



Discrimination of natural scenes in central and peripheral vision

M.P.S. To^{a,*}, I.D. Gilchrist^b, T. Troscianko^b, D.J. Tolhurst^a

^a Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

^b School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK

ARTICLE INFO

Article history:

Received 30 April 2010

Received in revised form 12 May 2011

Available online 26 May 2011

Keywords:

Peripheral vision

Crowding

Spatial vision

Visual Difference Predictor model

Colour

Orientation

ABSTRACT

We conducted suprathreshold discrimination experiments to compare how natural-scene information is processed in central and peripheral vision (16° eccentricity). Observers' ratings of the perceived magnitude of changes in naturalistic scenes were lower for peripheral than for foveal viewing, and peripheral orientation changes were rated less than peripheral colour changes. A V1-based Visual Difference Predictor model of the magnitudes of perceived foveal change was adapted to match the sinusoidal grating sensitivities of peripheral vision, but it could not explain why the ratings for changes in peripheral stimuli were so reduced. Perceived magnitude ratings for peripheral stimuli were further reduced by simultaneous presentation of flanking patches of naturalistic images, a phenomenon that could not be replicated foveally, even after M-scaling the foveal stimuli to reduce their size and the distances from the flankers. The effects of the peripheral flankers are very reminiscent of crowding phenomena demonstrated with letters or Gabor patches.

Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Peripheral vision provides a rich source of information in natural scene perception, both for recognition of features and objects (e.g. Henderson, Pollatsek, & Rayner, 1989; Pollatsek, Rayner, & Collins, 1984; Pollatsek, Rayner, & Henderson, 1990; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001) and for appreciation of scene gist (Larson & Loschky, 2009). While inspecting a scene, the decision about where next to direct the eyes is made using visual information from the periphery, and so peripheral vision is particularly important for the identification and localisation of objects for guiding eye movements to task-relevant locations in natural scenes (Torralba, Oliva, Castelhan, & Henderson, 2006).

However, visual ability falls off dramatically away from the fovea so that this important information is relatively crudely encoded. For example, the identification of a single isolated letter falls in a linear manner from 100% at 3° to 50% at 10°. Moreover, when the letter to be identified is flanked ('crowded' – see below) by two other letters, performance is already down to around 80% at just 1° and drops to 35% by 3° when the target and flankers were directly next to one another (Bouma, 1970). This concentration of visual ability in the central part of the visual field is reflected in the over representation of central vision in the cortex. Over 30% of primary visual cortex (V1) represents the central 5° of the visual field (Horton & Hoyt, 1991) and so, conversely, the peripheral visual field is under-represented in V1 (Azzopardi & Cowey, 1993;

Duncan & Boynton, 2003; Popovic & Sjostrand, 2001; Tolhurst & Ling, 1988; van Essen, Newsome, & Maunsell, 1984); i.e. the magnification factor (mm of V1 per deg of visual angle) is greater for foveal than for peripheral vision. In addition, V1 cells processing locations in peripheral vision have larger receptive fields than central ones (e.g. Hubel & Wiesel, 1974), and these cells' contrast sensitivity is biased to lower spatial frequencies (Tolhurst & Thompson, 1981; Yu et al., 2010). This presumably underlies the bias of human peripheral vision towards lower spatial frequencies compared to the fovea (Cannon, 1985; Hilz & Cavanus, 1974; Mullen & Kingdom, 2002; Peli, Yang, Goldstein, & Reeves, 1991; Pointer & Hess, 1989; Robson & Graham, 1981; Rovamo & Virsu, 1979; Wright & Johnston, 1983).

Some have suggested that the cortical magnification factor, neuronal receptive-field size, the density of retinal ganglion cells and measures of visual performance should be directly related (e.g. Hubel & Wiesel, 1974; Rolls & Cowey, 1970; Rovamo & Virsu, 1979). However, the representation of foveal vision in V1 seems to be grossly over-magnified compared to receptive-field size or measures of acuity (Azzopardi & Cowey, 1993; Tolhurst, 1989; van Essen et al., 1984). Differences in spatial scaling between peripheral and central vision alone do not explain the full extent of changes in visual ability with eccentricity – performance for different tasks falls off at different rates with eccentricity (Hess & Field, 1993; Levi, Klein, & Aitsebaomo, 1985; Levi, Klein, & Wang, 1994). Furthermore, targets in peripheral vision but not foveal vision are subject to "crowding" – the phenomenon where targets become more difficult to recognise (rather than to detect) when surrounded by other flanker or distractor stimuli (Bouma, 1970,

* Corresponding author. Fax: +44 (0) 1223 333840.

E-mail address: to@cantab.net (M.P.S. To).

see also reviews by Levi (2008) and Pelli (2008). Like many studies of peripheral vision, crowding has mostly been studied using well-defined stimuli such as Gabor patches or letters, although Martelli, Majaj, and Pelli (2005) have shown crowding by images of faces. In fact, crowding has been associated with the detrimental effect of clutter in complex natural images (Balas, Nakano, & Rosenholtz, 2009; Rosenholtz, Li, & Nakano, 2007), and the inability to recognise complex stimuli in the periphery (Pelli & Tillman, 2008).

There is, thus, a good body of neurophysiological and psychophysical evidence that peripheral V1 is not simply the same as foveal V1 but operating at a lower spatial-frequency scale; peripheral vision differs qualitatively as well as quantitatively from foveal vision. The situation is not, however, fully resolved since it is still often presumed that magnification factor, receptive field size, visual acuity and retinal ganglion cell density are directly related and that changes in these parameters are all that determine the difference between foveal and peripheral vision (Duncan & Boynton, 2003; Geisler, Perry, & Najemnik, 2006; Nugent, Keswani, Woods, & Peli, 2003; Thorpe et al., 2001).

An important step in investigating peripheral visual ability is to investigate visual performance using more naturalistic stimuli, the sorts of stimuli that might guide decisions about gaze in real-world tasks. Firstly, this is important because abilities at threshold do not always scale up to suprathreshold stimuli (e.g. Georgeson & Sullivan, 1975). Secondly, performance with simple stimuli does not necessarily predict performance with complex ones; the responses of V1 simple cells to natural stimuli are only partly explained by their responses to sinusoidal gratings (Smyth, Willmore, Thompson, Baker, & Tolhurst, 2003). It is timely to determine whether our great knowledge of visual performance with stimuli such as sinusoidal gratings is now mature enough to be extrapolated to the complexities of vision in real-world tasks.

One challenge in this area arises because naturalistic stimuli are heterogeneous, so that studies need to include a large set of stimuli in order to ensure that the results are likely to generalise. In addition it is not often possible to repeat the same stimuli many hundreds of times which would be normal in standard psychophysics. Thus, in order to study visual performance with natural images, we have developed a rating method (see To, Lovell, Troscianko, & Tolhurst, 2010) which allows individual images to be compared without the need for large numbers of repetitions. Here we use this method to investigate how visual ability with natural images in peripheral vision compares with foveal vision; for instance, is there evidence of phenomena such as crowding when we view naturalistic stimuli?

The development of this experimental method has allowed us to test our model of early visual processing (The Visual-Difference Predictor (VDP)) with these more naturalistic stimuli (To et al., 2010; Tolhurst et al., 2010) and so develop a model that is both neurally plausible and tested with complex natural images. Here we further test if the model can account for the changes in behaviour in our naturalistic image task with peripheral presentation of the stimuli. In particular, we ask whether a peripheral version of the VDP model will be successful if it is simply a port of the foveal model, differing from it only by virtue of the known differences in contrast sensitivity for sinusoidal gratings (e.g. Zelinsky, 2008), a kind of low-sensitivity ‘blurry’ fovea (Anstis, 1998). We would suspect this to be unlikely, given the existence of phenomena like crowding and the numerical mismatches between, say, magnification factors and receptive field size. However, it is important to know the degree to which such phenomena do apply to viewing of naturalistic stimuli.

Thus, using magnitude estimation ratings, we compare how suprathreshold differences in naturalistic image stimuli are perceived in the central and peripheral visual fields, and we ask whether changes in different domains (such as colour or shape) are differentially compromised in peripheral vision (compare van

den Berg, Roerdink, & Cornelissen, 2007). Second, we show that perception of differences is affected by the presence of flanking patches of natural image in periphery, and we examine how this effect is modulated by the similarity of flankers to the targets. We show that, like “crowding”, our effect is absent in the fovea for naturalistic stimuli (as with other kinds of stimuli – see Levi, 2008; Pelli, 2008), even when we “M-scale” (Párraga, Troscianko, & Tolhurst, 2005; Rovamo & Virsu, 1979) the foveal stimuli to match the peripheral ones. Third, we examine whether VDP models (Lovell, Párraga, Ripamonti, Troscianko, & Tolhurst, 2006; Párraga et al., 2005; Rohaly, Ahumada, & Watson, 1997; To et al., 2010; Watson & Solomon, 1997) can be ported to the periphery, simply by incorporating larger receptive fields and bias towards low spatial frequencies into the peripheral version of a foveal model (Zelinsky, 2008).

