Flexible contextual modulation of naturalistic texture perception in peripheral vision

Daniel Herrera-Esposito[1*], Ruben Coen-Cagli[2‡], Leonel Gomez-Sena[1‡]

1 Laboratorio de Neurociencias, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay
2 Dept. of Systems and Computational Biology and Dominick P. Purpura Dept. of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA

‡These authors contributed equally to this work
*__For correspondance:__ dherrera@fcien.edu.uy

Abstract

Peripheral vision comprises most of our visual field, and is essential in guiding visual behavior. Its characteristic low resolution has been explained by the most influential theory of peripheral vision as the product of representing the visual input using summary-statistics. Despite its success, this account may provide a limited understanding of peripheral vision, because it neglects processes of perceptual grouping and segmentation. To test this hypothesis, we studied how contextual modulation, namely the modulation of the perception of a stimulus by its surrounds, interacts with segmentation in human peripheral vision. We used naturalistic textures, which are directly related to summary-statistics representations. We show that segmentation cues affect contextual modulation, and that this is not captured by the summary-statistics model. We then characterize the effects of different texture statistics on contextual modulation, providing guidance for extending the model, as well as for probing neural mechanisms of peripheral vision.

Introduction

Central and peripheral vision fulfill different roles in visual perception, as reflected by their different information processing capabilities. The most influential model of peripheral visual processing is the summary statistics (SS) model [1, 23, 55, 69], which proposes that the peripheral visual input is represented using SS of the activations of feature detectors (Figure 1), computed over pre-specified regions of the visual field (termed pooling windows) whose size scales linearly with eccentricity. This model fits in the descriptive paradigm of vision as a hierarchical feedforward cascade of visual feature detectors [18, 66], and it is theoretically appealing because replacing a detailed representation of the visual input with a SS results in a significant compression of the visual input. Furthermore, this compression results in a loss of information that could parsimoniously explain the limitations of peripheral vision [69], including the impairment of target identification by surrounding stimuli (visual

crowding [1], often regarded as the most important factor in peripheral vision), as well as phenomena related to visual search [72], scene perception [20, 23], and subjective aspects of visual experience [17]. The SS framework has also been used to explain auditory perception of sound texture [48], suggesting a more general role of SS representations.

Despite providing a solid foundation, it has been hypothesized that phenomena involving segmentation and grouping in peripheral vision escape the standard SS model, and therefore more accurate models of peripheral vision should include recurrent processes of grouping and segmentation [18, 44, 46]. For instance, although the SS model explains crowding as arising from pooling surround statistics with target statistics, it is unclear if it could capture the "uncrowding" produced by surround elements that facilitate target segmentation [18, 44]. Current SS model implementations also fail to capture the peripheral appearance of natural scenes that contain strong grouping and segmentation cues [88]. However, the stimuli and tasks used in this past literature do not allow for a direct test of the hypothesis. For instance, tasks involving fine feature discrimination in simple artificial stimuli (e.g. inclination of a crowded vernier stimulus [18, 44]) or discriminating complex arrangements of features (e.g. scene distortions [88]), cannot be easily related to SS representations, which may be affected by more global properties of a stimulus. Thus, it is not clear whether the SS model could account for segmentation effects [22, 71].

Here we test the hypothesis more directly using naturalistic visual textures. SS representations have long been studied in relation to texture perception [33, 34, 83] (the SS model is also referred to as the texture-tiling model of vision [18]) because textures are statistically defined stimuli. We use naturalistic Portilla-Simoncelli (PS) textures [63], which have been instrumental to the recent success of the SS model [1, 20, 23, 72] and are a useful experimental tool for probing the model. PS textures are defined by a physiologically inspired set of SS that are the basis of the main implementation of the SS model of peripheral vision. This makes it possible to compare directly perception of PS textures to SS model predictions. Furthermore, it has been shown that, different from primary visual cortex (V1), neurons in higher cortical areas V2-V4 are selective for PS statistics [24, 53, 54, 102], offering a framework to relate the SS model and peripheral vision to neural mechanisms. Yet, no studies have addressed how peripheral naturalistic texture perception is affected by image surround (contextual modulation) and by segmentation cues (see [50, 51, 75, 76, 84] for examples with artificial stimuli, [87] for a study with natural images that does not explore segmentation).

Therefore, we use a PS texture discrimination task to study contextual modulation and segmentation in peripheral vision within the framework of the SS model. We evaluate how different texture surrounds affect texture perception, and study the influences of grouping and segmentation cues and of surround structure, as well as the relation between this contextual modulation and crowding.

Our results reveal an important role of segmentation processes in peripheral perception of naturalistic texture, highlight limitations of the feedforward framework of visual processing, and provide guidance for extending the standard SS model to include influences of segmentation and grouping. Our work offers new insights into the neural mechanisms of peripheral vision, and indicates how naturalistic textures could be used for probing them in the visual cortex.

Results

We used a PS texture discrimination task (details in Figure 2 and Methods) to study contextual modulation of texture perception in peripheral vision. PS textures are characterized by a physiologically inspired set of SS, including the correlations between the outputs of V1-like filters selective for orientation and spatial frequency. The corresponding SS model implementation consists of two stages:

the first stage computes the responses of the V1-like filters to the input image, and the second stage evaluates the PS statistics of those filter activations within fixed pooling windows (Figure 1).

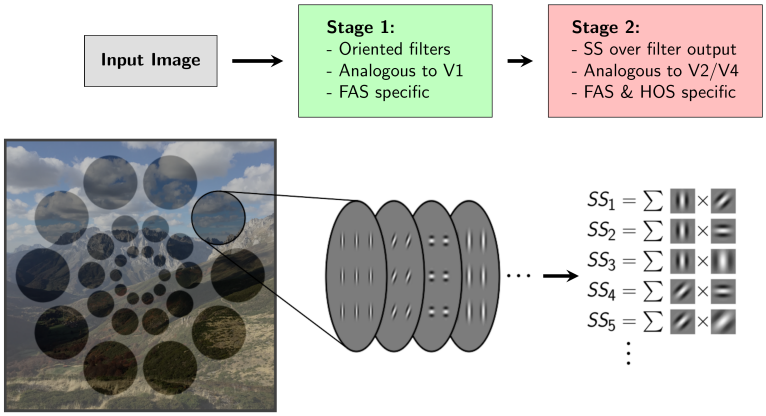

Figure 1: Summary-statistics representation model. Illustration of the main features of the standard SS model, and its relation to physiology and image properties. An input image is first filtered with a bank of oriented V1-like filters, whose activation power is determined by the Fourier amplitude spectrum (FAS) of the image in the pooling region. Then SS are computed over the activations of these filters in fixed pooling windows that tile the visual field. The SS in the second stage are referred to as higher-order statistics (HOS) (in contrast to the statistics contained in the FAS).

The task required discriminating patches of naturalistic PS texture from their corresponding phase-scrambled textures (see Figure 2) in a 3 alternative forced choice design (we refer to the patches to be discriminated as targets). These PS and phase-scrambled texture pairs have the same Fourier amplitude spectrum (FAS), which means they activate the V1-like filters of the SS model with the same average energy, and are thus matched in the first stage of the SS model. Unlike phase-scrambled textures, PS textures also have a more structured distribution of filters activations, corresponding to higher order statistics (HOS) that drive the second stage of the SS model and lead to a more natural appearance [63]. To study contextual modulation, we measured how task performance changed when we introduced uninformative textures surrounding the targets, and compared the effects of different surround parameters.

To evaluate whether our experimental observations could be captured by the feedforward SS model with fixed pooling windows, we used the PS statistics computed over a fixed area centered on the target to perform the task (Figure 2b, Methods), and compared qualitatively the model's discrimination performance to the participants. The radius of the pooling windows was chosen according to Bouma's law of crowding, which says that surrounding stimuli can interfere with target identification when they are within a distance of approximately 0.5 times the target eccentricity [58].

The results are divided into three sections. First, we report the effect of the surround on performance, and its dependence on target-surround grouping or segmentation. Then we explore the relevance of the statistical structure of the surround texture to contextual modulation. Lastly, we study the relation of this contextual interaction to crowding.

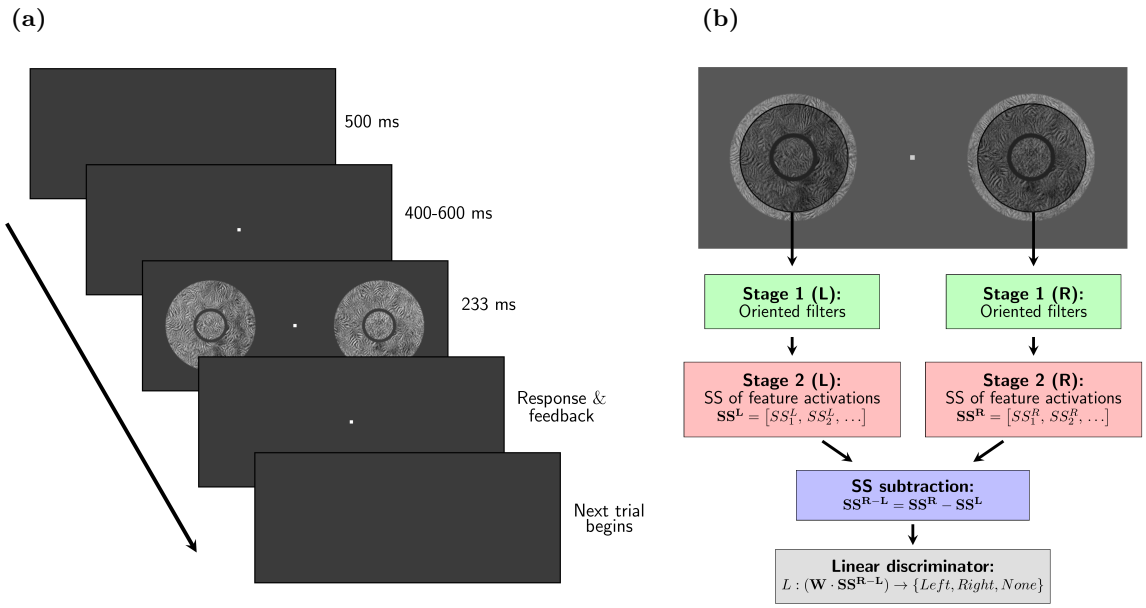

Figure 2: Task design and observer model. **(a)** Two targets centered at 12° to the left and to the right of the fixation point were displayed simultaneously in each trial for 233 ms. Either the left, the right, or none of the targets was sampled from the phase-scrambled texture (with the others sampled from the naturalistic texture), and the participant had to indicate with the arrows where (if) the scrambled texture was present (3 AFC). In most trials, we added uninformative surround textures around the target (in this example surround textures are present, separated from the target by a gap). In any given trial, the two targets always had the same kind of surround. To aid visibility, the size and color of the fixation dot in this image are not the same as in the experiments. **(b)** Diagram showing the architecture of the model observers based on the SS model used to simulate the experiments. The SS of the PS model are computed over circular pooling windows centered on each target (illustrated by the shaded regions). The difference between the SS of the two targets is used to predict the stimulus configuration (i.e. where the phase scrambled target texture is). See Methods for implementation details.

Contextual modulation and grouping

Target-surround grouping, or conversely segmentation, is a major modulator of contextual interactions in vision, especially for crowding [39, 44, 65, 73] in which identification of a target is hindered by nearby stimuli. It has been argued that these segmentation and grouping processes are an important missing component from pooling models of peripheral vision, including the SS model [18, 44, 88]. Despite considerable work using stimuli such as objects, shapes or features (e.g. [36, 44, 46, 74]), our understanding of how grouping processes affect peripheral perception is still incomplete because it is not clear how to relate those tasks that use non-texture stimuli to the SS model, which may be affected by more global stimulus information [71], and whether those results extend to texture processing.

Thus, to better understand the role of grouping and segmentation in the SS model, and how they influence perception of textures, we sought to determine whether contextual modulation of naturalistic texture perception is affected by segmentation or grouping cues.

