than the low-level processing of spatial frequencies drives the corridor illusion. *Vision Res.* **181**, 21–31 (2021).
39. Yildiz, G. Y., Sperandio, I., Kettle, C. & Chouinard, P. A. A review on various explanations of Ponzo-like illusions. *Psychon. Bull. Rev.* **29**, 293–320 (2022).
40. Flevaris, A. V., Bentin, S. & Robertson, L. C. Local or Global?: Attentional Selection of Spatial Frequencies Binds Shapes to Hierarchical Levels. *Psychol. Sci.* **21**, 424–431 (2010).
41. Flevaris, A. V., Bentin, S. & Robertson, L. C. Attention to hierarchical level influences attentional selection of spatial scale. *J. Exp. Psychol. Hum. Percept. Perform*. **37**, 12–22 (2011).
42. Flevaris, A. V., Martínez, A. & Hillyard, S. A. Attending to global versus local stimulus features modulates neural processing of low versus high spatial frequencies: an analysis with event-related brain potentials. *Front. Psychol.* **5**, 277 (2014).
43. Flevaris, A. V. & Robertson, L. C. Spatial frequency selection and integration of global and local information in visual processing: A selective review and tribute to Shlomo Bentin. *NEUROPSYCHOLOGIA* **83**, 192–200 (2016).
44. Morgan, M. J., Raphael, S., Tibber, M. S. & Dakin, S. C. A texture-processing model of the 'visual sense of number'. *Proc. R. Soc. B Biol. Sci.* **281**, 20141137 (2014).
45. Connor, C. E., Gallant, J. L., Preddie, D. C. & Van Essen, D. C. Responses in area V4 depend on the spatial relationship between stimulus and attention. *J. Neurophysiol.* **75**, 1306–1308 (1996).
46. Connor, C. E., Preddie, D. C., Gallant, J. L. & Van Essen, D. C. Spatial attention effects in macaque area V4. *J. Neurosci. Off. J. Soc. Neurosci.* **17**, 3201–3214 (1997).
47. Womelsdorf, T., Anton-Erxleben, K. & Treue, S. Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. *J. Neurosci. Off. J. Soc. Neurosci.* **28**, 8934–8944 (2008).
48. Fortenbaugh, F. C., Prinzmetal, W. & Robertson, L. C. Rapid changes in visual-spatial attention distort object shape. *Psychon. Bull. Rev.* **18**, 287–294 (2011).
49. Gobell, J. & Carrasco, M. Attention alters the appearance of spatial frequency and gap size. *Psychol. Sci.* **16**, 644–651 (2005).
50. Suzuki, S. & Cavanagh, P. Focused attention distorts visual space: an attentional repulsion effect. *J. Exp. Psychol. Hum. Percept. Perform*. **23**, 443–463 (1997).
51. Moutsiana, C. *et al.* Cortical idiosyncrasies predict the perception of object size. *Nat. Commun.* **7**, 12110 (2016).
52. Song, C., Schwarzkopf, D. S. & Rees, G. Variability in visual cortex size reflects tradeoff between local orientation sensitivity and global orientation modulation. *Nat. Commun*. **4**, 1–10 (2013).
53. Connor, C. E. Attention: beyond neural response increases. *Nat. Neurosci.* **9**, 1083–1084 (2006).
54. He, D., Mo, C., Wang, Y. & Fang, F. Position shifts of fMRI-based population receptive fields

in human visual cortex induced by Ponzo illusion. *Exp. Brain Res.* **233**, 3535–3541 (2015).