We will first present experiments that compare the perceived magnitude of changes in single, isolated stimuli in central and peripheral conditions; the stimuli exhibit various changes in the colour and/or spatial organisation of patches of natural images. Then we will discuss experiments where such stimulus targets are flanked by similar or different patches of natural image. Lastly, we will ask whether shrinking the foveal stimuli (M-scaling them) can simulate the peripheral results. We will consider whether a foveal Visual Difference Predictor model (To et al., 2010) based originally on Watson and Solomon (1997) can be modified to explain the differences between foveal and peripheral viewing.

2. Methods

2.1. Display equipment and stimulus construction

Our general methods have been described (To, Gilchrist, Troscianko, Kho, & Tolhurst, 2009; To, Lovell, Troscianko, & Tolhurst, 2008; To et al., 2010) and we reprise the details in the [Supplementary material](#).

Stimulus images were made from coloured photographs of natural scenes that contained animals, landscapes, objects, people, plants and garden/still-life scenes as described before (To et al., 2008, 2009, 2010). The target stimuli were circular patches cut from such coloured images and were presented in the centre of the display. Observers viewed the stimuli binocularly. Image patch diameters varied between the various experiments (see [Table 1](#)). When viewed foveally, the patches and the contained objects were large enough that their content could be recognised. However, the same-sized patches viewed peripherally were generally not recognisable as images of specific objects, but seemed to be a collection of shapes and colours. The pixels around the stimulus edges were blended into the grey surround, by squeezing the pixel values towards 128 with a Gaussian fall-off. For each experiment, several parent images were chosen and a number of variants of each parent were constructed. Observers were asked to give a magnitude estimate rating of the perceived difference between the target patch and a variant of it; sometimes, as a control, the target was paired against itself unchanged. Some variants were made from a second photograph of the same scene taken when, say, an item had moved; these image changes were entirely natural. Other variants were made from originals using code written in Matlab (The Mathworks). The image pairs were different along one or two of the following feature dimensions (see also [Table 1](#)):

- **Colour:** Image pairs where the variant was a uniform change in Hue or Saturation (see [Fig. 1A and B](#) in [Supplementary material](#)). These changes were performed with images transformed to an HSL space, but were not guaranteed to be psychophysically isoluminant. Hue angle of the whole stimulus patch could rotate by up to 45°, and saturation could change by up to ±45%.

Table 1

Summary of experimental set-up. For each experiment, we list: the number of stimuli in the central (C) and peripheral (P) conditions, the size of the images patches, the types of flankers (*Same*, *Diff1* and/or *Diff2*), the types of changes [Hue, Saturation (Sat), Shape, Flip, Rotate], and the target-flanker distances. In Experiment 1, each of the feature changes was performed with two different magnitudes to give eight variants from the parent.

Experiment (observers)	Number	Size (°)	Flanker	Feature changes	Distances (°)
1 (9)	C: 180 P: 180	3.8	n/a	Hue × 2, Sat × 2, Flip × 2, Rotate × 2	No flankers
2 (7)	C: 343 P: 343	5.2	<i>Same</i> , <i>Diff1</i>	Hue, Sat, Shape, Hue + Sat, Hue + Shape, Shape + Sat	5.2, 6.6, 8.2
3 (9)	C: 300 P: 300	3.8	<i>Same</i> , <i>Diff2</i>	Hue, Sat, Flip, Rotate	3.8
4 (8)	P: 420	5.2	<i>Same</i> , <i>Diff1</i> , <i>Diff2</i>	Hue, Sat, Shape, Hue + Sat, Hue + Shape, Shape + Sat	5.2, 6.6, 8.2
5 (8)	C: 343	1.5	<i>Same</i> , <i>Diff1</i>	Hue, Sat, Shape, Hue + Sat, Hue + Shape, Shape + Sat	1.5, 2.25, 3

- **Shape:** Image pairs where objects in the variant changed shape, posture or position; the image patches were cut from equivalent parts of two photographs of the same scene, taken at different times. Such changes obviously involved spatial changes in the target, but they might also have changed the amounts of differently coloured objects that were seen (e.g. Fig. 1C in Supplementary material).
- **Orientation:** Image pairs where the variant was a horizontal or vertical reflection of the parent image (Flip; Fig. 1D and E in Supplementary material), or a rotated transform of the parent (Rotate; Fig. 1F in Supplementary material). For the latter, the orientation of the whole circular patch could be rotated by up to $\pm 20^\circ$. The overall chromatic content of these variants was unchanged from the parent.

The magnitudes and ranges of these changes were chosen to ensure that there was a wide range of difficulty in judging a perceptual difference; some stimuli were difficult while others would have had very obvious changes. The orientation and colour changes were chosen so that, roughly, the ratings would cover the same ranges for the two change types. However, the magnitudes of the colour or shape changes do not directly inform on the likely ratings given (Tolhurst et al., 2010), since the basic stimulus patches were different in their initial colour and shape content.

In an experiment, the many stimulus pairs would have been derived from, say, seven original natural-image patches. Each patch was subject to the full variety of colour, shape or orientation changes being studied; usually, there were 7–8 variants of the original image, one of which was unchanged. Thus, one in 7–8 of the stimulus trials would have been of a pair with no difference. Each observer saw and rated each of these image pairs once foveally and once peripherally.

The target patches and their variants were either presented alone or amongst four flankers (unflanked vs. flanked conditions, respectively). Examples from Experiment 4 are presented in Fig. 1 (and in Fig. 2 in Supplementary material). In the flanked conditions, the target-flanker similarity and distance were varied between experiments (Table 1). The four flankers could be identical to the parent target of the current comparison-pair (*Same*, Fig. 1B), different from the target but identical between themselves (*Diff1*, Fig. 1C) or different from the target and different between themselves (*Diff2*, Fig. 1D). The target-flanker distances were measured from target centre to flanker centre. In most experiments, three distances were considered (but see Table 1).

2.2. Stimulus presentation protocol

In a trial, one target patch would be compared with itself or with one of its variants. The images were presented sequentially (To et al., 2008, 2009, 2010) for a duration of 833 ms each. Between

presentations was an inter-stimulus interval (ISI) of 83 ms. There were three presentation intervals: first image, second image and first image again. The parent image and its variant were randomly assigned as the first or second image. Only the target patch could be different during these presentations; any flankers, if present, would be the same in the three presentations (see Fig. 2 in Supplementary material).

The image patches could be presented in the central or peripheral field (see Table 1). In the central condition, observers gazed directly at the test images at the centre of the screen. In the peripheral condition, observers kept their gaze on a fixation dot in the top left end of the display ($\sim 7^\circ$ from top left corner) so that the target images (that were always presented at the centre of the screen) were located at 16° eccentricity in their lower right visual field (see Fig. 1). A fixation point was presented before and throughout each trial in the peripheral condition, but the point disappeared during image presentation in the central condition to ensure that it did not act as a marker to any small object shifts near the centre of the stimulus.

At the end of the trial, the observer made a numerical magnitude estimation rating of the perceived difference between the two target images (Gescheider, 1997, chap. 11, see To et al. (2010) for details). Observers were free to give a rating of zero if they perceived no difference in the stimuli since identity in a pair is a natural lowest bound, and the ratings for all other stimuli were made by comparison with a standard stimulus whose difference was deemed to be '20' (see Section 2.3). There was no upper limit to ratings, since there was no natural stimulus configuration that could be deemed more different than any other and since the observers were not specifically trained to recognise some image change combination as the greatest possible. In our experiments, the differences between image pairs were relatively small, and it would always have been possible to make another image pair that was more different than any that an observer judged to have the highest rating. In practice few stimuli ever received a rating > 60 , but different observers were differently exuberant in awarding high numerical ratings so that the standard errors tended to be higher for high ratings when we averaged all the observers' ratings together (To et al., 2010). A random number between 10 and 30 appeared at fixation, and the observer modified this number using a CRS CB6 response box until their choice of difference rating was reached. Image pairs were presented in a different random order for each observer.

2.3. Observer training and instructions, ratings and experimental design

Before an experiment, each observer underwent a training session, when they were asked to rate 40–60 image pairs of the various types of image differences that could be presented to them