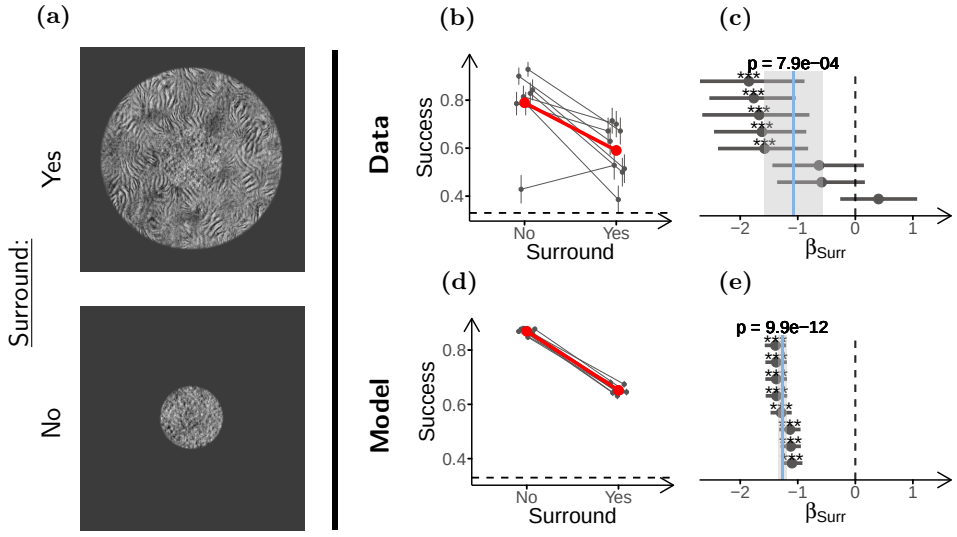

Figure 3: Surround textures impair texture discrimination performance. **(a)** Stimulus configurations used in the experiment (only scrambled targets shown). Top: Target with surround, Bottom: Target without surround. **(b)** Task performances for the two conditions. The gray dots and lines show the performance of individual participants. Vertical lines indicate the ±SD of the estimated performance. Horizontal jitter was applied to aid visualization. The larger red dots show mean performance across participants for each condition. The dashed horizontal line shows chance performance. **(c)** Log-odds-ratios (LOR) between the presence and absence of the surround (β_{Surr}), estimated from the performance data in **(b)**. Each dot shows the LOR for one participant (estimated by fitting a GLM), and the horizontal lines indicate their 95% c.i. Statistical significance of the LOR for the individual participants obtained by the Wald test is indicated as follows: $p < 0.05$ $(*)$, $p < 0.01$ $(**)$, $p < 0.001$ $(***)$. The vertical solid blue line indicates the estimated mean LOR for the population (estimated by fitting a GLMM), and the grey shade indicates its 95% c.i. The p-value of the mean LOR estimate as obtained by likelihood-ratio test (LRT) is indicated above the solid line. The dashed vertical line marks the LOR at which there is no difference between the conditions. **(d)** & **(e)** show the same as **(b)** & **(c)** but for the model observers. Participants (n=8) performed 70 trials in each condition, and model observers (n=8) discriminated 1500 stimuli per condition.

Experiment 1: Target-surround discontinuity reduces contextual modulation

First, we measured whether naturalistic texture perception is affected by contextual modulation. Based on the relevance of contextual modulation for target identification in peripheral vision, we expected task performance to be impaired by surrounding textures. To test this, we presented participants (n=8) with targets in isolation, and with targets surrounded by an uninformative texture ring that was sampled from the same PS texture (Figure 3a).

As expected, task performance was considerably worse for the surrounded targets (Figure 3b). To quantify the effect sizes and test for their statistical significance, we fitted a generalized linear mixed model (GLMM) to the data (which allows to take into account between-participant variability; see Methods and section S1) and reported the log-odds ratio between the conditions (LOR, denoted by β), which is a measure of their difference in success probability (see a guide for converting between the two in section S1). For example, β_{Surr} quantifies the effect of the surround around the target, and $\beta_{Surr}<0$ means that the surround hindered performance. Figure 3c shows that in our experiments the surround strongly impaired performance, and that the effect was statistically significant ($\beta_{Surr} = -1.07$, ci $= [-1.58, -0.57]$, p $= 8 \times 10^{-4}$). This effect was captured by our implementation of the SS model (Figure 3e).

We next tested whether segmentation affects this contextual modulation, and whether the effect can be captured by the feedforward SS model. To probe the effect of segmentation, we presented participants (n=9) with two kinds of stimuli, either with continuous target and surround, or with a visible gap that induced target surround segmentation (Figure 4a). Importantly, the gap was generated by shrinking the target of the continuous stimuli, keeping surround geometry the same in the two conditions. With this design, if pooling regions are constant, the two conditions would have the same amount of surround texture pooled with the target, but in the discontinuous condition there would be less target texture to be integrated (due to the smaller target size). Thus, based on the ratio of informative target texture and uninformative surround texture, the SS model based on fixed local pooling predicts worse performance in the discontinuous than in the continuous condition (Figures 4d,4e; a similar reasoning to that applied in [44]). This is in contrast to what we expect from previous studies using simple stimuli, in which segmentation reduced contextual modulation [36, 44, 65, 74]. Figure 4b shows that performance increased moderately when target and surround were discontinuous ($\beta_{Discont} = 0.62$, ci $= [0.37, 0.87]$, p $= 1 \times 10^{-4}$, Figure 4c). The SS model implementation was thus unable to capture the effect of segmentation (see section S6 for further discussion).

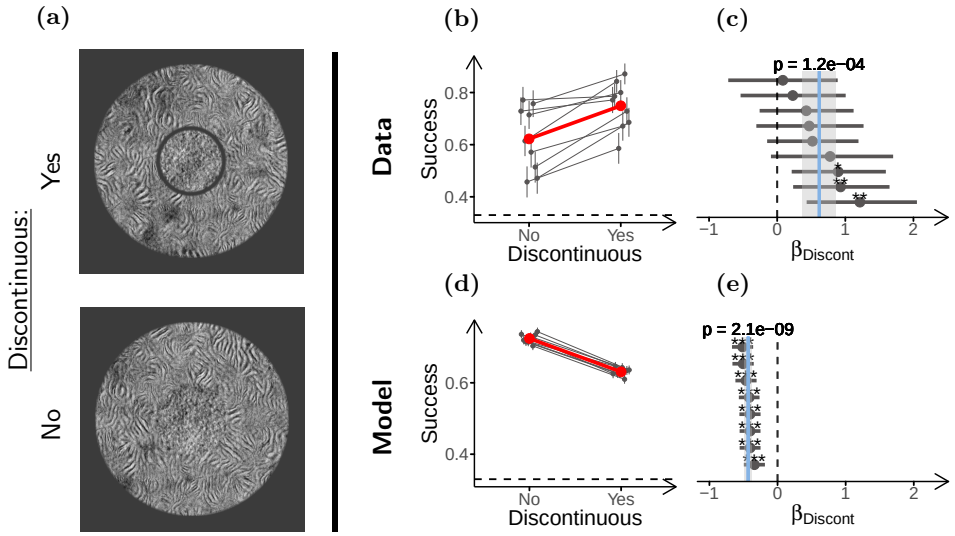

Figure 4: Segmentation reduces contextual modulation. **(a)** Stimulus configurations used in the experiment (only scrambled targets shown). Top: Discontinuous stimulus (smaller target size) Bottom: Continuous stimulus (larger target size). **(b)** Task performance for the two conditions. **(c)** LOR for discontinuity ($\beta_{Discont}$), estimated from the performance data in **(b)**. **(d)** & **(e)** Same as **(b)** & **(c)** but for the simulated observers. Participants (n=9) performed 70 trials in each condition, and model observers (n=8) discriminated 1500 stimuli per condition. Panels **(b)**-**(e)** use the same conventions as Figure 3.

We also found that the observed effect of discontinuity is sensitive to the size of the gap (see section S2, Figure S2), likely because the gap size affects gap visibility, and also the difference in target sizes between the conditions. Also, notice that segmentation did not completely remove contextual modulation, that is, performance was still lower for the discontinuous than for the unsurrounded condition ($\beta_{Surr}^{Discont} = -0.40$, ci $= [-0.68, -0.15]$, p $= 5 \times 10^{-3}$, Figure S3).

Experiment 2: The effect of target-surround discontinuity is mediated by segmentation

We reasoned that the gap between target and surround used to induce segmentation may also affect performance by other mechanisms, such as reducing the uncertainty of target location within the stimulus, or altering the SS of the stimulus in a way that is not captured by the model.

To control for factors related to the gap but not to segmentation, we introduced a different target shape (split-target) consisting of two adjacent semicircles with their straight sides facing outwards (Figure 5a). This split-target shape had approximately the same texture area as the original disk target, and a gap could be introduced around its curved sides, while preserving target-surround continuity on the straight sides of the target.

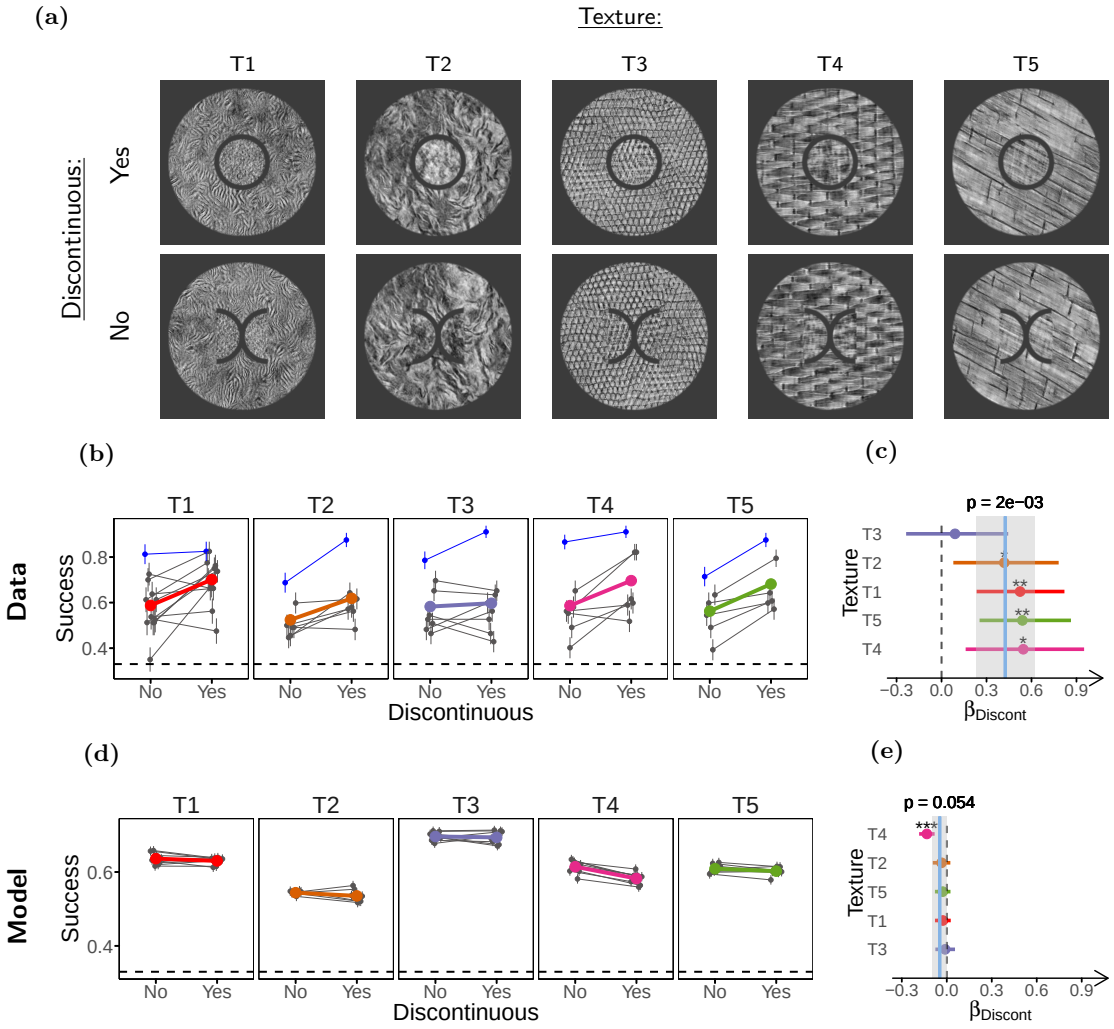

Figure 5: Low level properties of the gap do not explain the effect of discontinuity **(a)** Stimuli used in the experiment. Top: Disk targets (discontinuous), Bottom: Split-targets (continuous). **(b)** Task performance. Each panel shows the results for a different texture, with texture identity indicated above the panel. The layout of each panel is the same as in 3b, except a different color is used to identify the mean performance for each texture. The data of author DH are indicated by the blue symbols. **(c)** LOR for target-surround discontinuity ($\beta_{Discont}$). The colored dots show the LOR obtained by fitting a GLMM for each individual texture (color coded as in **(b)**) and the horizontal lines indicate their 95% c.i. The p-value for the ($\beta_{Discont}$) of each individual texture, estimated by LRT, is indicated as follows: $p < 0.05$ ($*$), $p < 0.01$ ($**$), $p < 0.001$ ($***$). The vertical solid blue line shows the mean $\beta_{Discont}$ across textures and participants estimated by a hierarchical GLMM model using all textures, and the shaded gray region shows its 95% c.i. The p-value for this estimate obtained using LRT is indicated above the line. The dashed vertical line marks the value at which there is no difference between conditions. **(d)** & **(e)** Same as **(b)** & **(c)** but for the model observers. Participants (n=25) completed 40 texture-sessions (see Methods), and performed between 80 and 112 trials per condition. Model observers (n=8) discriminated 1500 trials per condition.