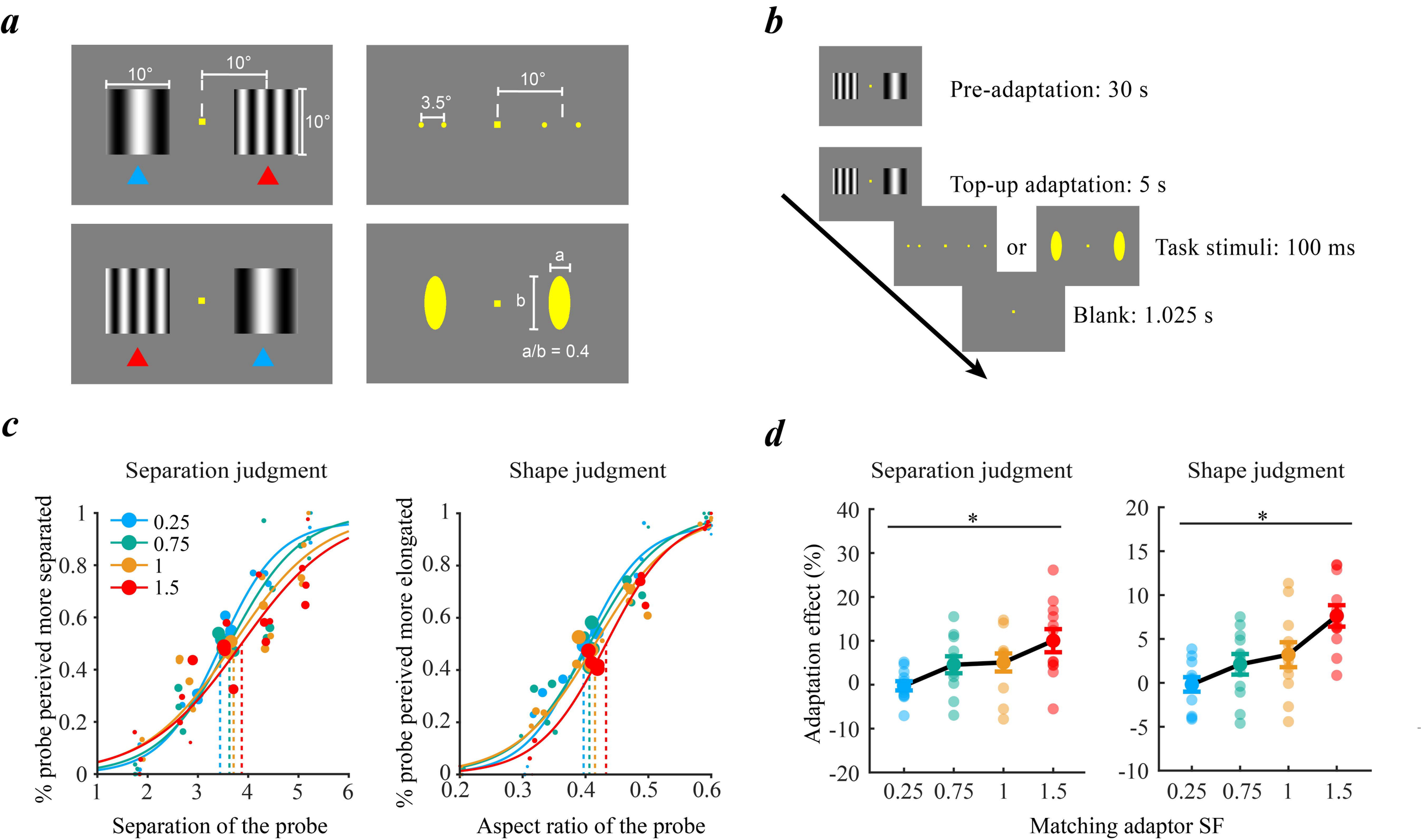
55. Nasr, shahin & Tootell, R. Columnar Neural Mechanisms Underlying Vertically Asymmetric Global Visual Processing. *J. Vis.* **18**, 122 (2018).
56. Ng, A. K. T. *et al.* Ultra-High-Field Neuroimaging Reveals Fine-Scale Processing for 3D Perception. *J. Neurosci.* **41**, 8362–8374 (2021).
57. Tsao, D. Y. *et al.* Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron* **39**, 555–568 (2003).
58. Özkan, M., Anstis, S., ’t Hart, B. M., Wexler, M. & Cavanagh, P. Paradoxical stabilization of relative position in moving frames. *Proc. Natl. Acad. Sci.* **118**, e2102167118 (2021).
59. Ross, J., Morrone, M. C. & Burr, D. C. Compression of visual space before saccades. *Nature* **386**, 598–601 (1997).
60. Zirnsak, M., Steinmetz, N. A., Noudoost, B., Xu, K. Z. & Moore, T. Visual space is compressed in prefrontal cortex before eye movements. *NATURE* **507**, 504-+ (2014).
61. Gregory, R. L. *Eye and Brain: The Psychology of Seeing - Fifth Edition*. (Princeton University Press, 1997). doi:10.2307/j.ctvc77h66.
62. Fründ, I., Haenel, N. V. & Wichmann, F. A. Inference for psychometric functions in the presence of nonstationary behavior. *J. Vis.* **11**, 16 (2011).
63. Engel, S. A., Glover, G. H. & Wandell, B. A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **7**, 181–192 (1997).
64. Mo, C., He, D. & Fang, F. Attention Priority Map of Face Images in Human Early Visual Cortex. *J. Neurosci.* **38**, 149–157 (2018).
65. Berens, P. CircStat: A MATLAB Toolbox for Circular Statistics. *J. Stat. Softw.* **31**, 1–21 (2009).
66. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
67. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. Nat Neurosci 20, 864–871 (2017).