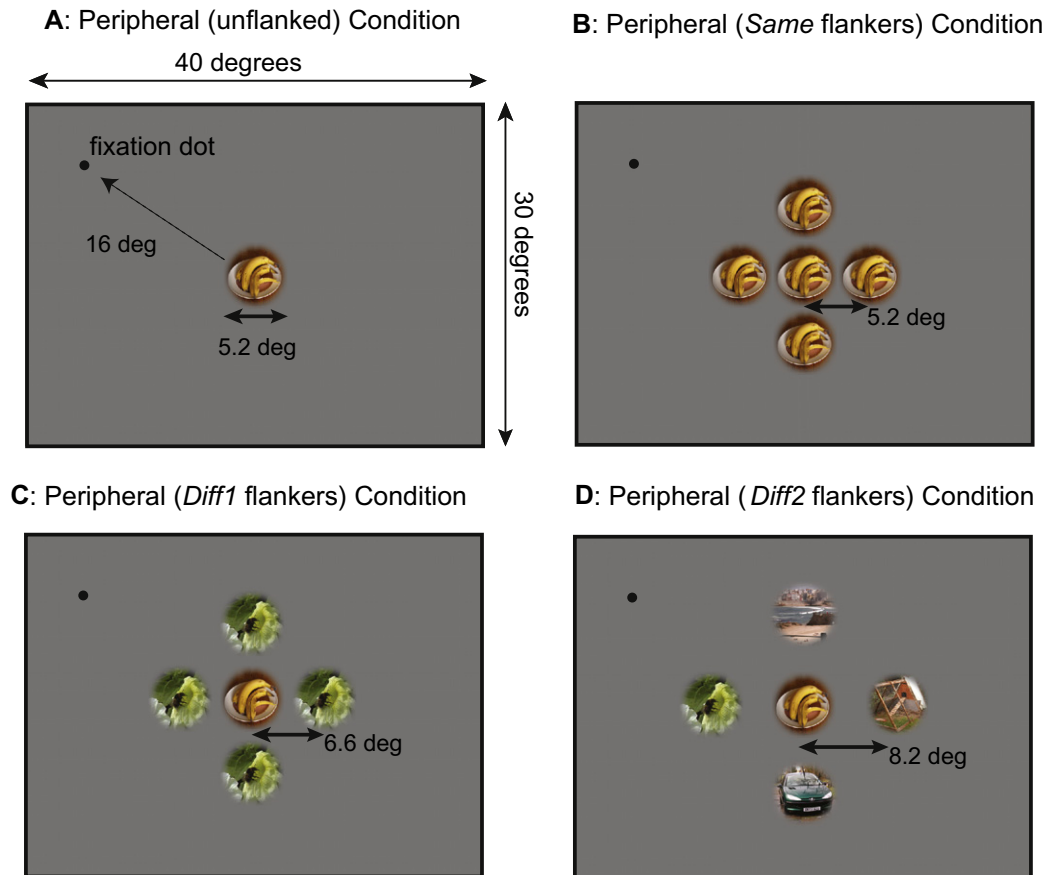


Fig. 1. Examples of parent stimuli for the peripheral condition in Experiment 4. This figure is monochrome; a coloured version of this figure can be found as [Supplementary Fig. 2 in the Supplementary material](#). (A) Target presented alone in the unflanked condition. A fixation point was located at the top left of the display, so that the patch was at 16° eccentricity. The width of each patch was 5.2°. Panels (B–D) present the same target in the different flanked conditions: *Same* flankers at near distance, *Diff1* flankers at mid distance, and *Diff2* flankers at far distance, respectively.

later. During both the training and the testing phases, observers were frequently presented with one particular *standard image pair* (see bottom of [Fig. 1 in Supplementary material](#)) whose magnitude difference was defined (the modulus) as '20'. Observers were instructed that their ratings of the subjective difference between any other image pair must be based on this standard pair using a ratio scale, even when the test pairs differed along different stimulus dimensions from the standard pair (details of the instructions and training are in [To et al. \(2010\)](#)). However, this technique is subject to context and range effects ([Gescheider, 1997](#)), such that the rating given to a particular stimulus difference may be influenced by what other stimulus differences are studied in the same experiment. It is fair to compare ratings given to different stimuli or stimulus conditions within one experiment, but similar stimuli may evoke different ratings in other experiments.

All images (apart from the standard image pair) used in demonstration or training phases were different from those in the testing phase, which was divided into blocks of 100–150 image pairs, each lasting 30–45 min. Each block started with the presentation of the standard image pair, which was subsequently presented after every 10 further trials to remind the observers of the standard difference of '20'.

We chose to collect magnitude estimation ratings instead of change-detection thresholds for two main reasons. First, our experience of collecting thresholds with natural image stimuli is that it can be extremely time consuming and only limited types of stimulus change can be studied ([Lovell et al., 2006](#); [Parraga et al., 2005](#)); we wanted to measure the perceived differences between as many image pairs as possible (686 pairs in Experiment 2 alone)

to allow generalisability of our results. Second, we are interested in realistic suprathreshold differences, such as those we are presented with in everyday situations.

Exactly the same standard image pair was used in both central and peripheral viewing conditions, and was always presented at fixation. When observers were trained to rate the peripheral images, they were told specifically to rate the difference as they perceived them in the periphery and *not* to imagine how different the targets might appear had they been presented at fixation (see [Supplementary material](#)). The central and peripheral conditions were blocked (about 100 trials per block) so that observers would always know whether the next stimulus pair would be presented foveally or peripherally (the peripheral stimuli were always presented at the same location). This differs from a deliberate change blindness paradigm, where there is uncertainty as to where the peripheral change might occur. Half the observers would have done a foveal block of 100 or so trials first, and half would have done a peripheral block first.

Observers were students at the University of Cambridge, UK. They all gave informed consent. They had normal vision after prescription correction, as verified using a Landolt C acuity chart and the Ishihara colour test (10th Edition). While some observers participated in more than one of our experiments, they remained naïve to the purpose of each. [Table 1](#) lists the number of observers.

2.4. Data collation and statistical analysis

In each experiment, between seven and nine observers rated each stimulus once. The ratings given by the several observers

for each stimulus were averaged together to give means and standard errors from seven to nine measurements. The ratings of each observer were, in fact, normalised by dividing by that observer's average or median rating within the experiment; the foveal and peripheral ratings were normalised separately (see [Supplementary material](#)). The averages were then multiplied by a scalar (the grand average of all the observers' ratings to all the stimuli) so that the presented data roughly centre on the standard value of '20' for foveal viewing (lower for peripheral viewing) rather than on unity. The data in the graphs thus show the mean of the ratings given to each stimulus, averaged across observers. The standard error of the mean rating was typically about 2.0 for foveal viewing (To et al., 2010), but was higher (2.8–3.0) for peripheral viewing of flanked or unflanked stimuli. Standard errors tended to be higher for the higher averaged ratings (see above; To et al., 2010). For foveal viewing, each observer's ratings were typically well correlated with those of other observers ($r = 0.7$ – 0.8) and with the average of the observers' ratings ($r = 0.9$). With peripheral viewing, the observer's ratings were less well correlated with other observers ($r = 0.45$ – 0.55) or with the average observer ratings ($r = 0.65$ – 0.75). Observers clearly found it harder to make consistent magnitude estimate ratings with peripheral vision.

The presence of flankers generally reduced the ratings given for peripherally-viewed targets, a phenomenon which may be the same as crowding. A common measure for the amount of crowding is threshold elevation computed as the ratio between crowded and uncrowded condition (e.g. Levi, Hariharan, & Klein, 2002; Pelli, Palomares, & Majaj, 2004; van den Berg et al., 2007). However, simple ratio measurements were unsuitable for the present discrimination tasks because observers' ratings could reasonably take the value of zero when they perceived no difference between the target pairs (often the case in peripheral viewing). One alternative would be to take the difference between ratings for images presented alone and amongst distractors, but this metric would not adequately allow for likely differences for image pairs containing large vs. small perceived differences. We have therefore (see [Supplementary material](#)) fitted a least-squares regression line constrained to go through the origin $[0, 0]$ to the ratings for targets presented among flankers plotted against the ratings alone (as in Fig. 5). The slope m of this regression is effectively an average 'crowding' ratio, where $m < 1$ indicates crowding.

2.5. Porting a V1-based Visual Difference Predictor (VDP) model to peripheral vision

We have developed a model to predict thresholds and supra-threshold perceived magnitudes in stimuli derived from natural images. The organisation of our present model (To et al., 2010) follows Watson and Solomon (1997) and Rohaly et al. (1997). Basically, the model of foveal magnitude ratings attempts to calculate how millions of V1 "neurons" would respond to each of the images under comparison in a pair, subjects those responses to suppressive nonlinearities and then it compares the paired responses neuron by neuron; finally the many response differences are pooled into a single number, which we postulate is proportional to the magnitude rating. The details are given in To et al. (2010) and Tolhurst et al. (2010), while the main steps in the model are given in the [Supplementary material](#). It is well-known that contrast sensitivity to gratings is lower in peripheral vision and is shifted towards lower spatial frequencies (Cannon, 1985; Hilz & Cavanaugh, 1974; Mullen & Kingdom, 2002; Peli et al., 1991; Pointer & Hess, 1989; Robson & Graham, 1981; Rovamo & Virsu, 1979; Wright & Johnston, 1983). We ported our foveal VDP model to an eccentricity of 16° by, essentially, modelling this difference in sensitivity to sinusoidal gratings in order to see whether this

fundamental difference in grating sensitivity can be an explanation for differences in foveal and peripheral vision.