Although the circular targets and the split targets had gaps with similar low level properties, we expected no segmentation for the split-target stimulus because target-surround continuity is maintained. Thus, if the effect observed in the previous section was mediated by segmentation, we should find lower performance for the grouped continuous stimulus (split-target) as compared to the segmented discontinuous stimulus (disk-target). If the effects were mostly due to other factors introduced by the low level properties of the gap, then we would expect similar performance for these two kinds of stimuli.

We presented participants (n=25) with the disk-target and split-target stimuli using 5 different textures to verify that the results did not depend on a specific texture (most participants were shown only some of the textures, see Methods). Participants completed 40 texture-sessions (defined as the unique combination of one participant and one texture) that satisfied the inclusion criterion (see Methods). Consistent with a role of segmentation in contextual modulation of texture perception, performance was moderately worse for the continuous (split-target) than for the discontinuous stimulus ($\beta_{Discont} = 0.42$, ci $= [0.23, 0.62]$, p $= 2 \times 10^{-3}$, Figure 5c). In contrast to this observation, the SS model showed little difference between the stimuli, showing again a failure to capture the segmentation effect (Figure 5e).

We also verified, using additional stimuli for texture T1 (see Figures S4, S5a), that splitting the target had a small and non-significant effect on performance ($\beta_{Split} = -0.09$ ci $= [-0.31, 0.14]$, p $= 0.43$, Figure S5c) validating the use of this experiment to control for low-level gap properties. Furthermore, the estimated effect of the gap after accounting for segmentation was also close to 0 ($\beta_{Gap} = 0.04$, ci $= [-0.15, 0.22]$, p $= 0.71$, Figure S5d), suggesting that effects of the gap other than inducing target-surround segmentation are negligible in our task.

We conclude from these experiments that target-surround segmentation is an important factor in mediating contextual modulation of texture perception, that a discontinuity between target and surround induces segmentation and thus reduces contextual modulation, and that this effect is not observed in the standard SS model.

Effect of surround statistics

Besides the geometric cue (the gap) we considered above, another important factor that can reduce contextual modulation is target-surround dissimilarity. This breakdown in statistical similarity is known to enhance perceptual saliency [40, 41], and in some cases is suggested to act through segmentation [44, 90]. This is well reported for crowding, where the effects of the surround on target identification can be reduced if the two differ in aspects such as color, orientation, or higher-level attributes [21, 36, 42, 45, 64, 90], thus increasing target saliency [28]. Understanding the effects of surround structure on contextual modulation of texture perception is important because during natural scene perception there is abundant variability in texture properties and arrangement. Furthermore, different levels of surround structure are often used as proxies for different stages of neural processing [21, 29, 42, 45], which could provide insights on the mechanisms behind our observations. For these reasons, we next asked how target-surround dissimilarity affects peripheral texture perception, and how it interacts with segmentation.

Experiment 3: FAS dissimilarity but not HOS dissimilarity strongly reduces contextual modulation through segmentation

We focused on target-surround dissimilarity at the FAS and HOS levels because they are related to the SS model [23, 24] and to physiology [1, 23, 24, 53, 54, 102], as discussed above. Previous work [45,

90, 95] suggests that dissimilar surrounds should exert less contextual modulation, but the relative effects of FAS and HOS dissimilarity, and their relation to segmentation depend on the unknown underlying mechanisms.

To test the effects of FAS and HOS dissimilarity, we compared 3 different surround textures (Figure 6a): 1) The same PS texture as the target (none dissimilar), 2) a different PS texture with FAS and pixel histogram matched to the target PS texture (HOS dissimilar), and 3) a different PS texture with only its pixel histogram matched to the target (FAS & HOS dissimilar). Furthermore, to study the interaction of FAS and HOS dissimilarity with segmentation, we showed these surround textures in both the continuous and discontinuous conditions. In this experiment, target size was the same for the continuous and discontinuous conditions, and the gap was generated by enlarging the surround for the discontinuous condition (increasing inner and outer diameter to maintain its width).

We presented participants (n=22) with 4 different target textures (Figure 6a), adding to 31 texture-sessions. To analyze the data we fitted a GLMM with parameters for FAS dissimilarity (β_{FAS}) and HOS dissimilarity (β_{HOS}) separately to the continuous and discontinuous conditions.

First, we asked whether the two levels of dissimilarity had an effect for the continuous stimulus. The effect of HOS dissimilarity was close to 0 and not significant ($\beta_{HOS}^{Cont} = -0.04$, ci $= [-0.69, 0.61]$, p $= 0.87$, Figure 6d), whereas FAS dissimilarity generated strong improvements in performance overall ($\beta_{FAS}^{Cont} = 0.84$, ci $= [0.50, 1.16]$, p $= 2 \times 10^{-3}$, Figure 6d). We note that the effect of HOS showed considerable variability between textures. In particular, for texture T4 performance was strongly reduced for dissimilar HOS, contrary to expectations. This is likely because the surround without dissimilarity for this texture has a high regularity that introduces a phase effect at the target-surround boundary which could act as a segmentation cue.

To better understand the relation between dissimilarity and segmentation, we then asked whether dissimilarity interacted with discontinuity. If the effects of dissimilarity are mediated simply by surround statistics pooled over fixed regions, we would expect dissimilarity effects for the discontinuous condition comparable to those of the continuous condition (assuming, as we do, a pooling area with the radius of Bouma's law such that the small change in surround geometry is negligible). On the other hand, if dissimilarity effects are mediated by segmentation, we expect the effects to be reduced in the discontinuous condition where segmentation is already induced by the gap. Consistent with the second mechanism, we found that target-surround dissimilarity had little effect on contextual modulation in the discontinuous condition (Figure 6c) for both HOS ($\beta_{HOS}^{Discont} = 0.03$, ci $= [-0.36, 0.40]$, p $= 0.84$, Figure 6e), and FAS ($\beta_{FAS}^{Discont} = 0.03$, ci $= [-0.29, 0.34]$, p $= 0.80$, Figure 6e), although there was considerable variability between textures. We verified that the change of the effect of FAS dissimilarity for the discontinuous condition was significant (see Supplementary S4 and Figure S6).

Our analysis therefore suggests that FAS dissimilarity effects are strong and mediated by segmentation, whereas HOS dissimilarity effects show considerable variability across textures but are on average weak. We then tested whether these results could be captured by the SS model. First, in the continuous condition the model showed a strong improvement in performance when there was HOS dissimilarity, and much weaker changes for FAS dissimilarity (Figures 6f, 6h). Second, these effects were mostly unchanged for the discontinuous condition (Figures 6f, 6h, S6), due to the lack of explicit segmentation processes. Therefore, the SS model was not able to capture the patterns observed in the human data.

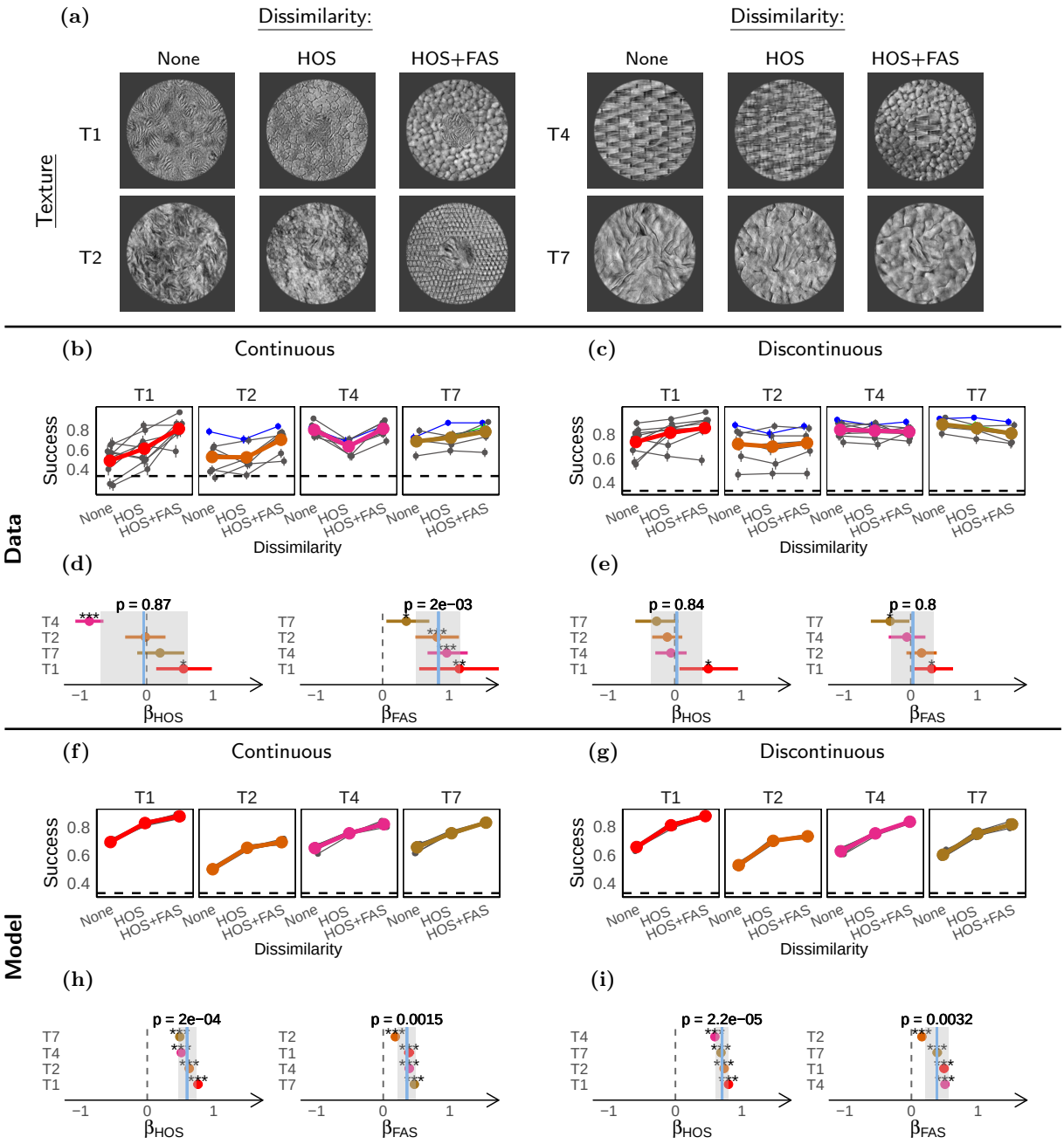

Figure 6: Target-surround dissimilarity reduces contextual modulation. **(a)** Samples of the stimuli used in this experiment, showing for target textures the 3 levels of target-surround dissimilarity used in the experiment (discontinuous stimuli not shown). **(b)** & **(c)** Task performances for the different target-surround dissimilarities in the continuous and discontinuous conditions respectively. **(d)** & **(e)** LOR for HOS (β_{HOS}) and FAS (β_{FAS}) dissimilarity in the continuous and discontinuous conditions respectively. **(f)**-**(i)** Same as **(b)**-**(e)** but for the model observers. Participants (n=22) completed 31 texture-sessions, and performed between 60 and 120 trials per condition. Model observers (n=8) discriminated 1500 stimuli per condition. The plots in this figure use the same conventions as the corresponding plots in Figure 5.