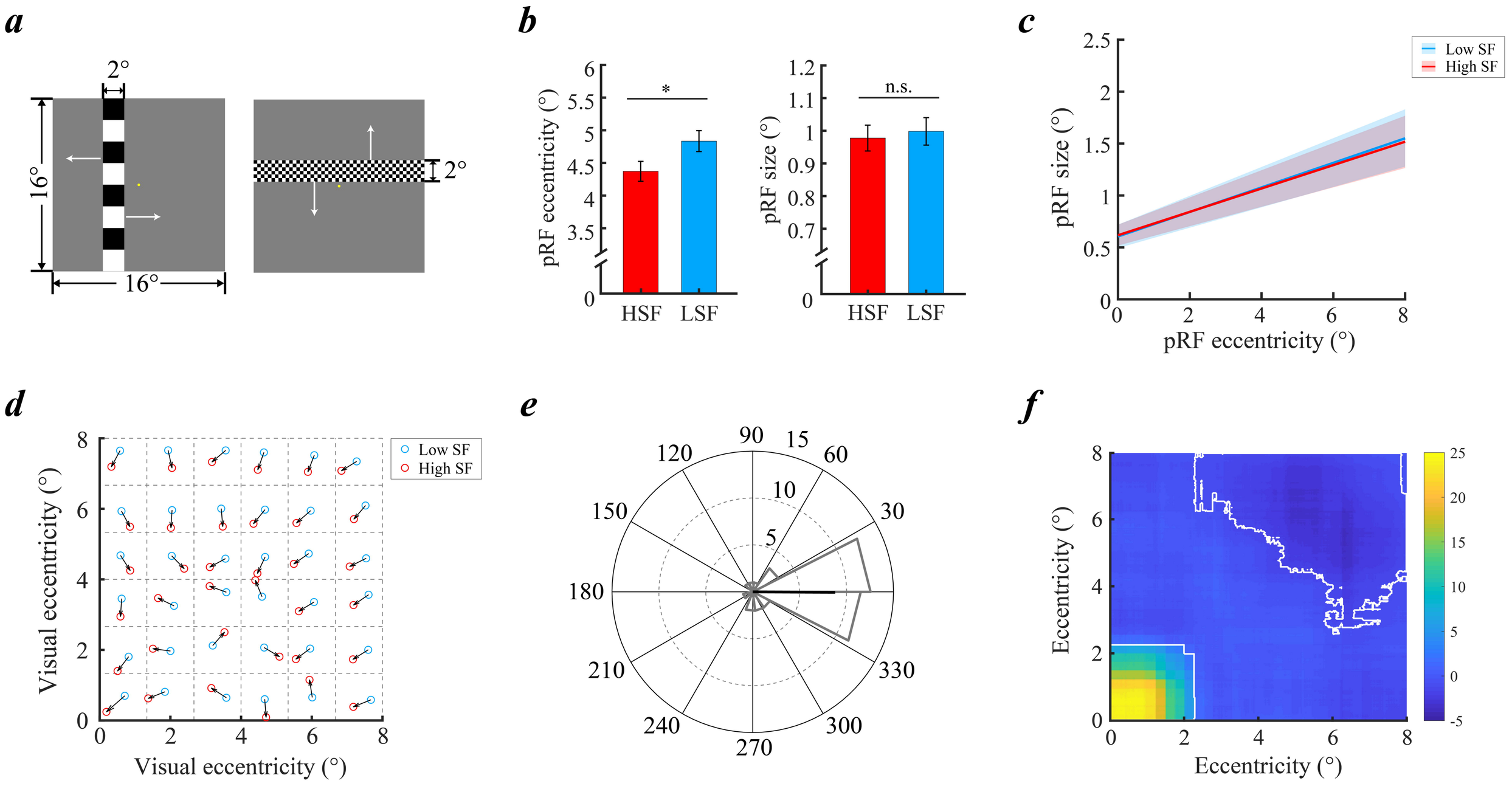


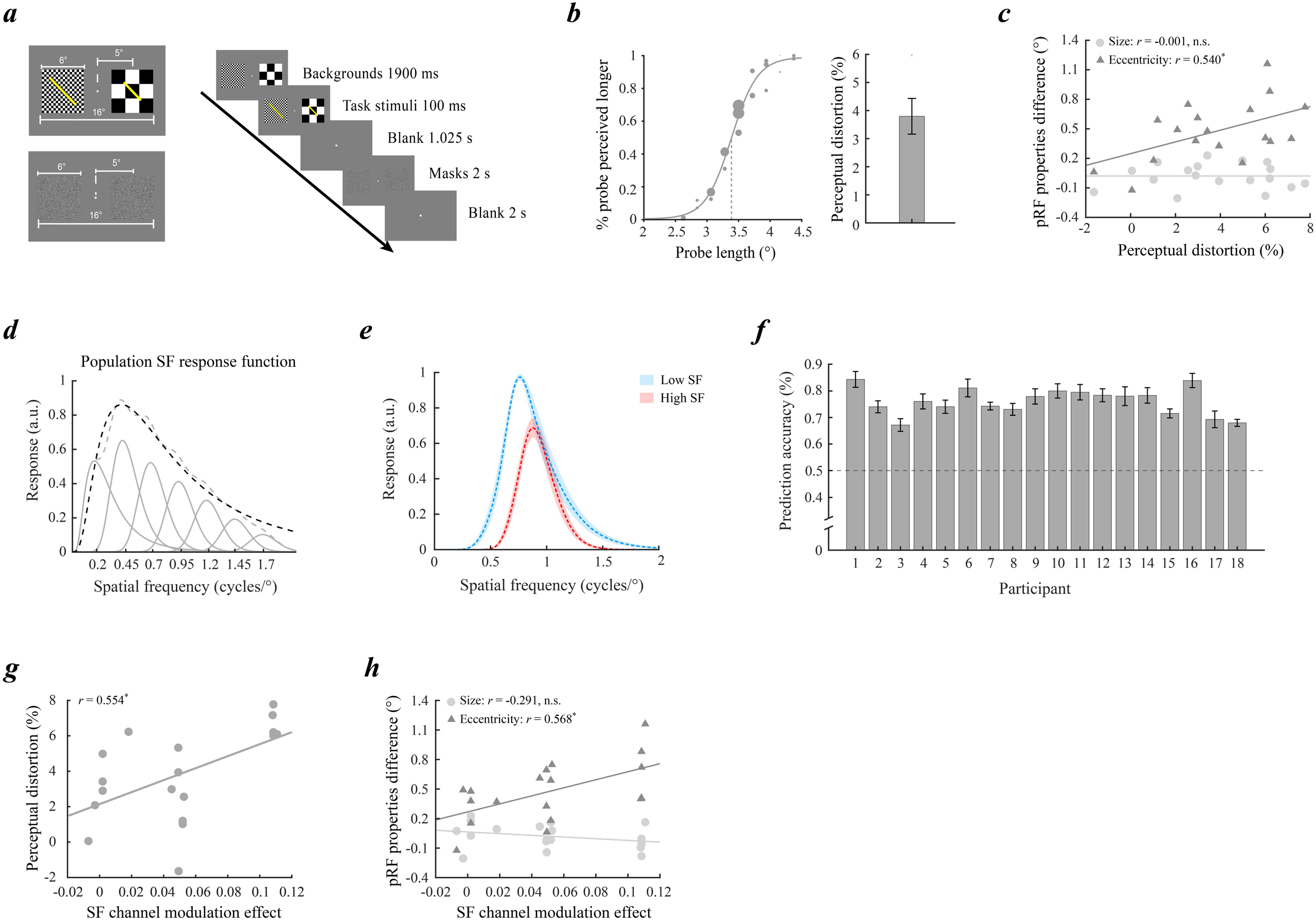
Perceptual inflation

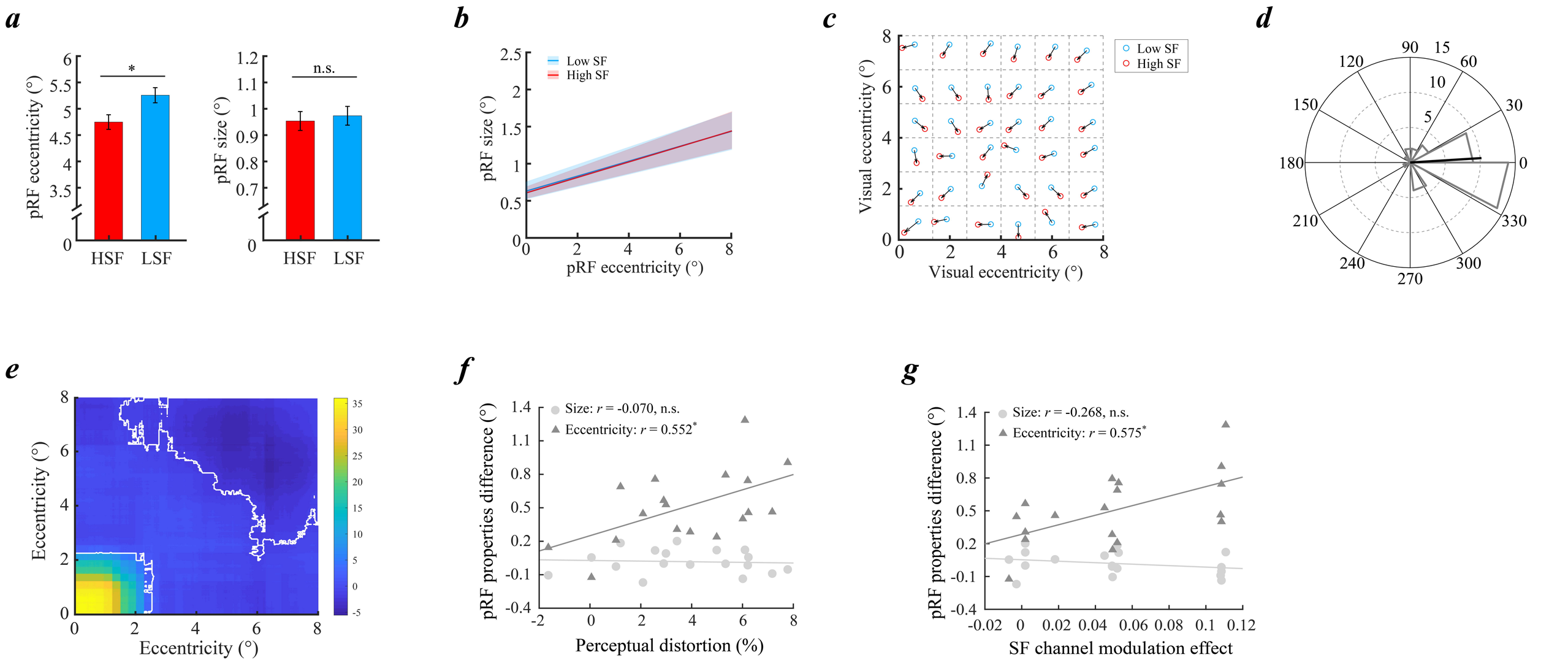


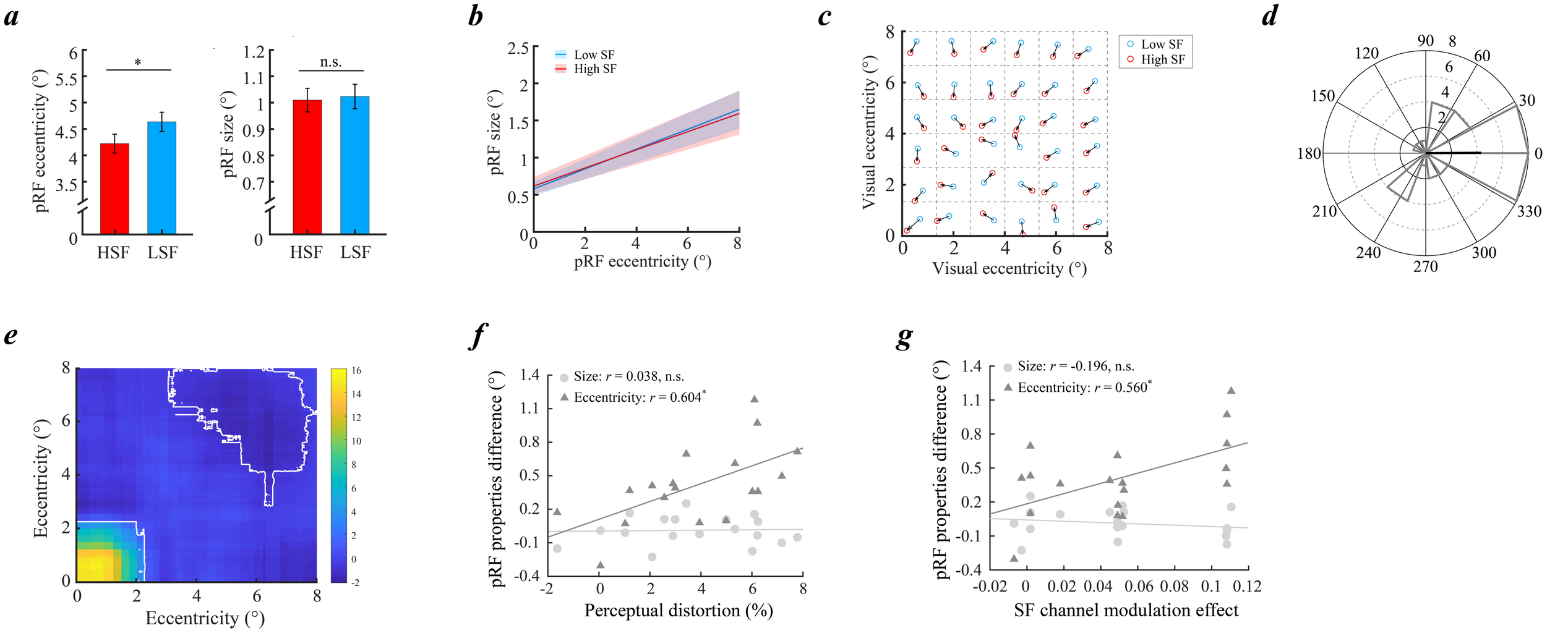
Perceptual shrinkage


a
10°
10°
10°
3.5°
10°
a
b
a/b = 0.4
b
Pre-adaptation: 30 s
Top-up adaptation: 5 s
or
Task stimuli: 100 ms
Blank: 1.025 s
c
Separation judgment
Shape judgment
0.25
0.75
1
1.5
% probe pereived more separated
% probe pereived more elongated
Separation of the probe
Aspect ratio of the probe
d
Separation judgment
Shape judgment
Adaptation effect (%)
*
*
0.25 0.75 1 1.5
Matching adaptor SF