For the foveal version of the model, stimulus images were cropped to 512 by 512 pixels centred on the target stimulus, and were convolved with 60 elongated Gabor receptive-fields with Gaussian envelope that was 1.5 times longer along the axis of preferred orientation (Foley, Varadharajan, Koh, & Farias, 2007; Tolhurst & Thompson, 1981). At each of six orientations (30° steps), there were five spatial frequencies ranging from 8 to 128 cycles across the 512 pixel stimuli, with even- and odd-symmetric fields for each of the 30 bands. The fields were elongated and were not self-similar: neuronal spatial frequency and orientation tuning gets sharper as spatial frequency is increased (Tolhurst & Thompson, 1981; Yu et al., 2010). The bandwidths were graded with optimal frequency: for fields with optima of 0.31, 0.62, 1.25, 2.5, 5 cycles per deg, the frequency bandwidths were 2.12, 1.43, 0.93, 0.64, 0.45 octaves and the orientation bandwidths were 43.4° , 34.5° , 22.8° , 17.7° , 11.6° . Now, V1 receptive fields are about three times larger at 16° periphery than in central vision (e.g. Hubel & Wiesel, 1974; van Essen et al., 1984), and there is a shifted, parallel relationship between spatial frequency bandwidth and optimal spatial frequency (Tolhurst & Thompson, 1981; Yu et al., 2010). Our model therefore needs to implement receptive fields with three times lower centre frequencies (of 0.1, 0.2, 0.4, 0.8, 1.6 cycles per deg), but with the same bandwidths as in the fovea (i.e. 2.12, 1.43, 0.93, 0.64, 0.45 octaves, and 43.4° , 34.5° , 22.8° , 17.7° , 11.6°). We cropped 600 pixel high by 768 pixel wide segments of the stimulus images (centred on the target patch) and padded the top and bottom 84 pixels with uniform grey. We then sub-sampled the 768 by 768 image by 1-in-3 rows and by 1-in-3 columns to give a 256 by 256 image for analysis. The receptive-field templates occupied the same number of pixels as in the foveal version of the model, so that the centre frequencies now ranged from 4 to 64 cycles across these (smaller) 256 pixel images. The 1-in-9 sub-sampling meant that, all other things staying equal, the model output would be reduced by a factor of 9^{-M} (where M is the Minkowski exponent); we compensated for that.

All the parameters of the peripheral model were the same as those optimised for the foveal model (see [Supplementary material](#)). The only important difference was that the spatial point of maximum grating sensitivity (the "fovea") was in the top right of the image map, rather than in the centre. The fall-off in sensitivity to sinewave gratings from this foveal point was modelled after Pointer and Hess (1989) and Mullen and Kingdom (2002) as described in the [Supplementary material](#). We measured contrast thresholds for luminance gratings 16° peripherally, and used those values (along with chromatic estimates from Mullen and Kingdom (2002)) to ensure that the model predicted the thresholds for peripherally-viewed gratings in the same way that the foveal model predicted foveal grating thresholds; i.e. the peripheral model's numerical output for peripherally viewed gratings at threshold had to be the same as the foveal model's output for a foveally-viewed grating at threshold.

3. Experimental results

3.1. Central vs. peripheral ratings

Experiments 1, 2 and 3 collected difference ratings in both central and peripheral regions of the visual field, allowing direct stimulus-by-stimulus comparison of the ratings. Particular image changes presented in the central condition were more salient compared to when the same changes were presented in the periphery. Overall, observers' ratings for central changes were higher than those for peripheral changes [$F(1, 1644) = 132.07$, $p \approx 0$]. Fig. 2

plots the ratings given to feature changes presented in the unflanked condition; the peripheral ratings are plotted against the corresponding central ratings. Each point is the average of the ratings for a specific stimulus given by the observers participating in the experiment; typical standard errors were 2.0 for foveal viewing and 3.0 for peripheral viewing. The cluster of points near the origin (0, 0) represent small difference ratings for image pairs that contain subtle differences or image pairs with no difference that served as controls in each experiment.

The majority of data points are beneath the identity line in Fig. 2, demonstrating a significant trend for lower peripheral ratings, and regressions through the data for the three experiments all have slopes significantly less than unity (see figure legend): changes in natural scenes are less visible and/or less salient when viewed eccentrically. These findings are consistent with our earlier study (To et al., 2009). Results for three experiments are shown in Fig. 2, and the slopes of the three regressions are all significantly less than 1.0 but the slopes are different from each other (ranging from 0.41 to 0.8). There are two possible reasons for this discrepancy. First, we show below (Fig. 4) that different kinds of stimulus change (e.g. colour vs. shape) are affected differently by peripheral viewing, and the different stimulus sets in the three experiments may have differed in their balance between different stimulus changes. Second, magnitude estimation rating experiments are unfortunately subject to context and range effects (Gescheider, 1997), so that the ratings given to a particular stimulus difference may be different in different experiments.

Lower ratings for peripheral stimuli might come as no surprise given the decreased sensitivity to simple sinusoidal gratings in the periphery (Cannon, 1985; Hilz & Cavanaugh, 1974; Mullen & Kingdom, 2002; Peli et al., 1991; Pointer & Hess, 1989; Robson & Graham, 1981; Rovamo & Virsu, 1979; Wright & Johnston, 1983), although it is not necessarily the case that low grating sensitivity

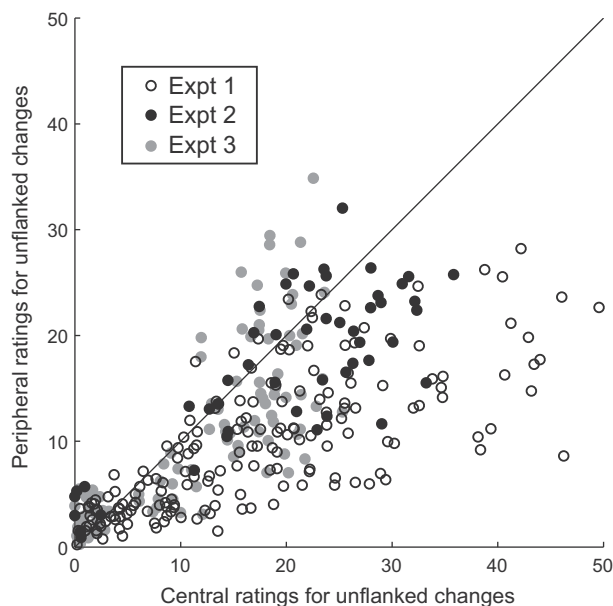


Fig. 2. Comparing ratings for differences between pairs of natural scene patches presented alone (unflanked) at fixation and at 16° eccentricity. The peripheral ratings are plotted against the central ratings. Results from Experiments 1, 2 and 3 are shown in white, black and grey, respectively. The majority of data points are located beneath the identity line (solid line), meaning that differences in the periphery are less salient than foveal ones. A two-parameter least squares regression through the whole data set had slope of only 0.49 (standard error SE = 0.27; $n = 328$) and was highly significantly different from 1 [$t(326) = 18.39$; $p \approx 0$]. In Experiment 1, the slope was 0.41 (95% confidence 0.35–0.46); in Experiment 2, 0.61 (95% confidence 0.47–0.75); and in Experiment 3, 0.80 (95% confidence 0.67–0.93).

should affect the percept of suprathreshold stimuli. Thus, in order to determine whether the fall in peripheral magnitude estimation ratings is due simply to this difference in peripheral grating sensitivity, we have compared the ratings to the predictions of a foveal and a peripheral VDP model based on the behaviour of V1 neurons (see Section 2).

Fig. 3 plots the observed magnitude ratings against the predictions of the appropriate version of the model for the cases where the target stimuli were presented without flankers. The model predictions and observers' ratings were well correlated in the central condition ($r = 0.71$ – 0.83 , depending on the experiment; Table 2 and Fig. 3A). When we ported the VDP model to the periphery by implementing the lower peripheral contrast sensitivities and shift towards lower spatial frequencies (Mullen & Kingdom, 2002; Tolhurst & Thompson, 1981, see Section 2), the correlations between model predictions and observers ratings dropped to only 0.52–0.78 (Fig. 3B). The peripheral model did generate lower predictions in general but the slopes of the regressions in the two parts of Fig. 3 are clearly not the same (see legend for tests of significance), i.e. the peripheral ratings are consistently less than is predicted by the V1-based model.

3.1.1. Orientation vs. colour changes

Experiment 1 examined potential differences in observers' perception of colour (saturation and hue) and orientation (reflection and rotation) changes; the latter involved flipping or rotating the image patches, and so there were changes in spatial arrangement of features *without* any change in the overall chromaticity of the patches (see Section 2). Fig. 4A plots the peripheral and central ratings against each other, the regression slope for colour changes was larger than the slope for orientation changes, suggesting that changes in orientation become comparatively more difficult to perceive at higher eccentricities. The slopes of the two-parameter regressions were 0.71 (standard error SE = 0.057, $n = 80$) for the colour change ratings and 0.43 (SE = 0.034, $n = 80$) for the orientation change ratings. These slopes are significantly different [$t(156) = 4.09$; $p \approx 0$].

Ideally, we would have been able to choose stimuli where the ratings for colour changes covered the same range as the ratings for orientation changes. Given that colour and orientation seem to be rated differently in the fovea and peripherally, this cannot be achieved completely. For peripheral vision at least, the ratings were roughly in the same range (0–25, compared to the foveal standard of value '20').

Fig. 4B compares the foveal ratings for colour and orientation changes with the predictions of the foveal VDP model; Fig. 4C similarly compares the peripheral ratings with the peripheral VDP. Both the peripheral regressions (Fig. 4C) are shallower than the foveal ones (Fig. 4B) as expected from Fig. 3. The model seems to be more consistent in the fovea in that the two foveal regressions (Fig. 4B) are rather similar. However, in the periphery (Fig. 4C) the lines are more divergent. A two-factor ANCOVA showed that the four regressions in Fig. 4B and C were significantly different. Foveal lines were steeper than peripheral [$F(1, 315) = 216.0$, $p \approx 0$] as expected from Fig. 3. The two lines for colour changes were steeper than those for orientation changes [$F(1, 315) = 49.1$, $p \approx 0$]. Interestingly, the interaction between the two factors was also significant [$F(1, 315) = 41.4$, $p \approx 0$], implying that peripheral ratings for orientation changes (open circles, thin line) are even more discrepant from the VDP model than are colour changes.