Experiment 4: Naturalistic structure in the surround is important to recruit contextual modulation

The results of the previous section show that a surround with different naturalistic HOS than the target can still exert substantial contextual modulation. Interestingly, other studies have previously shown that contextual modulation can be reduced by removing the natural HOS from the surround. Perceptually, this has been observed for tasks involving recognition and discrimination of natural scenes in peripheral vision [29, 86], and for local orientation processing during scene perception [52]. Neurally, it has been shown that phase-scrambling the surround (i.e. the HOS are removed but the FAS maintained) strongly affects contextual modulation of neural activity in response to natural images in V1 [16, 31, 57] and to naturalistic textures in V2 [101]. In the context of this literature, our finding of a relatively weak effect of HOS dissimilarity in the previous experiment raises the question of whether the presence of natural HOS is necessary for recruiting contextual modulation for textures.

To address this question, we compared the effects of naturalistic and phase-scrambled surrounds continuous to the target (Figure 7a). Because our experiments required to identify the phase-scrambled target, we reasoned that target-surround similarity with the texture to be identified could affect contextual modulation and lead to unpredictable confounding effects. Therefore, to balance out this possible effect of similarity, we asked half the participants to identify the phase-scrambled texture and the other half to identify the naturalistic texture (modifying the task accordingly, see Methods), and we report the results from both task variants together.

Participants (n=28) were presented with 5 textures, adding to 43 texture-sessions. Consistent with previous studies suggesting a role of natural HOS in recruiting contextual modulation processes, we observed that performance was worse with natural HOS in the surround ($\beta_{Nat} = -0.91$, ci $= [-1.25, -0.56]$, p $= 10 \times 10^{-4}$, Figure 7c). Taken together with the previous experiment, this result suggests that the presence of naturalistic HOS in the surround is important for recruiting this contextual modulation phenomenon, whereas HOS similarity with the target is of secondary importance.

Although the observed effect of naturalness agrees with previous reports suggesting a role of naturalistic statistics in recruiting contextual modulation [16, 57, 101], we observed a qualitatively similar effect of naturalness in the SS model for the continuous condition (Figure 7e). This means that at least part of the effect of naturalness could be mediated by simple pooling. Nonetheless, for textures T1 and T2 we also studied the interaction between naturalness and segmentation (see Supplementary S5, Figure S7), and found that adding a discontinuity reduced the effect of naturalness ($\beta_{Nat:Discont} = 0.43$, ci $= [0.06, 0.83]$, p $= 0.04$, Figure S7d), while this effect was not captured by the SS model ($\beta_{Nat:Discont} = 0.05$, ci $= [-0.03, 0.12]$, p $= 0.15$, Figure S7h). Also, further analysis of the model shows that the observed naturalness effect in the model is not due to surround naturalness itself, but rather due to some specific features of our stimulus generating process (see Supplementary S6, Figure S12). Thus, it is likely that pooling is not the only mediator of the effect of naturalness in our experiments.

In conclusion, these results suggest that naturalistic HOS are important in recruiting contextual modulation phenomena. This is compatible with suggestions that contextual modulation phenomena are tuned to the structure of natural images [16, 57], and more specifically, with the results observed for other contextual modulation phenomena in V1 and V2, that may be mechanistically related to our results (see Discussion).

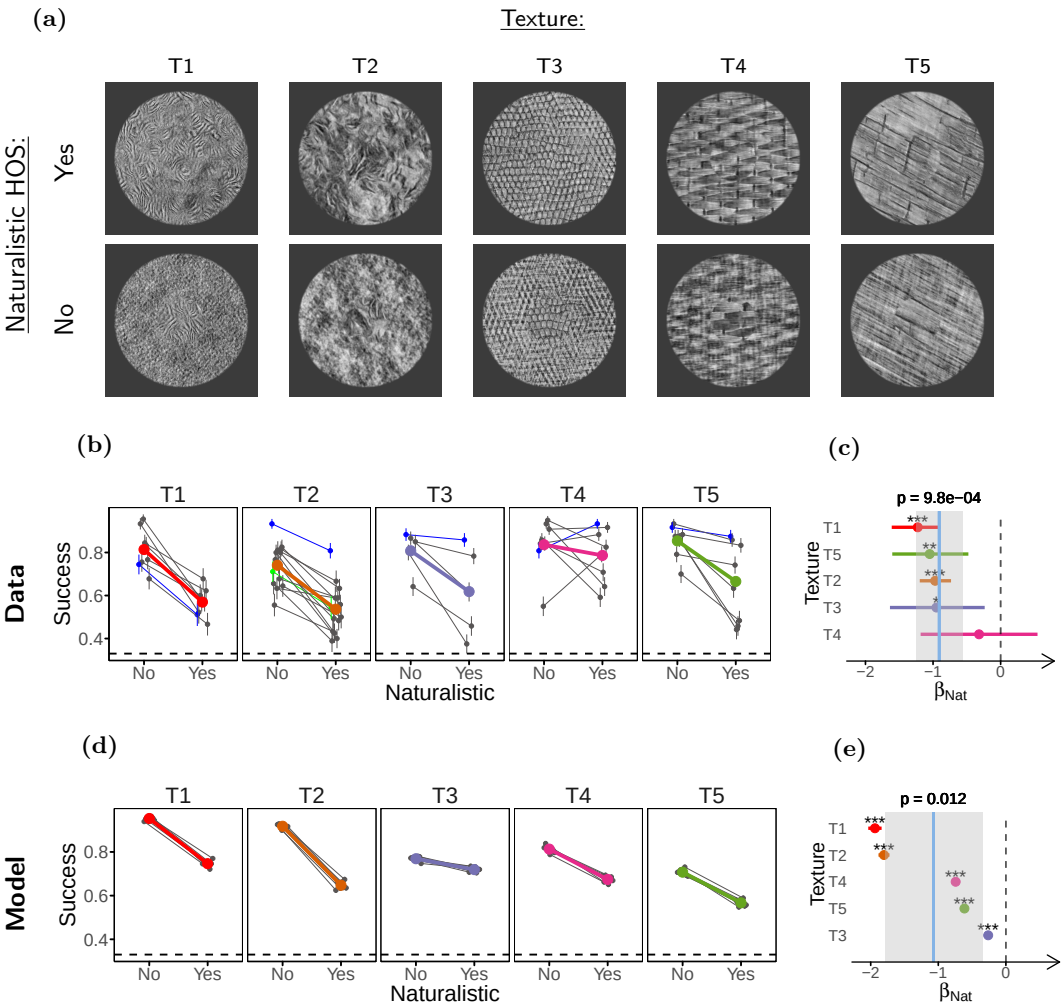

Figure 7: Naturalistic HOS increase contextual modulation. **(a)** Stimuli used in the experiment. Top row: naturalistic surrounds. Bottom row: phase-scrambled surrounds (only naturalistic targets are shown). **(b)** Task performance. **(c)** LOR for the presence of naturalistic HOS (β_{Nat}). **(d)**, **(e)** Same as **(b)** and **(c)** but for the model observers. Participants (n=43) completed 43 texture-sessions and performed between 90 and 120 trials per condition. Model observers (n=8) discriminated 1500 trials per condition. The plots in this figure use the same conventions as the corresponding plots in Figure 5.

Texture crowding

We have thus far shown that texture perception is affected by contextual modulation, and influenced by segmentation and target surround dissimilarity. These characteristics are consistent with a possible role of visual crowding, a contextual modulation phenomenon often regarded as the most important factor of peripheral vision [69]. The SS model explains crowding as a loss of information from pooling together target and surround features when computing local SS [1, 23, 24, 90]. However, it is not clear whether this explanation, that is often applied to tasks on non-texture stimuli, should hold for our task. Thus, we decided to test whether the contextual modulation we observed is due to crowding.

There are two main diagnostic criteria for crowding. One is compliance with Bouma's law, which

states that the critical distance at which surrounds interfere with target perception scales linearly with eccentricity with a slope of approximately 0.5 [58]. The other is an inwards-outwards asymmetry in which surrounds more eccentric (outwards) to the target exert a stronger modulation than surrounds more central (inwards) to the target [21, 58, 62, 69, 90]. Probing Bouma's law with textures poses experimental challenges, such as changing target-surround distance without breaking continuity or altering target size, and determining how to measure distance between texture stimuli (e.g. [68]). Therefore, we decided to probe the characteristic inwards outwards asymmetry of crowding.

Experiment 5: Effect of surround position is small, highly variable and task dependent

To test for inwards-outwards asymmetry in our task, we used half-ring shaped surrounds (Figure 8a) placed inwards or outwards of the target. Participants (n=21) were presented with 5 different textures, completing 37 texture-sessions. Opposite to what has been reported for crowding, performance in our task was consistently lower when the surround was inwards of the target ($\beta_{In} = -0.32$, ci $= [-0.47, -0.18]$, p $= 2 \times 10^{-3}$, Figure 8c). This suggests that crowding as reported for classical letter detection or orientation discrimination may not be the main contextual modulation phenomenon in our experiments.

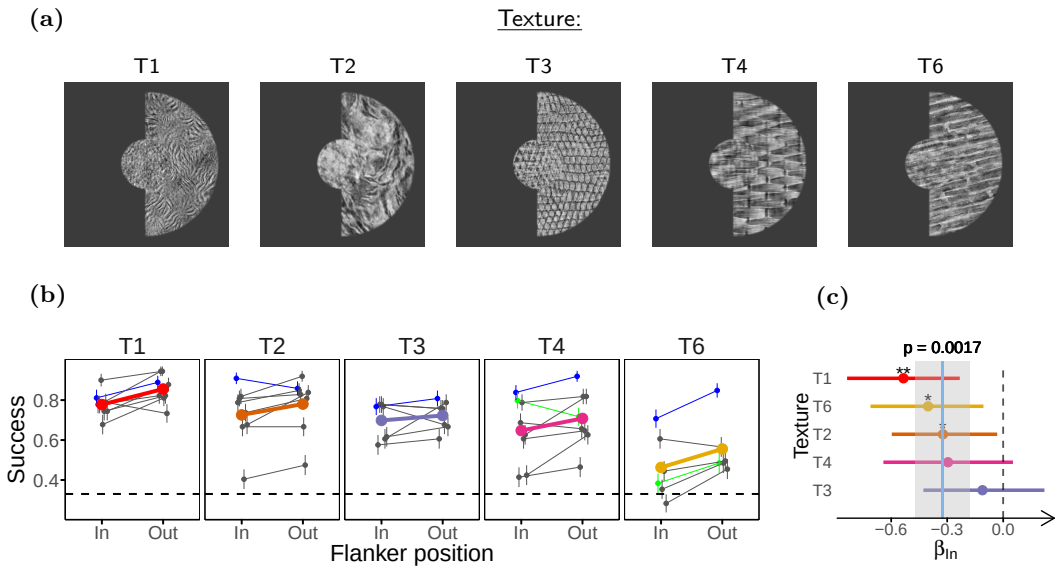

Figure 8: Inwards surrounds affect performance more than outwards surrounds. **(a)** Stimuli used in the experiment. Inwards and outwards surround conditions differ in the position of the half ring of surround texture relative to the fixation point. **(b)** Task performance for the different surround positions. **(c)** LOR for inwards versus outwards surround (β_{In}). Participants (n=21) completed 40 texture-sessions, and performed between 90 and 99 trials per condition. The figure uses the same conventions as Figure 5.

Nonetheless, unlike the task used here, most reports of inwards-outwards asymmetry use only one target [2, 21, 43, 62]. The exception is a study [61] in which two targets were presented simultaneously as in our task, showing that the usual asymmetry (stronger crowding for outwards surround) could be reversed by biasing attention towards the fovea. To verify that the previous result is not only due to this task-related effect, we repeated the experiment for texture T1 using only one target, presented to the right of the fixation point. Participants (n=15) had to report whether the target was naturalistic or phase scrambled. Using this new task we observed an effect of surround position close to 0 ($\beta_{In} = 0.02$, ci $= [-0.32, 0.38]$, p $= 0.88$, Figure 9b). This is not what would be expected from

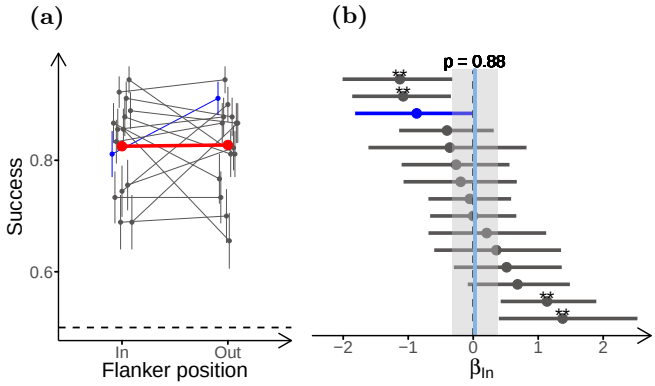

Figure 9: Reduced inwards-outwards asymmetry with single target. **(a)** Task performance for the different surround positions for the task using only one target, for texture T1. **(b)** LOR of the position of the surround for the task using only one target (β_{In}). Participants (n=15) performed 90 trials per condition. The plots in this figure use the same conventions as the corresponding plots in Figure 3.

the classical asymmetry in crowding, and thus supports the conclusion from the experiment using two targets. Nonetheless, we also note that the difference between the results from the two tasks is in agreement with an effect of task and attention on inwards-outwards asymmetry, such as that reported in [61].