a
2°
16°
16°
2°
b
pRF eccentricity (°)
*
HSF
LSF
pRF size (°)
n.s.
HSF
LSF
c
Low SF
High SF
pRF size (°)
pRF eccentricity (°)
d
Low SF
High SF
Visual eccentricity (°)
Visual eccentricity (°)
e
90
15
60
30
330
300
270
240
210
180
150
120
10
5
f
Eccentricity (°)
Eccentricity (°)


a
6°
5°
16°
Backgrounds 1900 ms
Task stimuli 100 ms
Blank 1.025 s
Masks 2 s
Blank 2 s
b
% probe perceived longer
Probe length (°)
Perceptual distortion (%)
c
Size: r = -0.001, n.s.
Eccentricity: r = 0.540*
pRF properties difference (°)
Perceptual distortion (%)
d
Population SF response function
Response (a.u.)
Spatial frequency (cycles/°)
e
Low SF
High SF
Response (a.u.)
Spatial frequency (cycles/°)
f
Prediction accuracy (%)
Participant
g
r = 0.554*
Perceptual distortion (%)
SF channel modulation effect
h
Size: r = -0.291, n.s.
Eccentricity: r = 0.568*
pRF properties difference (°)
SF channel modulation effect


a
pRF eccentricity (°)
*
HSF
LSF
pRF size (°)
n.s.
b
Low SF
High SF
pRF size (°)
pRF eccentricity (°)
c
Low SF
High SF
Visual eccentricity (°)
Visual eccentricity (°)
d
90
15
60
30
0
330
300
270
240
210
180
150
120
10
5
e
Eccentricity (°)
Eccentricity (°)
f
Size: r = -0.070, n.s.
Eccentricity: r = 0.552*
pRF properties difference (°)
Perceptual distortion (%)
g
Size: r = -0.268, n.s.
Eccentricity: r = 0.575*
pRF properties difference (°)
SF channel modulation effect


a
pRF eccentricity (°)
*
HSF
LSF
pRF size (°)
n.s.
b
Low SF
High SF
pRF size (°)
pRF eccentricity (°)
c
Low SF
High SF
Visual eccentricity (°)
Visual eccentricity (°)
d
e
Eccentricity (°)
Eccentricity (°)
f
Size: r = 0.038, n.s.
Eccentricity: r = 0.604*
pRF properties difference (°)
Perceptual distortion (%)
g
Size: r = -0.196, n.s.
Eccentricity: r = 0.560*
pRF properties difference (°)
SF channel modulation effect