3.2. Unflanked vs. flanked conditions

3.2.1. Central viewing

In Experiment 2, targets were presented alone or among four flankers that could either be similar to the targets (*Same*) or

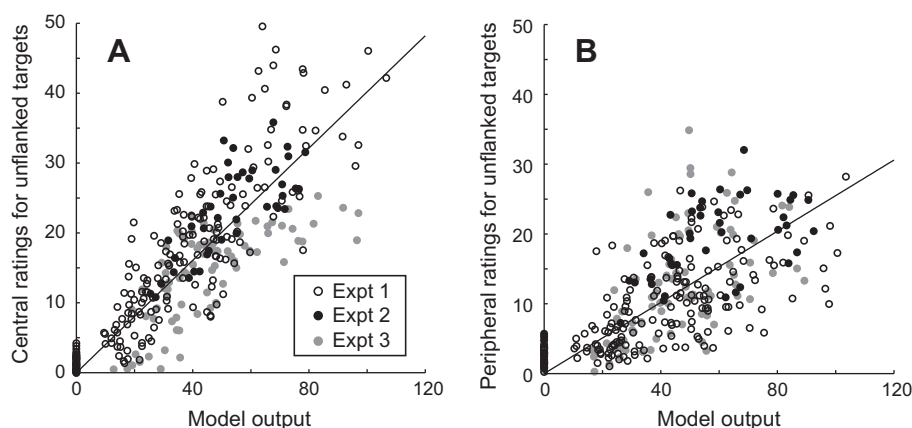


Fig. 3. Observers' ratings vs. model predictions. Panels (A and B) plot observers' ratings against model predictions from the central and peripheral conditions, respectively. The model's numerical output is a number which we postulate should be directly proportional to the observer's ratings. Data from Experiments 1, 2 and 3 are presented in white, black and grey data points. The regression lines (constrained to pass through the origin) are shown in black. For conventional two-parameter regressions, the foveal slope was 0.37 (standard error SE = 0.013, $n = 328$) and the peripheral slope was less at 0.21 (SE = 0.012, $n = 328$), showing that observers' peripheral ratings were generally much lower than predicted by the model. The slopes are significantly different [$t(652) = 9.03$; $p \approx 0$].

Table 2

The correlation coefficients between model predictions and the averages of the observers' ratings for target pairs presented in the central and peripheral conditions. For each experiment, we show r values for the subset of targets that were presented alone, as well as an r value for the whole set of flanked and unflanked stimuli.

Expt.	Unflanked vs. model (central)	All vs. model (central)	Unflanked vs. model (peripheral)	All vs. model (peripheral)
1	0.86	n/a	0.67	n/a
2	0.91	0.83	0.78	0.78
3	0.85	0.81	0.68	0.64
4	n/a	n/a	0.51	0.52
5	0.77	0.71	n/a	n/a

different from the targets but similar between themselves (*Diff1*). The results for the central condition are presented in the top part of Fig. 5. The graphs plot the ratings for target changes in the Flanked condition (at three distances of the flankers) against the control rating in the Unflanked (alone) condition. Fig. 5A and B displays data from the *Same* and *Diff1* flanker conditions, respectively. The solid lines are the lines of identity. The dashed regression lines constrained to pass through the origin were almost the same as the lines of identity; their slopes are a measure of the amount of interaction between target and flankers (see Section 2). When the targets were presented in the central field, observers' ability to perceive feature differences between target pairs was not affected by flankers, i.e. ratings for target pairs presented alone were similar to those for target pairs presented among flankers (Fig. 5A and B; slope of the constrained regression line $m \approx 1$, see Table 3). If this interaction has the same origin as the crowding seen with letters of Gabor patches, then this is consistent with previous research that has shown little or no crowding in foveal vision (Strasburger, Harvey, & Rentschler, 1991). Target-flanker distance had no effect, even in the nearest conditions when the centre-to-centre length was equal to the targets' diameter (black data points). The degree of target-flanker similarity or the degree of similarity amongst flankers also made no difference (compare Fig. 5A and B). These negative results were confirmed in all experiments that measured the effects of flankers in central viewing, where the target size ranged between 1.5° and 5.2° .

3.2.2. Peripheral viewing

Fig. 5C and D shows that, when the natural-image targets were presented with flankers in the periphery (16° eccentricity), there

was a noticeable reduction in the magnitude estimation ratings (dashed regression lines in Fig. 5C and D; $m < 1$, see Table 3). This is similar to reports of crowding in the periphery with other kinds of stimuli. The results also show that the effects of flankers were similar at different flanker-target distances in the range we have used (up to 8.2°). The extent of crowding is greater when the flankers are similar to the target (*Same*; Fig. 5C), and in many conditions, there is little effect when flankers differ from the targets (e.g. *Diff1*; Fig. 5D). The dashed lines in Fig. 5 represent the regression lines for the near target-flanker distances; it is clear that the slope for the *Same* flankers is lower than for the *Diff1* flankers ($m = 0.81$ and 0.94 , respectively). These findings are consistent with previous reports of similar crowdings being more detrimental than dissimilar ones (Kooi, Toet, Tripathy, & Levi, 1994; To et al., 2009).

The amount of interaction between flankers and target was generally small, even in the peripheral field, and we wondered whether the stimulus design meant that the shortest centre-to-centre target-flanker distance (being the diameter of the image patches) was too large to observe the full effects of flankers. The sizes of image patches were reduced in Experiment 3 to study smaller target-flanker distances. The results in Experiment 3 demonstrate that, when the distance between target and distractor images was reduced, there was still no interaction in central viewing ($m \approx 1$), but the amount of interaction did increase substantially in the periphery [$m = 0.48$ and 0.61 in the conditions when the flankers were identical to the target (*Same*) and when the flankers were different from the target and different between themselves respectively (*Diff2*); see Table 3].

Experiment 4 examined specifically the role of target-flanker similarity by randomly interleaving the three flanker types within a single experiment. In our experiment, targets could be surrounded by four flanker patches that were similar to the target (*Same*), different from the target but identical between themselves (*Diff1*) or different from the target and from each other (*Diff2*) to study whether there are any differences between the different flanked conditions (see Fig. 1B–D for examples). The results are presented in Table 3. As in Experiment 2, the amount of target-flanker interaction was largest when target and flankers were similar (*Same*). The results also demonstrate slightly more interaction in cases where target-flankers are completely different (*Diff2*) compared to when flankers are similar between themselves (*Diff1*; see Table 3). In other crowding experiments, *Diff1* flankers have been shown to group with each other, and not group with the target, thus reducing crowding (Livne & Sagi, 2007; Saarela, Sayim,

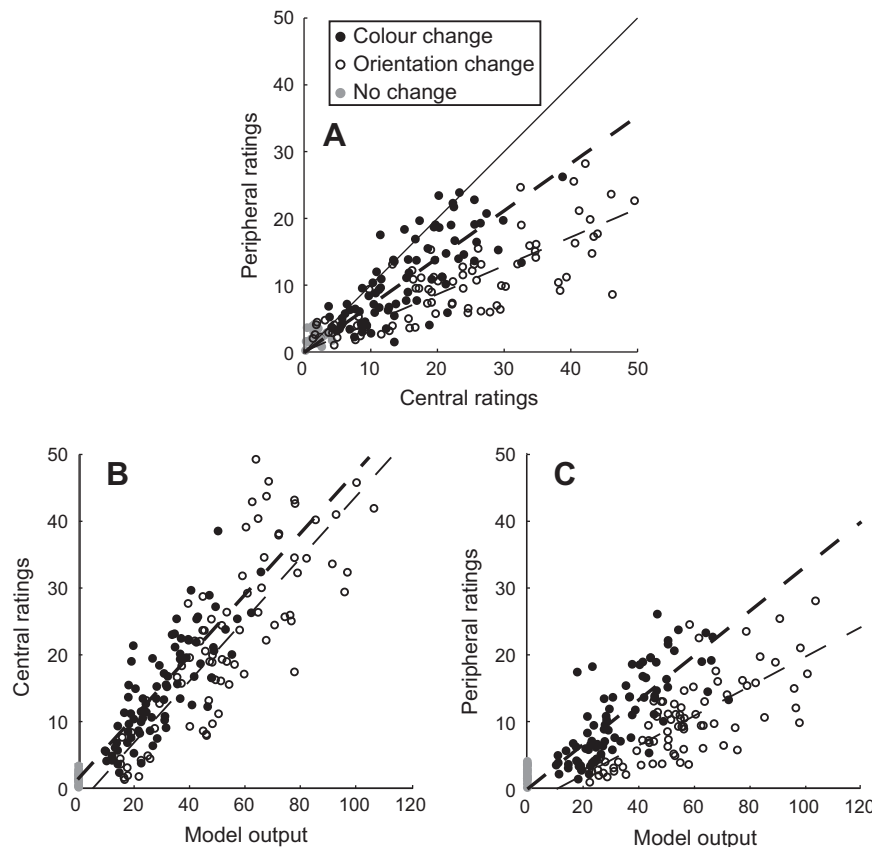


Fig. 4. Results from Experiment 1. Panel (A) compares ratings for differences between pairs of patches of natural scenes presented at fixation and at 16° eccentricity. The peripheral ratings are plotted against the central ratings. The ratings for colour and orientation changes are presented by the black and white data points, respectively; the ratings for control targets pairs with no change are clustered near the origin (grey). The thin solid line is the line of equality; the thin and thick dashed lines represent the regression lines (constrained to pass through the origin) for the orientation and colour changes, respectively. The slopes of these regressions show that changes in orientation are comparatively less salient compared to changes in colour ($m = 0.71$ vs. 0.43). Observers' central and peripheral ratings are plotted against model predictions in panels (B and C), respectively. Note that the regression lines in (B and C) are not constrained to pass through the origin; they result from the ANCOVA.