Despite the lack of a clear asymmetry in the average performance, variation between participants was high, and some individual participants showed strong effects of surround position in both directions. One plausible interpretation of this result is that contextual modulation in our task arises from different contributing processes, and that participants with stronger crowding effects would show worse performance for outwards targets, whereas participants more affected by other processes would show little or opposite asymmetry. This hypothesis is in line with previous studies reporting substantial variability in sensitivity to crowding between observers [36, 38, 60, 85]. In addition, we hypothesize that this variability in sensitivity to contextual modulation phenomena could arise from the use of different strategies for solving the task, possibly contributing to the considerable between-participant variability that we observed in the results of the previous experiments.

In conclusion, these results suggest that crowding as observed in classical paradigms such as letter recognition, interacts with other processes of at least comparable relevance to contextual modulation of texture perception.

Discussion

Although the SS model of peripheral vision has had considerable success [69], studies using complex scenes [88] and simple object-like stimuli [18, 44, 46] suggest that including processes of segmentation and grouping together with contextual modulation is crucial for a more accurate understanding of peripheral vision. Here we showed that PS texture perception in the periphery is modulated by spatial context, and that contextual modulation is strongly reduced by segmentation engaged both by a gap between target and surround, and by target surround dissimilarity (Figures 4, 5, 6). Although the relevance of segmentation and target-surround dissimilarity for contextual modulation has been studied for discrimination tasks using simple features or objects [36, 44, 45, 65, 73, 74, 90, 100], this is to our knowledge the first report of such effects for texture discrimination, which likely involves

different processing of the visual input [10, 11, 14, 70]. Furthermore, while the simple feature and object stimuli are more difficult to relate to the SS model [71], our choice of stimuli and task allowed for a direct comparison with the SS model. Our analysis showed that the standard SS model with fixed pooling windows failed to reproduce our experimental observations on segmentation. Together, these results thus strongly support the view that the two-stage model with filtering followed by fixed pooling windows cannot fully explain peripheral perception. We note however that this does not argue against the importance of SS as a general framework for understanding peripheral vision, but rather for the need to incorporate segmentation and flexible pooling processes more explicitly.

Our stimulus design also allowed us to study the perceptual relevance of texture properties (specifically, FAS and HOS) related to the different stages of the SS model and of early visual processing. Previous studies using artificial textures have reported that FAS is a stronger segmentation cue than HOS, and that some HOS induce moderate and others weak or no segmentation [34, 84, 98]. Our study is the first to test the role of FAS and HOS in segmentation of naturalistic textures. We found that dissimilarity in FAS was a strong segmentation cue, but we did not observe clear evidence that dissimilarity in the HOS of the PS model induces segmentation in the periphery (Figure 6). This seems also in agreement with simple inspection of our stimuli, in which the targets strongly pop out when the surround is dissimilar in FAS and HOS, but not when it is only dissimilar in HOS. The weak effect of HOS dissimilarity in peripheral vision is particularly interesting if we note that the textures with HOS dissimilarity were noticeably different under foveal inspection. It is also noteworthy that we did not observe FAS dissimilarity effects when we induced segmentation by a discontinuity between target and surround. If the surround were pooled with the target for the discontinuous condition, as the fixed pooling regions model would suggest, we would expect more similar statistics to interfere more, contrary to what we observed. Indeed, this reasoning was supported by our implementation of the SS model, which showed similarity effects for the discontinuous condition and a lack of interaction between continuity and dissimilarity (Figure 6). A possible explanation for this discrepancy between the model and our data is that pooling windows are flexible, and when the surrounds are segmented from the target they are not pooled equally than when grouped together [47, 88].

Given the selectivity of areas V2 and V4, but not V1, for the HOS of the PS model [24, 53, 54, 102] one possible interpretation of this result is that the segmentation process acts primarily in V1, corresponding roughly to the first stage of the SS model. However, V2 and V4 neurons are also highly selective to the FAS of images [24, 53, 54, 102], so the effect of FAS could be mediated by these same areas. Furthermore, the space of PS statistics is high-dimensional, and there is large variability in the efficacy of subsets of PS statistics in driving V2 and V4 neurons [24, 54]. It is therefore possible that dissimilarity of a targeted subset of HOS of the PS model will induce stronger segmentation. For instance, selectivity for other simpler HOS that can support texture segmentation [84] emerges primarily in V2 [97]. Therefore, although our results are consistent with the idea that the segmentation process in the periphery may not use HOS dissimilarities between naturalistic textures (different from other synthetic textures, [84]), a more complete understanding of their role would require probing the space of PS statistics more exhaustively.

What neural mechanisms might underlie the contextual modulation we observe? One candidate is V1 surround suppression, which appears linked to our experimental results in several ways: both strongly depend on FAS similarity [13] and on segmentation cues [16], and it has been proposed that V1 surround suppression underlies perceptual surround suppression [12, 99], which affects texture perception [15, 49] and is relatively strong in peripheral vision [59, 95]. Also, we showed that contextual modulation of naturalistic texture perception is tuned to the naturalness of the HOS, in agreement with previous perceptual [29, 52, 86] and physiological [16, 31, 57, 101] studies in contextual modulation. This too could reflect V1 surround suppression, which has been shown to be reduced for scrambled surrounds (i.e. lacking natural HOS) compared to natural images in V1 [16, 31, 57] (though unpublished recordings indicate this might not be the case for naturalistic textures [103]). Overall,

our experimental results on contextual modulation and segmentation appear consistent with flexible V1 surround suppression [16], in which suppression strength is reduced when center and surround are inferred to be segmented on the basis of image statistics. Another possible mechanism is surround facilitation at the level of V2, which is tuned to natural statistics and appears to counteract surround suppression of V2 responses to textures [101]. Although that study only used uniform textures, unlike our mixed center-surround stimuli, this facilitation mechanism could reduce the difference between the responses to the two kinds of center textures. If this were the case, then according to our results we would predict V2 surround facilitation to be reduced by target surround segmentation, which could be readily tested experimentally.

Lastly, as mentioned above, another possible mechanism behind our results is pooling over flexible windows shaped by segmentation, as proposed in studies of natural scene perception in peripheral vision [88] and orientation discrimination in central vision [47]. Flexible surround suppression, facilitation, and flexible pooling windows could therefore be integrated at the corresponding stages of the SS model, leading to a broader framework within which to interpret our results and guide further studies of peripheral vision.

Although our results on contextual modulation and segmentation appeared consistent with perceptual crowding, we did not observe a clear inwards-outwards asymmetry as is often reported for crowding [62] (Figures 8, 9), indicating that other processes are at least as important as crowding for limiting texture perception in our task. In addition we found large variability across participants, particularly in the comparison between inwards and outwards surrounds, suggesting that different participants are affected differently by those processes, which is also consistent with other crowding studies [36, 38, 60, 85]. To our knowledge, this is the first report of the effects of crowding on texture perception, though crowding has been frequently described as objects undergoing 'forced texture processing' [69]. Given that some of the tasks most associated with peripheral vision such as scene perception [7, 20, 30], guidance of eye movements [25, 56, 77] and the control body of movement [3, 6, 8, 32, 79] have also been connected to texture perception, our results raise the question of to what extent crowding may be the major limitation in some of these natural tasks. Similarly, our results suggest that previous studies of crowding in natural scenes [29, 87] might also have measured, to an unknown degree, other contextual modulation processes affecting texture perception. Crowding has been mostly associated with excessive pooling of target and surround features, which can be easily accommodated by the standard SS model. As explained above, our work points to additional processes such as flexible surround suppression and facilitation, highlighting the importance of further studying their relative contributions.

Methods

Texture synthesis

Portilla-Simoncelli synthesis and phase-scrambling

We synthesized grayscale naturalistic textures using the Portilla-Simoncelli (PS) texture synthesis algorithm [63] in Octave [19]. The algorithm first computes a set of statistics over an input image, including mean luminance, contrast and higher-order moments of the pixel histogram; and the means and pairwise correlations of the activations of multi-scale, multi-orientation filters (steerable pyramid [78]) analogous to V1 cells. Then it iteratively modifies a white noise image until its statistics match those of the input image. We used as input images natural textures from the Brodatz texture database, the Amsterdam Library of Textures [9] and from the database presented by Lazebnik et.al. [37]. We

refer to an image synthesized this way as a naturalistic texture or PS texture. We used filters with 4 scales and 4 orientations, and 9 by 9 pixels neighborhood for computing the spatial correlations of the filter responses. We synthesized two 1024 x 1024 PS textures for each input image.

For each PS texture we also synthesized a phase-scrambled texture. This was achieved by first generating a uniform noise image and then replacing its Fourier amplitude spectrum (FAS) for the FAS of the naturalistic texture.

Pixel histogram matching

Phase-scrambling a naturalistic image can change the histogram of pixel activations (e.g. changing the minimum and maximum intensities). To prevent participants from using aspects of the pixel histogram (e.g. brightness) as cues to solve the task, we matched the pixel histograms of the naturalistic and phase-scrambled images to an average of the two, using the SHINE package for Octave [94] with 30 iterations. In each iteration their FAS was also matched to the original FAS, and the structural similarity index (SSIM) with respect to the original image was also optimized in order to reduce alterations to image structure [89, 94]. Images produced by this method appeared very similar to the starting textures (besides changes in pixel intensities), suggesting it did not produce noticeable structural alterations.

Matching pixel histogram and FAS of different naturalistic textures

In experiment 3, to generate the surround image that was dissimilar to the target only in HOS, we started by generating a new PS texture using a different input image than the one used for the target. Then we matched its FAS and pixel histogram to those of the target PS texture with the SHINE package, using 30 iterations. In each iteration the SSIM with respect to the original surround PS texture was also optimized. For the surround texture that was dissimilar in both FAS and HOS, the same procedure was used but without matching the FAS to the target PS texture.

Texture selection

We applied the synthesis procedure using several natural textures as input, and then selected those where the PS and phase-scrambled textures subjectively appeared to be more discriminable, so as to make the task easier. We also selected textures that had different kinds of structures, in order to better probe the texture space (e.g. strongly oriented, weakly oriented, regular, irregular).

In experiment 3 most textures to which we applied the FAS matching procedure acquired a phase-scrambled appearance, so we selected for further use those that maintained a naturalistic appearance after this procedure.

Stimuli and task

The task was performed in a dark room, using a 27 inch LCD screen (ASUS, model PG278QR) with a refresh rate of 60 Hz. Participants used a chinrest to maintain a viewing distance of 40 cm, at which 1° of the visual field subtended 30 pixels. Experiments were ran on Psychtoolbox-3 [35] running in Octave Version 4.0.0 in Ubuntu 14.04.

The background gray had a luminance of $8.7\,\mathrm{cd\,m^{-2}}$, and the textures used in the experiments had a range of mean luminances of $50.9\,\mathrm{cd\,m^{-2}}$ to $67.3\,\mathrm{cd\,m^{-2}}$, and a range of standard deviations in the luminance of the pixels of $26.9\,\mathrm{cd\,m^{-2}}$ to $34.1\,\mathrm{cd\,m^{-2}}$, as determined with a screen calibration performed with a colorimeter (Cambridge Research Systems, model ColorCAL II).

Texture cropping and rotation

All textures shown in the experiments were patches cropped from the larger synthesized images, with a linear transparency gradient at their border, allowing for a smooth fading with their neighboring surface e.g. the background or a neighboring texture). These gradients had a length of 4 pixels, roughly equivalent to 0.15°.

For each texture patch displayed, the cropped region was randomly selected over the whole image on a trial by trial basis. We note that since the PS statistics were matched over the large synthesized images, the random sampling of patches from these images introduced some trial by trial variation in the texture statistics displayed. Although testing the effect of this image variability on our results would require additional experiments, we think this variability is unlikely to have significant effects on the participants performance. First, participants could readily generalize from being shown the textures and immediately learned to discriminate naturalistic and scrambled textures despite this variability. Second, participant errors could not be predicted from this image variability (analysis not shown). Third, we studied the behavior of the SS model for the task without image sampling variability, and although there were some differences with the behavior for the original task, the conclusions when comparing to the participants remained unaltered, except for the comparison between naturalistic and scrambled surround as explained in the main text (see Supplementary S6 for the results, and further discussion on image sampling variability).

In each individual trial an angle multiple of 90° was randomly chosen and all textures were rotated by this angle before being cropped for display. This was done to reduce participants adaptation to low level properties of the textures.

Texture discrimination task

Our task is a variation of that described by Freeman et.al. [24], and consisted in discriminating between the naturalistic and the phase-scrambled versions of a texture. Because we did not use eye-tracking, we used a task design that discouraged participants from looking away from fixation.

The target stimuli (targets) consisted of two circular patches of texture presented simultaneously for 233 ms, centered at 12° to the right and to the left of the fixation point (Figure 2). We used three different target configurations: 1) phase-scrambled target to the right (PS texture to the left), 2) phase-scrambled target to the left (PS texture to the right), or 3) no phase-scrambled target (PS texture in both targets). The three configurations were shown an equal amount of times, in random order. Participants were instructed to report the location of the phase-scrambled target with the arrow keys, and to use the upwards arrow to indicate the absence of phase-scrambled targets.

Trial dynamics

The sequence of events in any given trial was the following (see Figure 2): 1) Start with the gray screen, 2) after 500 ms a red fixation dot appeared at the center of the screen, which participants were

instructed to fixate, 3) after a time interval sampled uniformly from 400 ms to 600 ms, the two targets were presented simultaneously for 233 ms (14 frames), 4) after the targets disappeared the participant responded (without a time limit), 5) auditory feedback was provided and the fixation dot disappeared, returning to step 1). Participants were told to use the response stage (step 4) to rest as needed by delaying the response.

Surround textures

In all experiments we included surrounding textures with varying shapes and texture contents. In any given trial the two targets shared the same kind of surround. These surrounds were also sampled randomly (and independently from each other and from the targets) from the larger synthesized textures. Unless indicated otherwise in the text, the surrounds were sampled from the PS texture of the texture pair to be discriminated in the targets.

In most cases, surrounds were rings (or half-rings) with a width (i.e. distance between inner and outer edges) equal to target diameter. Experiment 1 and texture T1 in experiment 3 were an exception, having a surround width 1.4 times the diameter of the target. The surrounds of the split-disk targets in experiment 2 were not rings, but they had the same outer diameter as the surrounds for the corresponding disk-shaped targets.

Surrounds could be contiguous to the target or separated by a gap showing the gray background. The gap had a width of 0.5° in all cases except experiment 1, where it had a width of 0.35° and texture T1 in experiment 3, where both gaps of 0.5° and 1° were used (although these were grouped together for the analysis, see Supplementary S5).

Organization of trials into experiment and texture blocks

Experiment sessions were divided into 2 to 4 experimental blocks (with balanced conditions) separated by 30 s resting periods. Within these blocks, some conditions were also blocked (with their order balanced across participants) to avoid confusions during the task. The different conditions contained within a block were randomly interleaved and each was presented an equal amount of times.

Each experimental session used either a single texture or two different textures. When two textures were used, the experiment was first completed with one texture and then with the other (with their order balanced across participants).

The total duration of the experiments, including training and instructions, was between 30 and 45 minutes.

Training and difficulty adjustment

Before the experiment, participants were provided with training opportunity. Auditory feedback was used in all stages of training, as well as in the main experiment. In the first training session, targets were shown without surround and remained on the screen until the participant responded. The second training session also used targets only, but had the same dynamics as the experiment. Both sessions were terminated at will by the participant by pressing a special key.

After running experiments 1 and 3 with texture T1 with a target diameter of 3.5°, we observed

considerable variability between participants in task performance. Therefore we adjusted task difficulty to each participant, in order to drive participants to a more informative performance range (preventing saturation with very high or with chance-level performances). To this aim, we presented a sequence of trials with unsurrounded targets in which target diameter was adaptively adjusted using the accelerated stochastic approximation procedure [81] to drive participant performance to a predetermined level of 90% correct responses (see Supplementary S7 for details on the procedure and final sizes distribution). If the final target diameter was larger than 5.3° (160 pixels), we used a diameter of 5.3° in the experiments. Some of the textures had a mean target diameter around 5.3°, while others were near 4° or even lower. The widths of the surrounds were then set equal to the target diameter.

After size adjustment, we repeated the static and dynamic training stages as described above, including also the surrounds, and instructed participants to perform the task ignoring the surrounds. Again, participants terminated these sessions at will.

Participants and exclusion criteria

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Research Ethics Committee of the Faculty of Psychology of the Universidad de la República. Participants gave signed consent to participate in the experiment, and to have the anonymized data from the experiments made available online. Participants were given no economic or course credit reward for their participation in the experiment.

A total of 98 adult individual participants (including the authors DH and LG, denoted in the figures by colors blue and green respectively), participated in the experiments, of which 34 were women. All participants had normal or corrected to normal vision. Some participants performed more than one experimental session, although no participant performed the same experiment with the same texture more than once. We call the combination of a participant performing an experiment with a given texture a texture-session (as described above). In total, participants completed 189 texture-sessions.

We excluded from analysis texture-sessions in which the participant performed below 45% correct for all conditions (chance level performance is 33%), to avoid strong floor effects. This criterion discarded 14 of the total 189 texture-sessions. In the main text we report for each experiment the number of texture-sessions that satisfied the inclusion criterion. All results and analysis are robust to removing this exclusion criterion, as well as to excluding the main author from the analysis.

Summary-statistics model observer

We implemented an image-computable observer model based on the feedforward SS model with fixed pooling windows [24]. This model first computes PS statistics over the two stimuli, then computes their difference and feeds it to a linear discriminator to solve the task (see Figure 2b). The weights of the discriminator were optimized to maximize discrimination performance on a training set, and the model is then tested on a separate test set (cross-validation). We were interested in testing whether the model reproduced qualitatively the trends in our data, therefore we did not fit model parameters to match human behavior quantitatively. However, we added noise to the PS statistics computed by the model in both training and testing stages, to roughly match the performance of the human participants on average across stimuli.

Computing the SS

We first generated sample images of single stimuli such as those used in the experiments, with either phase-scrambled or naturalistic targets diameter 110 pixels, and with the different surrounds. We adapted the code of [23] to compute PS statistics over a circular fixed pooling area centered on the target. We used a pooling area with a diameter of 360 pixels, equivalent to 12° of visual field. We based this pooling size on Bouma's law of crowding [90], which says that surround elements hinder target perception when they are within a distance of about 0.5 times the eccentricity, thus we used this distance ($12^\circ \times 0.5$) as the radius of integration around the target center.

We computed PS statistics using 4 scales, 4 orientations, and a neighborhood for computing spatial correlations of 7 pixels (smaller than for texture synthesis to reduce the number of model parameters). This procedure leads to 782 SS per stimulus (after removing the repetitions of symmetric parameters from the correlation matrices).

Stimulus pairing

To mimic the experimental task we arranged the stimuli (which either had naturalistic or phase-scrambled target) into 3 kinds of ordered pairs, equivalent to those shown in the experiment. Using *Nat* and *Scr* to refer to stimuli with naturalistic and scrambled targets respectively, the 3 kinds of ordered pairs were: {*Scr*, *Nat*}, {*Nat*, *Scr*} or {*Nat*, *Nat*}. As in the experiment, the stimuli from a given pair had the same surround. Then, we subtracted the SS of the second stimulus to each corresponding SS of the first stimulus, resulting in 782 differences in SS (or predictors) for each stimulus pair.

Observer training

The observer consisted of a linear discriminator trained to predict the class of the stimulus pair (e.g. {*Scr*, *Nat*}, {*Nat*, *Scr*} or {*Nat*, *Nat*}) from the SS difference of the pair.

First, for an observer trained for a given experiment, we generated 750 stimulus pairs (250 of each class), or trials, for each different surround condition in the experiment, and computed the difference in SS (predictors) for each generated pair. We then added Gaussian noise to the predictors, with a standard deviation equal to the standard deviation of the predictor across the training dataset containing all the conditions for the simulated experiment). Next, we normalized each predictor to have unit variance (using the default setting of the fitting package, glmnet [26]). Lastly, we trained multiclass logistic regression on the normalized predictors with L2 penalization, and optimized the hyperparameter that weights the penalization by 10-fold cross-validation (i.e. the default in the glmnet package). For each experiment, we trained 8 different models (observers), using different noise samples and different samples for the training set, leading to some variability between model observers.

After training the models, we tested their discrimination performance on a test set comprising 1500 texture pairs (500 of each class) for each surround condition.

We verified that all the trends and conclusions are robust to the choices of target size, penalization (we tested also elasticnet, which uses a mixed L1 and L2 penalization), and noise level. Furthermore, we also ran the model with a variation of the task that involved no stimulus sampling variability (see Supplementary S6).

Statistical analysis

All experiments were first performed with texture T1, and all but experiment 1 were then reproduced with other textures. Experiments performed with T1 sometimes had more conditions than experiments with the other textures. These conditions exclusive to T1 are analyzed separately in the supplementary analysis.

We analyzed the data of the experiments and the simulations using generalized linear mixed models (GLMM) of the binomial family [27]. In these models we included a fixed effect for each parameter of interest and an offset term. For each of the fixed effects we added random effects. When applying the GLMM to multiple textures to estimate the mean effect across textures (e.g. Figure 5), we included for each fixed effect a random effect for texture and a random effect for participants nested within texture. We also applied the GLMM to individual textures, both for the analysis of data that was only collected with one texture (e.g. experiment 1), and for estimating the effects of the different manipulations on each texture. In the plots showing the effects for multiple textures (e.g. Fig. 5), the estimate for each individual texture was obtained by fitting a GLMM to that texture individually. In these cases we only used a random effect for participants. Correlations between random effects in the model were always set to 0, to avoid overly complex models [5].

All the GLMM fitted by maximum likelihood using the R package lme4 [4]. The reported p-value for each effect was obtained by a likelihood ratio test (LRT) between the full model and the null model, in which that fixed effect is set to 0. The 95% confidence intervals of the fixed effects were obtained by likelihood profiling.

The analysis in the text is based on the parameters fitted by these models, which are in log-odds ratio (LOR) units. Although less intuitive than simple differences between success probabilities, this is a more adequate measure for the experimental effects, especially given the variability in performance between participants and textures.

In some cases we fitted a GLM to the data of each participant in an experiment in order to display the actual observed LOR for each individual (e.g. Figure 3). These models contained no random effects. The confidence intervals for the parameters were obtained by the Wald method, and their p-values by the Wald test.

Data analysis was performed in R 3.4.4 [80] using the packages lme4 1.1-19 [4], dplyr 0.7.6 [93], tidyr 0.8.1 [92], ggplot2 3.0.0 [91], broom 0.5.0 [67], MASS 7.3-50 [82], and this document was generated using knitr 1.20 [96].

Data availability

The anonymized raw data of the experiments, together with the analysis code, and the code for running the experiments, are available in the Open Science Framework (`https://osf.io/8zr5h/`). All participants gave informed written consent for their anonymize data to be publicly shared.

Acknowledgments. We thank Corey Ziemba for useful discussions on an earlier version of this manuscript. We also thank Adam Kohn for comments on an earlier version of this manuscript. This work was funded by the studentships awarded to D. Herrera and L. Gómez-Sena by the Comisión Académica de Posgrados, UdelaR, Uruguay, and by travelships awarded to D. Herrera by PEDECIBA, UdelaR, Uruguay, and CSIC, Uruguay. Ruben Coen-Cagli was supported in part by National Institutes of Health grant EY031166.

References

[1] Benjamin Balas, Lisa Nakano, and Ruth Rosenholtz. "A summary-statistic representation in peripheral vision explains visual crowding". In: *Journal of vision* 9.12 (2009), p. 13. DOI: 10.1167/9.12.13.

[2] William P. Banks, Douglas W. Larson, and William Prinzmetal. "Asymmetry of visual interference". In: *Perception & Psychophysics* 25.6 (1979), pp. 447–456. DOI: 10.3758/BF03213822.

[3] Benoit G. Bardy, William H. Warren, and Bruce A. Kay. "The role of central and peripheral vision in postural control during walking". In: *Perception & psychophysics* 61.7 (1999), pp. 1356–1368. DOI: 10.3758/BF03206186.