Westheimer, & Herzog, 2009) compared to *Diff2* flankers. Our results seem to be consistent with this.

3.2.3. M-scaling the foveal stimuli

It has been argued that reduced visual performance in peripheral vision might be compensated by “M-scaling” the stimuli (Rovamo & Virsu, 1979), i.e. by increasing the sizes of peripheral stimuli in inverse proportion to some measure of cortical magnification in the periphery. However, it is well documented that crowding with, e.g., letters and Gabor patches does not occur at all with foveal vision (Bouma, 1970; Levi, 2008; Pelli & Tillman, 2008). Thus, it is natural to ask whether the difference in interaction between flankers and target that we have found for central and peripheral vision is related to differences in cortical magnification; will an appropriate reduction in the sizes and target-flanker distances for foveal targets result in as much target-flanker interaction as the larger peripheral targets? However, the size-scaling factor for a given peripheral eccentricity can be calculated in different ways, depending upon what assumptions are made about retinal and cortical anatomy, and how those factors relate to visual performance (Rovamo & Virsu, 1979; Tolhurst & Ling, 1988; Tolhurst & Thompson, 1981). Furthermore, different scaling factors are needed for different tasks, and not all tasks can be compensated (Bijl, Koenderink, & Kappers, 1992; Levi et al., 1985; Párraga et al., 2005; Saarinen, 1988).

We have already seen that a reduction of the nearest flanker distance to 3.8° (Experiment 3, Table 3) produced no crowding in foveal fixation. Experiment 5 examined whether further reducing

the size of target patches at foveal fixation could simulate the crowding effects reported in the periphery. The patches used were 1.5° wide and were shrunk versions of the stimuli used in Experiment 2; the flankers could be presented correspondingly closer. The foveal stimulus sizes and distances were thus reduced by a factor of 3.87 compared to the peripheral stimuli of Experiments 2 and 4. This factor is smaller than would be proposed by Rovamo & Virsu's (1979) M-scaling algorithm, but is consistent with the factor of 3 that we argue (see Section 2) is an appropriate scaling of receptive field size in the VDP model. Table 3 shows that, despite the reduction in size, the presence of very close flankers did not reduce the ratings for small foveal stimuli ($m \approx 1$), demonstrating that M-scaling alone is unlikely to explain the lower ratings when target patches are presented among flankers in the periphery. This, again, implies that the effects of flankers in our peripheral experiments are revealing phenomena like crowding.

3.2.4. Modelling ratings in the flanked conditions

As shown above, the VDP model was reasonably well correlated with the actual observers' ratings for unflanked target stimuli (e.g. Fig. 3A); the correlation coefficients range was 0.77–0.91 for foveal targets (see Table 2). However, the models' performance was diminished when their output was compared with *all* the stimuli in an experiment, i.e. the flanked stimuli as well as the unflanked ones (Table 2). In the worst cases, the correlation between model and ratings was as low as 0.52 for the peripheral VDP model. This implies that the relatively poor performance of the peripheral

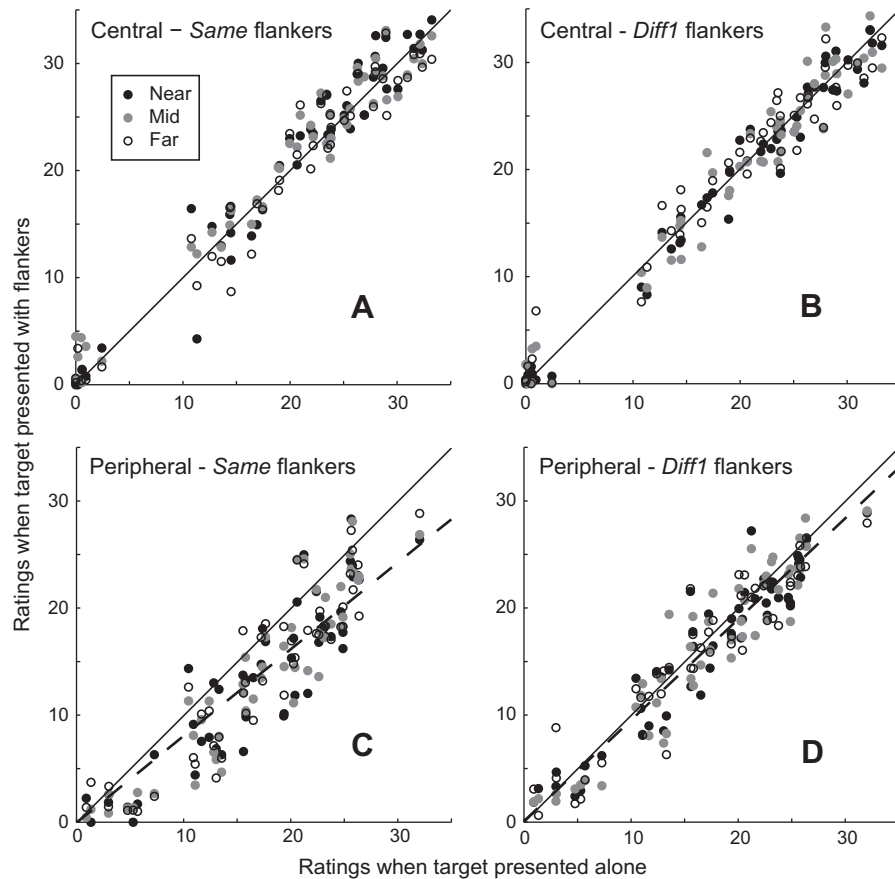


Fig. 5. Results from Experiment 2. The top (A and B) and bottom (C and D) show data from the central and peripheral conditions, respectively. The *Same* flanker conditions are on the left (A and C) and the *Diff1* flankers conditions are on the right (B and D). The different target-flanker distances 5.2°, 6.6° and 8.2° are represented by the black, grey and white data points, respectively. The solid lines are the lines of identity. In the central condition, the presence of flankers made no difference to observers' ratings; the degree of target-flanker similarity or distance (compare A and B). In the peripheral condition, observers' ratings were significantly lower when targets were surrounded by *Same* flankers (C) compared to the *Diff1* flankers (D) [$t(146) = -9.72, p \approx 0$]; the dashed line shows the lower regression slope for near target-flanker distances in the periphery ($m = 0.81$ and 0.94 in graphs (C and D)).

Table 3

Slope m of the regression line of ratings for targets presented among flankers plotted against the ratings alone; $m < 1$ indicates an interaction like crowding (alone ratings > flanked ratings) while $m \approx 1$ signifies little or no crowding (alone ratings \approx crowded ratings). Slopes for the *Same*, *Diff1* and *Diff2* flanker conditions are labelled S, D1 and D2, respectively. The bold numbers are measured slopes, the ones in parentheses are the estimated standard errors of the slopes, and the italic ones below are model predictions. The distances represented by 'near', 'mid' and 'far' vary between experiments; but, within an experiment, the distances were the same in the central and peripheral conditions.

Expt.	Central			Peripheral		
	Near	Mid	Far	Near	Mid	Far
2	S: 1.03 (.014) 0.84 D1: 0.99 (.010) 0.85	S: 1.01 (.013) 0.93 D1: 0.99 (.014) 0.91	S: 1.00 (.015) 0.99 D1: 0.99 (.014) 0.95	S: 0.81 (.026) 0.79 D1: 0.94 (.018) 0.78	S: 0.83 (.024) 0.79 D1: 0.95 (.020) 0.78	S: 0.83 (.024) 0.81 D1: 0.94 (.019) 0.83
3	S: 0.98 (.016) 0.74 D2: 0.97 (.015) 0.66	n/a	n/a	S: 0.49 (.022) 0.74 D2: 0.61 (.026) 0.69	n/a	n/a
4	n/a	n/a	n/a	S: 0.87 (.022) 0.72 D1: 1.01 (.020) 0.75 D2: 0.95 (.020) 0.69	S: 0.89 (.019) 0.73 D1: 0.99 (.014) 0.78 D2: 0.96 (.020) 0.78	S: 0.94 (.021) 0.74 D1: 1.01 (.018) 0.81 D2: 0.98 (.022) 0.84
5	S: 1.00 (.017) 0.85 D1: 0.98 (.018) 0.89	S: 1.00 (.019) 0.88 D1: 1.01 (.018) 0.91	S: 0.98 (.020) 0.91 D1: 0.98 (.015) 0.95	n/a	n/a	n/a

model becomes aggravated when coping with the results of flankers.