[4] Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Fabian Scheipl, Gabor Grothendieck, Peter Green, and John Fox. *lme4: Linear Mixed-Effects Models using 'Eigen' and S4*. Version 1.1-21. Mar. 5, 2019.

[5] Douglas Bates, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. "Parsimonious Mixed Models". In: *arXiv:1506.04967 [stat]* (June 16, 2015). arXiv: 1506.04967.

[6] Andrea Berencsi, Masami Ishihara, and Kuniyasu Imanaka. "The functional role of central and peripheral vision in the control of posture". In: *Human Movement Science* 24.5 (2005), pp. 689–709. DOI: 10.1016/j.humov.2005.10.014.

[7] Timothy F. Brady, Anna Shafer-Skelton, and George A. Alvarez. "Global Ensemble Texture Representations are Critical to Rapid Scene Perception". In: *Journal of Experimental Psychology: Human Perception and Performance* 43.6 (2017), pp. 1160–1176. DOI: 10.1037/xhp0000399.

[8] Th. Brandt, J. Dichgans, and E. Koenig. "Differential effects of central versus peripheral vision on egocentric and exocentric motion perception". In: *Experimental Brain Research* 16.5 (Mar. 1973), pp. 476–491. DOI: 10.1007/BF00234474.

[9] Gertjan J. Burghouts and Jan Mark Geusebroek. "Material-specific adaptation of color invariant features". In: *Pattern Recognition Letters* 30.3 (2009). Publisher: Elsevier B.V., pp. 306–313. DOI: 10.1016/j.patrec.2008.10.005.

[10] Jonathan S. Cant and Yaoda Xu. "Object Ensemble Processing in Human Anterior-Medial Ventral Visual Cortex". In: *Journal of Neuroscience* 32.22 (2012). ISBN: 1529-2401 (Electronic)\r0270-6474 (Linking), pp. 7685–7700. DOI: 10.1523/JNEUROSCI.3325-11.2012.

[11] Jonathan S Cant, Mary-Ellen Large, Lindsay McCall, and Melvyn A Goodale. "Independent Processing of Form, Colour, and Texture in Object Perception". In: *Perception* 37.1 (Jan. 1, 2008), pp. 57–78. DOI: 10.1068/p5727.

[12] Matteo Carandini and David J Heeger. "Normalization as a canonical neural computation". In: *Nature Reviews Neuroscience* 13 (2012), pp. 51–62. DOI: 10.1038/nrn3136.

[13] James R. Cavanaugh, Wyeth Bair, and J. Anthony Movshon. "Selectivity and Spatial Distribution of Signals From the Receptive Field Surround in Macaque V1 Neurons". In: *Journal of Neurophysiology* 88.5 (Nov. 1, 2002), pp. 2547–2556. DOI: 10.1152/jn.00693.2001.

[14] C. Cavina-Pratesi, R. W. Kentridge, C. A. Heywood, and A. D. Milner. "Separate Channels for Processing Form, Texture, and Color: Evidence from fMRI Adaptation and Visual Object Agnosia". In: *Cerebral Cortex* 20.10 (Oct. 1, 2010), pp. 2319–2332. DOI: 10.1093/cercor/bhp298.

[15] Charles Chubb, George Sperling, and Joshua A Solomon. "Texture interactions determine perceived contrast". In: *Proc. Natl. Acad. Sci. USA* 86.23 (1989), pp. 9631–9635. DOI: 10.1073/pnas.86.23.9631.

[16] Ruben Coen-Cagli, Adam Kohn, and Odelia Schwartz. "Flexible gating of contextual influences in natural vision". In: *Nature Neuroscience* 18.11 (2015). Publisher: Nature Publishing Group, pp. 1648–1655. DOI: 10.1038/nn.4128.

[17] Michael A. Cohen, Daniel C. Dennett, and Nancy Kanwisher. "What is the Bandwidth of Perceptual Experience?" In: *Trends in Cognitive Sciences* 20.5 (2016), pp. 324–335. DOI: 10.1016/j.tics.2016.03.006.

[18] Adrien Doerig, Alban Bornet, Ruth Rosenholtz, Gregory Francis, Aaron M. Clarke, and Michael H. Herzog. "Beyond Bouma's window: How to explain global aspects of crowding?" In: *PLOS Computational Biology* 15.5 (May 10, 2019), e1006580. DOI: 10.1371/journal.pcbi.1006580.

[19] John W. Eaton, David Bateman, Soren Hauberg, and Rik Wehbring. *GNU Octave version 4.0.0 manual: a high-level interactive language for numerical computations*. 2015.

[20] Krista A Ehinger and Ruth Rosenholtz. "A general account of peripheral encoding also predicts scene perception performance". In: *Journal of Vision* 16.2016 (2016), pp. 1–19. DOI: 10.1167/16.2.13.doi.

[21] Faraz Farzin, Susan M. Rivera, and David Whitney. "Holistic crowding of Mooney faces". In: *Journal of Vision* 9.6 (June 1, 2009), pp. 18–18. DOI: 10.1167/9.6.18.

[22] Gregory Francis, Mauro Manassi, and Michael H Herzog. "Neural Dynamics of Grouping and Segmentation Explain Properties of Visual Crowding". In: *Psychological Review* 124.4 (2017), pp. 483–504. DOI: 10.1037/rev0000070.

[23] Jeremy Freeman and Eero P. Simoncelli. "Metamers of the ventral stream". In: *Nature Neuroscience* 14.9 (Sept. 2011), pp. 1195–1201. DOI: 10.1038/nn.2889.

[24] Jeremy Freeman, Corey M Ziemba, David J Heeger, Eero P Simoncelli, and J Anthony Movshon. "A functional and perceptual signature of the second visual area in primates." In: *Nature neuroscience* 16.7 (2013), pp. 974–981. DOI: 10.1038/nn.3402.

[25] Hans-Peter Frey, Peter König, and Wolfgang Einhäuser. "The role of first- and second-order stimulus features for human overt attention". In: *Perception & Psychophysics* 69.2 (Feb. 1, 2007), pp. 153–161. DOI: 10.3758/BF03193738.

[26] Jerome Friedman, Trevor Hastie, Rob Tibshirani, Balasubramanian Narasimhan, Noah Simon, and Junyang Qian. *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. Version 3.0-2. Dec. 11, 2019.

[27] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Google-Books-ID: c9xLKzZWoZ4C. Cambridge University Press, Dec. 18, 2006. 651 pp.

[28] Carolina Gheri, Michael J. Morgan, and Joshua A. Solomon. "The relationship between search efficiency and crowding". In: *Perception* 36.12 (2007), pp. 1779–1787. DOI: 10.1068/p5595.

[29] Mingliang Gong, Yuming Xuan, L. James Smart, and Lynn A. Olzak. "The extraction of natural scene gist in visual crowding". In: *Scientific Reports* 8.1 (Sept. 19, 2018), p. 14073. DOI: 10.1038/s41598-018-32455-6.

[30] Iris I A Groen, Edward H Silson, and Chris I Baker. "Contributions of low- and high-level properties to neural processing of visual scenes in the human brain". In: *Philosophical transactions of the Royal Society B* 372.1714 (2017), p. 20160102. DOI: 10.1098/rstb.2016.0102.

[31] Kun Guo, Robert G. Robertson, Sasan Mahmoodi, and Malcolm P. Young. "Centre-surround interactions in response to natural scene stimulation in the primary visual cortex". In: *European Journal of Neuroscience* 21.2 (2005), pp. 536–548. DOI: 10.1111/j.1460-9568.2005.03858.x.

[32] Thomas L. Harrington, Marcia K. Harrington, Denise Quon, Rick Atkinson, Robert Cairns, and Kriss Kline. "Perception of Orientation of Motion as Affected by Change in Divergence of Texture, Change in Size, and in Velocity". In: *Perceptual and Motor Skills* 61.3 (Dec. 1985), pp. 875–886. DOI: 10.2466/pms.1985.61.3.875.

[33] Bela Julesz. "Visual pattern discrimination". In: *IRE Transactions on Information Theory* 8.2 (1962). ISBN: 0340-1200, pp. 84–92. DOI: 10.1109/TIT.1962.1057698.

[34] Bela Julesz and Terry Caelli. "On the limits of Fourier decompositions in visual texture perception". In: *Perception* 8.1 (1979), pp. 69–73. DOI: 10.1068/p080069.

[35] Mario Kleiner, David H Brainard, Denis G Pelli, Chris Broussard, Tobias Wolf, and Diederick Niehorster. "What's new in Psychtoolbox-3?" In: *Perception* 36 (2007), S14. DOI: 10.1068/v070821.

[36] Frank L. Kooi, Alex Toet, Srimant P. Tripathy, and Dennis M. Levi. "The effect of similarity and duration on spatial interaction in peripheral vision". In: *Spatial Vision* 8.2 (1994), pp. 255–279. DOI: 10.1163/156856894X00350.

[37] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. "A sparse texture representation using local affine regions". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (Aug. 2005), pp. 1265–1278. DOI: 10.1109/TPAMI.2005.151.

[38] Maria Lev and Uri Polat. "Space and time in masking and crowding". In: *Journal of Vision* 15.13 (2015), p. 10. DOI: 10.1167/15.13.10.

[39] Dennis M. Levi. "Crowding—An essential bottleneck for object recognition: A mini-review". In: *Vision Research* 48.5 (Feb. 1, 2008), pp. 635–654. DOI: 10.1016/j.visres.2007.12.009.

[40] Zhaoping Li. "A saliency map in primary visual cortex". In: *Trends in Cognitive Sciences* 6.1 (2002), pp. 9–16. DOI: 10.1016/S1364-6613(00)01817-9.

[41] Zhaoping Li. "Contextual influences in V1 as a basis for pop out and asymmetry in visual search". In: *Proceedings of the National Academy of Sciences* 96.18 (Aug. 31, 1999), pp. 10530–10535. DOI: 10.1073/pnas.96.18.10530.

[42] Elizabeth G. Louie, David W. Bressler, and David Whitney. "Holistic crowding: Selective interference between configural representations of faces in crowded scenes". In: *Journal of Vision* 7.2 (Jan. 2, 2007), pp. 24–24. DOI: 10.1167/7.2.24.

[43] Mauro Manassi, Bilge Sayim, and Michael H. Herzog. "Grouping, pooling, and when bigger is better in visual crowding". In: *Journal of Vision* 12.10 (Sept. 1, 2012), pp. 13–13. DOI: 10.1167/12.10.13.

[44] Mauro Manassi, Bilge Sayim, and Michael H. Herzog. "When crowding of crowding leads to uncrowding". In: *Journal of Vision* 13.13 (2013), p. 10. DOI: 10.1167/13.13.10.

[45] Mauro Manassi and David Whitney. "Multi-level Crowding and the Paradox of Object Recognition in Clutter". In: *Current Biology* 28.3 (Feb. 5, 2018), R127–R133. DOI: 10.1016/j.cub.2017.12.051.

[46] Mauro Manassi, Sophie Lonchampt, Aaron Clarke, and Michael H. Herzog. "What crowding can tell us about object representations". In: *Journal of Vision* 16.3 (Feb. 1, 2016), pp. 35–35. DOI: 10.1167/16.3.35.

[47] Isabelle Mareschal, Michael P. Sceniak, and Robert M. Shapley. "Contextual influences on orientation discrimination: binding local and global cues". In: *Vision Research* 41.15 (July 2001), pp. 1915–1930. DOI: 10.1016/s0042-6989(01)00082-7.

[48] Josh H. McDermott and Eero P. Simoncelli. "Sound Texture Perception via Statistics of the Auditory Periphery: Evidence from Sound Synthesis". In: *Neuron* 71.5 (Sept. 8, 2011), pp. 926–940. DOI: 10.1016/j.neuron.2011.06.032.

[49] James Scott McDonald and Yoav Tadmor. "The perceived contrast of texture patches embedded in natural images". In: *Vision Research* 46.19 (Oct. 1, 2006), pp. 3098–3104. DOI: 10.1016/j.visres.2006.04.014.

[50] C Meinecke and L Kehrer. "Peripheral and foveal segmentation of angle textures." In: *Perception & psychophysics* 56.3 (1994), pp. 326–334. DOI: 10.3758/BF03209766.

[51] Kazunori Morikawa. "Central performance drop in texture segmentation: the role of spatial and temporal factors". In: *Vision Research* 40.25 (Jan. 2000), pp. 3517–3526. DOI: 10.1016/S0042-6989(00)00170-X.

[52] Peter Neri. "Object segmentation controls image reconstruction from natural scenes". In: *PLoS biology* 15.8 (2017), e1002611. DOI: 10.1371/journal.pbio.1002611.

[53] Gouki Okazawa, Satohiro Tajima, and Hidehiko Komatsu. "Gradual Development of Visual Texture-Selective Properties Between Macaque Areas V2 and V4". In: *Cerebral Cortex* 27.10 (Sept. 20, 2016), pp. 4867–4880. DOI: 10.1093/cercor/bhw282.

[54] Gouki Okazawa, Satohiro Tajima, and Hidehiko Komatsu. "Image statistics underlying natural texture selectivity of neurons in macaque V4." In: *Proceedings of the National Academy of Sciences of the United States of America* 112.4 (2015), E351–60. DOI: 10.1073/pnas.1415146112.

[55] Laura Parkes, Jennifer Lund, Alessandra Angelucci, Joshua A. Solomon, and Michael Morgan. "Compulsory averaging of crowded orientation signals in human vision." In: *Nature neuroscience* 4.7 (2001), pp. 739–744. DOI: 10.1038/89532.

[56] Derrick J. Parkhurst and Ernst Niebur. "Texture contrast attracts overt visual attention in natural scenes". In: *European Journal of Neuroscience* 19.3 (Feb. 2004), pp. 783–789. DOI: 10.1111/j.0953-816X.2003.03183.x.

[57] Michael Pecka, Yunyun Han, Elie Sader, and Thomas D. Mrsic-Flogel. "Experience-Dependent Specialization of Receptive Field Surround for Selective Coding of Natural Scenes". In: *Neuron* 84.2 (Oct. 22, 2014), pp. 457–469. DOI: 10.1016/j.neuron.2014.09.010.

[58] Denis G. Pelli, Melanie Palomares, and Najib J. Majaj. "Crowding is unlike ordinary masking: distinguishing feature integration from detection." In: *Journal of vision* 4.12 (2004), pp. 1136–1169. DOI: 10.1167/4.12.12.

[59] Yury Petrov, Matteo Carandini, and Suzanne McKee. "Two Distinct Mechanisms of Suppression in Human Vision". In: *The Journal of Neuroscience* 25.38 (2005), pp. 8704–8707. DOI: 10.1523/JNEUROSCI.2871-05.2005.

[60] Yury Petrov and Olga Meleshkevich. "Asymmetries and idiosyncratic hot spots in crowding". In: *Vision Research* 51.10 (2011), pp. 1117–1123. DOI: 10.1016/j.visres.2011.03.001.

[61] Yury Petrov and Olga Meleshkevich. "Locus of spatial attention determines inward–outward anisotropy in crowding". In: *Journal of Vision* 11.4 (Apr. 1, 2011), pp. 1–1. DOI: 10.1167/11.4.1.

[62] Yury Petrov, Ariella V. Popple, and Suzanne P. McKee. "Crowding and surround suppression: Not to be confused". In: *Journal of Vision* 7.2 (2007). ISBN: 1534-7362 (Electronic)\n1534-7362 (Linking), p. 12. DOI: 10.1167/7.2.12.

[63] Javier Portilla and Eero P. Simoncelli. "A parametric texture model based on joint statistics of complex wavelet coefficients". In: *International Journal of Computer Vision* 40.1 (2000), pp. 49–70. DOI: 10.1023/A:1026553619983.

[64] Endel Põder. "Effect of colour pop-out on the recognition of letters in crowding conditions". In: *Psychological Research* 71.6 (Sept. 12, 2007), pp. 641–645. DOI: 10.1007/s00426-006-0053-7.

[65] Cheng Qiu, Daniel Kersten, and Cheryl A Olman. "Segmentation decreases the magnitude of the tilt illusion". In: *Journal of Vision* 13.13 (2013), p. 19. DOI: 10.1167/13.13.19.

[66] Maximilian Riesenhuber and Tomaso Poggio. "Hierarchical models of object recognition in cortex". In: *Nature Neuroscience* 2.11 (Nov. 1999), pp. 1019–1025. DOI: 10.1038/14819.

[67] David Robinson and Alex Hayes. *broom: Convert Statistical Analysis Objects into Tidy Tibbles*. Version 0.5.0. 2018.

[68] Sarah Rosen, Ramakrishna Chakravarthi, and Denis G Pelli. "The Bouma law of crowding, revised: Critical spacing is equal across parts, not objects." In: *Journal of vision* 14.6 (2014), p. 10. DOI: 10.1167/14.6.10.

[69] Ruth Rosenholtz. "Capabilities and Limitations of Peripheral Vision". In: *Annual Review of Vision Science* 2.1 (2016), pp. 437–457. DOI: 10.1146/annurev-vision-082114-035733.

[70] Ruth Rosenholtz. "Texture Perception". In: *Oxford Handbook of Perceptual Organization*. Vol. 167. Oxford University Press, USA, 2014, p. 186. DOI: 10.1016/B978-012375731-9/50081-1.

[71] Ruth Rosenholtz, Dian Yu, and Shaiyan Keshvari. "Challenges to pooling models of crowding: Implications for visual mechanisms". In: *Journal of Vision* 19.7 (July 1, 2019), pp. 15–15. DOI: 10.1167/19.7.15.

[72] Ruth Rosenholtz, Jie Huang, Alvin Raj, Benjamin J. Balas, and Livia Ilie. "A summary statistic representation in peripheral vision explains visual search". In: *Journal of Vision* 12.4 (2012). ISBN: 9788578110796, p. 14. DOI: 10.1167/12.4.14. arXiv: 1011.1669v3.

[73] Toni P. Saarela and Michael H. Herzog. "Size tuning and contextual modulation of backward contrast masking." In: *Journal of vision* 9.11 (2009), p. 21. DOI: 10.1167/9.11.21.

[74] Toni P. Saarela, Gerald Westheimer, and Michael H. Herzog. "The effect of spacing regularity on visual crowding". In: *Journal of Vision* 10.10 (2010), p. 17. DOI: 10.1167/10.10.17.

[75] Ursula Schade and Cristina Meinecke. "Spatial distance between target and irrelevant patch modulates detection in a texture segmentation task". In: *Spatial Vision* 22.6 (Dec. 1, 2009), pp. 511–527. DOI: 10.1163/156856809789822998.

[76] Ursula Schade and Cristina Meinecke. "Texture segmentation: Do the processing units on the saliency map increase with eccentricity?" In: *Vision Research* 51.1 (2011), pp. 1–12. DOI: 10.1016/j.visres.2010.09.010.

[77] Anita M. Schmid and Jonathan D. Victor. "Possible functions of contextual modulations and receptive field nonlinearities: Pop-out and texture segmentation". In: *Vision Research* 104 (2014), pp. 57–67. DOI: 10.1016/j.visres.2014.07.002.

[78] Eero P. Simoncelli, William T. Freeman, Edward H. Adelson, and David J. Heeger. "Shiftable multiscale transforms". In: *IEEE Transactions on Information Theory* 38.2 (1992), pp. 587–607. DOI: 10.1109/18.119725.

[79] M. J. Sinai, W. K. Krebs, R. P. Darken, J. H. Rowland, and J. S. McCarley. "Egocentric Distance Perception in a Virutal Environment Using a Perceptual Matching Task". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 43.22 (Sept. 1999), pp. 1256–1260. DOI: 10.1177/154193129904302219.

[80] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018.

[81] Bernhard Treutwein. "Adaptive psychophysical procedures". In: *Vision Research* 35.17 (1995), pp. 2503–2522. DOI: 10.1016/0042-6989(95)00016-X.

[82] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. New York: Springer, 2002.

[83] Jonathan D. Victor. "Images, statistics, and textures: implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: a comment". In: *J. Opt. Soc. Am.* 11.5 (1994), pp. 1680–1684. DOI: 10.1364/JOSAA.11.001680.

[84] Jonathan D. Victor, Daniel J. Thengone, and Mary M. Conte. "Perception of second- and third-order orientation signals and their interactions". In: *Journal of Vision* 13.4 (Mar. 1, 2013), pp. 21–21. DOI: 10.1167/13.4.21.

[85] Julian M. Wallace, Michael K. Chiu, Anirvan S. Nandy, and Bosco S. Tjan. "Crowding during restricted and free viewing". In: *Vision Research* 84 (May 24, 2013), pp. 50–59. DOI: 10.1016/j.visres.2013.03.010.

[86] Thomas S A Wallis, Matthias Bethge, and Felix A Wichmann. "Testing models of peripheral encoding using metamerism in an oddity paradigm". In: *Journal of Vision* 16.2 (2016), p. 4. DOI: 10.1167/16.2.4.

[87] Thomas S. A. Wallis and Peter J. Bex. "Image correlates of crowding in natural scenes". In: *Journal of Vision* 12.7 (July 2, 2012), pp. 6–6. DOI: 10.1167/12.7.6.

[88] Thomas S. A. Wallis, Christina M. Funke, Alexander S. Ecker, Leon A. Gatys, Felix A. Wichmann, and Matthias Bethge. "Image content is more important than Bouma's Law for scene metamers". In: *eLife* 8 (2019), e42512. DOI: 10.7554/eLife.42512.

[89] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image quality assessment: from error visibility to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (Apr. 2004), pp. 600–612. DOI: 10.1109/TIP.2003.819861.

[90] David Whitney and Dennis M. Levi. "Visual crowding: A fundamental limit on conscious perception and object recognition". In: *Trends in Cognitive Sciences* 15.4 (2011), pp. 160–168. DOI: 10.1016/j.tics.2011.02.005.

[91] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[92] Hadley Wickham, Lionel Henry, and RStudio. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. Version 0.8.1. 2018.

[93] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*. Version 0.7.6. 2018.

[94] Verena Willenbockel, Javid Sadr, Daniel Fiset, Greg O. Horne, Frédéric Gosselin, and James W. Tanaka. "Controlling low-level image properties: the SHINE toolbox". In: *Behav Res Methods* 42.3 (2010), pp. 671–684. DOI: 10.3758/brm.42.3.671.

[95] Jing Xing and David J. Heeger. "Center-surround interactions in foveal and peripheral vision." In: *Vision research* 40.22 (2000), pp. 3065–3072. DOI: 10.1016/S0042-6989(00)00152-8.

[96] Xie Yihui. *Dynamic Documents with {R} and knitr*. 2nd. Boca Raton, Florida: Chapman and Hall/CRC, 2015.

[97] Yunguo Yu, Anita M. Schmid, and Jonathan D. Victor. "Visual processing of informative multipoint correlations arises primarily in V2". In: *eLife* 4 (2015), pp. 1–13. DOI: 10.7554/eLife.0.6604.001.

[98] Elizabeth Zavitz and Curtis L Baker. "Higher order image structure enables boundary segmentation in the absence of luminance or contrast cues". In: *Journal of Vision* 14.2014 (2014), p. 14. DOI: 10.1167/14.4.14.

[99] Barbara Zenger-Landolt and David J. Heeger. "Response suppression in v1 agrees with psychophysics of surround masking." In: *Journal of Neuroscience* 23.17 (2003), pp. 6884–6893. DOI: 10.1523/JNEUROSCI.23-17-06884.2003.

[100] Barbara Zenger-Landolt and Christof Koch. "Flanker effects in peripheral contrast discrimination—psychophysics and modeling". In: *Vision Research* 41.27 (Dec. 1, 2001), pp. 3663–3675. DOI: 10.1016/S0042-6989(01)00175-4.

[101] Corey M. Ziemba, Jeremy Freeman, Eero P. Simoncelli, and J. Anthony Movshon. "Contextual modulation of sensitivity to naturalistic image structure in macaque V2". In: *Journal of Neurophysiology* 120.2 (Aug. 2018), pp. 409–420. DOI: 10.1152/jn.00900.2017.

[102] Corey M. Ziemba, Jeremy Freeman, J. Anthony Movshon, and Eero P. Simoncelli. "Selectivity and tolerance for visual texture in macaque V2". In: *Proceedings of the National Academy of Sciences* 113.22 (2016), E3140–E3149. DOI: 10.1073/pnas.1510847113.

[103] Corey. M. Ziemba, Tim Oleskiw, Richard K. Perez, Eero P. Simoncelli, and J. Anthony Movshon. "Selectivity of contextual modulation in macaque V1 and V2". In: *Annual Meeting, Neuroscience*. 2017.