We examined how the foveal and peripheral VDP models predicted the effects of flankers in the stimuli by plotting the models' predictions in the same way as the true experimental results in Fig. 5. Table 3 (numbers in parentheses) show the slopes of all the model regressions for comparison with the interaction slopes of the actual experimental results. Two predictions of the VDP model are clearly inconsistent with observers' ratings: (i) the model seems to predict target-flanker interactions for foveal stimuli which is clearly counter to the experimental results and (ii) the predicted amount of interaction in the periphery is similar across the different flanker types – *Same*, *Diff1* and *Diff2* even though there was actually relatively little experimental interaction with *Diff1* flankers. Indeed, the model predicts an interaction slope of about 0.75 in the periphery, whatever the nature of the flankers and whatever the target-flanker distance. Section 4 will consider whether the behaviour of the VDP models can be changed to rectify these false predictions.

4. Discussion

In this paper, we have described suprathreshold discrimination experiments that examine how visual information about natural images is processed and integrated in central and peripheral vision. We are interested in the degree to which peripheral vision of natural scenes differs qualitatively from foveal vision, as has been shown in studies with other kinds of stimulus. We have attempted to show the extent of quantitative differences by modifying a foveal Visual Difference Predictor (VDP) model (To et al., 2010; Tolhurst et al., 2010), changing the components of the model to explain the change in sinewave grating sensitivity in the periphery. If the periphery acted like the fovea but at a low spatial scale (Anstis, 1998; Rovamo & Virsu, 1979; Thorpe et al., 2001; Zelinsky, 2008), then an appropriately adjusted foveal VDP should explain any changes in peripheral performance. However, that our porting of the VDP has not been successful shows that, in natural visual tasks, the contribution of peripheral vision must be subject to different rules than the fovea. We suggest that the well-studied phenomenon of crowding in peripheral vision is a key reason for the reduced perception of naturalistic stimulus changes in peripheral vision.

4.1. Perceived differences in peripheral natural-image targets

Overall, when rating the perceived differences between two patches of natural images presented at fixation and at 16° eccentricity, observers' ratings were lower for peripheral viewing. The VDP model could generate good predictions for observers' central ratings; in fact, the correlations between model predictions and observers' ratings were higher than those reported previously ($r = 0.77$ – 0.91 here vs. $r = 0.58$ – 0.72 in To et al. (2010)). The better performance here might be attributed to a reduction in the stimulus set size and the smaller number and variety of change dimensions – which we have previously shown to improve model/rating correlations (To et al., 2010).

However, VDP model performance was diminished for peripheral viewing of the same image pairs; the correlation between model predictions and observer' ratings dropped to only $r = 0.51$ – 0.78 . Observers' ratings were consistently overestimated by the model even though we ported the VDP model to the periphery by applying the lower peripheral contrast sensitivities (Mullen & Kingdom, 2002) and a shift towards lower spatial frequencies in the relation between bandwidth and optimal frequency (Tolhurst & Thompson, 1981; Yu et al., 2010). This suggests that decreased spatial acuity

and sensitivity alone cannot account for the diminished ability to detect differences in natural images in the periphery (Párraga et al., 2005), just as poor performance in tasks such as vernier acuity cannot be explained by changes in peripheral contrast discrimination (e.g. Hess & Field, 1993; Levi et al., 1994; Saarinen, 1988). Perhaps, other aspects of the model should be changed to allow a proper port of a foveal VDP to the periphery, but studies of the form of peripheral contrast “dippers” with gratings (Bradley & Ohzawa, 1986; Legge & Kersten, 1987) suggest that the numerical parameters underlying the non-linear transducer function (Eq. (1) in Supplementary material) will be much the same. During the whole evolution of our modelling (from Párraga et al. (2005)) with a variety of assumptions about receptive field shapes, transducer functions or suppressive interactions, we have never been able to generate a model that would correctly predict the relative thresholds or ratings for foveal and peripheral stimuli.

We investigated whether there are any differences between observers' abilities to perceive changes in colour and feature orientation in central and peripheral vision, and found that changes in orientation are comparatively more difficult to perceive peripherally. Now, van den Berg et al. (2007) reported different amounts of crowding for different features: crowding effects for orientation were stronger compared to those for colour (hue and saturation). Perhaps, this suggests an explanation for our results: features *within* an image patch in our experiments may crowd one another and so the lower difference ratings given to unflanked peripheral targets (especially for shape and orientation) could be due to internal self-crowding (Martelli et al., 2005). This would be difficult to test explicitly, since our natural stimuli do not often consist of discrete elements which we can remove or separate as in a formal crowding experiment. Furthermore, the types and numbers of any such elements would vary very much between stimuli, so that even dissected stimuli might behave differently from each other since the different types of element may have different crowding effects. Natural stimuli are very varied and it may not be possible to appreciate the rules governing their perception by attempting to break the scenes down into arbitrary elements.

In the periphery, we found that the ratings given to particular natural image differences were reduced by the simultaneous presentation of four flanking natural-image patches. Many of the properties of the target-flanker interaction suggest that we were measuring an aspect of typical crowding (e.g. Bouma, 1970; Levi, 2008; Pelli, 2008). First, although the presence of flankers made no difference in central vision, there was noticeable crowding in the periphery and observers found it more difficult to detect image differences. M-scaling the foveal stimuli (i.e. making them much smaller and bringing the flankers much closer) did not induce any crowding in our foveal experiments. Second, the most effective flankers were those similar to the targets. Third, crowding effects were strongest when the target-flanker distances were short, as great as a 50% reduction in perceived magnitude for the closest flankers. However, the effects were generally not large, and they did not fall off as much as the distance between target and flankers increased. Perhaps, the small magnitude relates to our presentation protocol where there were two separate time intervals and where there was no uncertainty in the target location; in a change detection experiment, Freeman and Pelli (2007) found least crowding when the target location was cued.

4.2. VDP model performance and the origin of crowding

The VDP model did a poor job of describing the perceived difference ratings for unflanked peripheral targets. The correlations between ratings and model predictions fell even further when in the flanked peripheral condition, in circumstances where crowding (e.g. Bouma, 1970; Levi, 2008; Pelli, 2008) might have been

expected. It is tempting to believe that there is a single explanation for why the model over-estimated the ratings for unflanked peripheral targets and for why it failed to capture the additional effects of flankers. The idea that features within the targets might crowd each other (Martelli et al., 2005) becomes more attractive. There were two systematic inconsistencies between model predictions and observers' ratings: (i) the model predicted interactions between target and flankers even in central vision, and (ii) the model predicted that the *Same*, *Diff1* and *Diff2* flanker conditions should be equally effective in the peripheral condition. In other words, the VDP models predicted a fairly general (eccentricity-independent) "masking" from surround stimuli in the fovea and periphery rather than what is generally described as "crowding" (Chung, Levi, & Legge, 2001; Pelli et al., 2004).

The VDP model predicted more interaction between flankers and target than we actually found experimentally, a particular embarrassment for the foveal model. The over-prediction of such lateral masking is probably prominent here because the radius of the surround suppression in our present model (rad_f see Supplementary material) was so large: 4.35 periods of the carrier sinewave in the Gabor receptive fields. This radius was not well constrained by the experimental data against which we ran the model optimisation; in To et al. (2010) the model with only slightly different assumptions gave a radius of only 0.59 periods. We re-ran the present model with surround suppression radius of 0.8 or 0.4 periods; this had little effect on the way that the model dealt with the stimuli being optimised (a subset of the stimuli described in To et al. (2010)), but it did remove the VDP prediction of interaction between target and flanker for foveal viewing, except at the closest flanking distance.

In order to address the model's inability to discriminate between the *Same*, *Diff1* and *Diff2* flanker conditions in the periphery, we further optimised the model parameters on Experiment 2, our most complete experiment with peripheral stimuli, where observers were presented with all three types of flankers. However, despite a small improvement in fit, dissimilar flankers were still predicted to be as destructive as similar ones. The difficulty in generating a model that can differentiate between flanker types has also been recently reported by van den Berg, Roerdink, and Cornelissen (2010). A simple model based on the behaviour of stylised V1 neurons, does not seem likely to be able to explain crowding phenomena (Flom, Heath, & Takahashi, 1963; Levi et al., 2002).

The present findings suggest that in natural scene perception, differences between central and peripheral vision cannot be attributed solely to lower spatial sensitivities or M-scaling. In fact, the effects similar to typical crowding found in the peripheral condition show that, when porting a VDP model (based on the presumed behaviour of V1 neurons) to the periphery, it will be necessary to include mechanisms known to generate crowding into the VDP. Many different mechanisms have been proposed and these mostly do not fit with the known behaviour of V1 neurons, such as spatial uncertainty (Krumhansl & Thomas, 1977; Nandy & Tjan, 2007; Pelli, 1985; Popple & Levi, 2005; Pöder & Wagemans, 2007; Strasburger, 2005; Wolford, 1975), imprecise integration (Neri & Levi, 2006; van den Berg et al., 2010), signal pooling or feature integration (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Pelli & Tillman, 2008; Pelli et al., 2004) or limited field of view (Tjan, 2009). The common feature of uncertainty or imprecision in such models might suggest that observers would find it harder to assign a number to a peripheral percept than to a foveal one, perhaps explaining why, in our experiments, the standard errors of the ratings were higher in peripheral than foveal vision. In the signal pooling model, features within the integration field are proposed to combine, thus forming a single texture. Since elements of identical colour tend to group together, it has been suggested (Chung et al., 2001; Pelli et al., 2004) that they may form a single

integrated "texture" more easily than some arbitrary collection of differently-coloured shapes; this might explain why colour changes in peripheral stimuli evoked higher ratings than orientation changes in our experiments. We have attempted to model spatial uncertainty or imprecise integration by adding various degrees of randomisation about where individual receptive fields originate in our VDP model so that feature information becomes spatially shuffled between the two images in the paired comparison. Indiscriminate shuffling of receptive-field locations over the large distances needed to explain crowding gave very unsatisfactory results. And, most significantly, any such procedure that could cause disruption to model the effects of our *Same* flankers was at least as disruptive with the *Diff* ones; van den Berg et al. (2010) have met the same problem. It appears that any shuffling would have to be applied arbitrarily and *only* between parts of image that are already deemed to be similar. This could therefore suggest that crowding is limited to within feature maps.

Alternatively, an adequate model of these results may need to include some uncertainty in the location of the "limited field of view" (Tjan, 2009), i.e. some uncertainty about which parts of the image are the parts to be compared. In our experiments, it is possible that, when observers are asked to judge the perceived difference between two targets, their response is partly based on information from the unchanging flankers. In attentional terms this would correspond to a misdirection of an attentional spotlight. Observers' decisions could be diluted by the zero-changes in the nearby non-changing flankers. However, this *again* only applies when the target and flankers are similar. The real challenge to developing a model of peripheral vision is, then, to devise a simple computational means of recognising that some features are similar to others.

Our attempts to model the known behaviour and interactions of V1 neurons have not been successful at illuminating the differences between foveal and peripheral vision, and our attempts to add spatial noise to the decision processes were unable to explain the effects of flanking stimulus patches. While it is possible that our model of peripheral V1 may need attention, this suggests that crowding occurs at levels higher than V1, whether immediately above V1 or even higher (e.g. Chung, Li, & Levi, 2007; Levi, 2008). Our results demonstrate effects of crowding on judgements of the appearance of natural scene stimuli in the periphery. Previous work has shown these effects with simpler stimuli, and our results are generally in keeping with that earlier work but, importantly, we show how these phenomena may carry over to the subjective assessment of complex natural scenes.

Acknowledgments

This work has been funded by project Grants (S56399, S56405) to D.J. Tolhurst and to I. Gilchrist and T. Troscianko from the EPSRC/Dstl of the UK. M.P.S.T. was employed on the grant to D.J.T. We are grateful for the helpful comments and suggestions of two anonymous reviewers.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2011.05.010.

References

- Anstis, S. M. (1998). Picturing peripheral acuity. *Perception*, 27, 817–825.
- Azzopardi, P., & Cowey, A. (1993). Preferential representation of the fovea in the primary visual cortex. *Nature*, 361(6414), 719–721.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12), 11–18 [13].

- Bijl, P., Koenderink, J. J., & Kappers, A. M. L. (1992). Deviations from strict M scaling. *Journal of the Optical Society of America A – Optics Image Science and Vision*, 9(8), 1233–1239.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241), 177–178.
- Bradley, A., & Ohzawa, I. (1986). A comparison of contrast detection and discrimination. *Vision Research*, 26(6), 991–997.
- Cannon, M. W. Jr., (1985). Perceived contrast in the fovea and periphery. *Journal of the Optical Society of America A – Optics Image Science and Vision*, 2(10), 1760–1768.
- Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research*, 41(14), 1833–1850.
- Chung, S. T., Li, R. W., & Levi, D. M. (2007). Crowding between first- and second-order letter stimuli in normal foveal and peripheral vision. *Journal of Vision*, 7(2), 11–13 [10].
- Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4), 659–671.
- Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science*, 142, 979–980.
- Foley, J. M., Varadharajan, S., Koh, C. C., & Farias, M. C. (2007). Detection of Gabor patterns of different sizes, shapes, phases and eccentricities. *Vision Research*, 47(1), 85–107.
- Freeman, J., & Pelli, D. G. (2007). An escape from crowding. *Journal of Vision*, 7(2): 22, 1–14.
- Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6, 858–873.
- Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology*, 252, 627–656.
- Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception and Psychophysics*, 45(3), 196–208.
- Hess, R. F., & Field, D. (1993). Is the increased spatial uncertainty in the normal periphery due to spatial undersampling or uncalibrated disarray? *Vision Research*, 33(18), 2663–2670.
- Hilz, R., & Cavonius, C. R. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research*, 14(12), 1333–1337.
- Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in human striate cortex. A revision of the classic Holmes map. *Archives of Ophthalmology*, 109(6), 816–824.
- Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *Journal of Comparative Neurology*, 158(3), 295–305.
- Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, 8(2), 255–279.
- Krumhansl, C. L., & Thomas, E. A. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception and Psychophysics*, 21, 269–279.
- Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, 9(10), 1–16 [6].
- Legge, G. E., & Kersten, D. (1987). Contrast discrimination in peripheral-vision. *Journal of the Optical Society of America A – Optics Image Science and Vision*, 4(8), 1594–1598.
- Levi, D. M. (2008). Crowding – An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48(5), 635–654.
- Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*, 2(2), 167–177.
- Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, 25(7), 963–977.
- Levi, D. M., Klein, S. A., & Wang, H. (1994). Discrimination of position and contrast in amblyopic and peripheral vision. *Vision Research*, 34(24), 3293–3313.
- Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, 7(2), 1–12 [4].
- Lovell, P. G., Párraga, C. A., Ripamonti, C., Troscianko, T., & Tolhurst, D. J. (2006). Evaluation of a multi-scale color model for visual difference prediction. *ACM Transactions on Applied Perception*, 3, 155–178.
- Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, 5(1), 58–70.
- Mullen, K. T., & Kingdom, F. A. (2002). Differential distributions of red–green and blue–yellow cone opponency across the visual field. *Visual Neuroscience*, 19(1), 109–118.
- Nandy, A. S., & Tjan, B. S. (2007). The nature of letter crowding as revealed by first- and second-order classification images. *Journal of Vision*, 7(2), 1–26 [5].
- Neri, P., & Levi, D. M. (2006). Spatial resolution for feature binding is impaired in peripheral and amblyopic vision. *Journal of Neurophysiology*, 96(1), 142–153.
- Nugent, A. K., Keswani, R. N., Woods, R. L., & Peli, E. (2003). Contour integration in peripheral vision reduces gradually with eccentricity. *Vision Research*, 43, 2427–2437.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744.
- Párraga, C. A., Troscianko, T., & Tolhurst, D. J. (2005). The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model. *Vision Research*, 45(25–26), 3145–3168.
- Peli, E., Yang, J. A., Goldstein, R., & Reeves, A. (1991). Effect of luminance on suprathreshold contrast perception. *Journal of the Optical Society of America A – Optics Image Science and Vision*, 8(8), 1352–1359.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A – Optics Image Science and Vision*, 2(9), 1508–1532.
- Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, 18(4), 445–451.
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4(12), 1136–1169.
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11(10), 1129–1135.
- Pöder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, 7(2), 1–12 [23].
- Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, 29(9), 1133–1151.
- Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, 113(3), 426–442.
- Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1), 199–210.
- Popovic, Z., & Sjostrand, J. (2001). Resolution, separation of retinal ganglion cells, and cortical magnification in humans. *Vision Research*, 41(10–11), 1313–1319.
- Popple, A. V., & Levi, D. M. (2005). The perception of spatial order at a glance. *Vision Research*, 45(9), 1085–1090.
- Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, 21(3), 409–418.
- Rohaly, A. M., Ahumada, A. J., Jr., & Watson, A. B. (1997). Object detection in natural backgrounds predicted by discrimination performance and models. *Vision Research*, 37(23), 3225–3235.
- Rolls, E. T., & Cowey, A. (1970). Topography of the retina and striate cortex and its relationship to visual acuity in rhesus monkeys and squirrel monkeys. *Experimental Brain Research*, 10, 298–310.
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), 11–22 [17].
- Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, 37(3), 495–510.
- Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision*, 9(2), 1–11 [5].
- Saarienen, J. (1988). Perception of spatial order in extrafoveal vision. *Scandinavian Journal of Psychology*, 29(3–4), 162–167.
- Smyth, D., Willmore, B., Thompson, I. D., Baker, G. E., & Tolhurst, D. J. (2003). The receptive-field organisation of simple cells in primary visual cortex (V1) of ferrets under natural scene stimulation. *Journal of Neuroscience*, 23, 4746–4759.
- Strasburger, H. (2005). Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, 5(11), 1024–1037.
- Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception and Psychophysics*, 49(6), 495–508.
- Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bulthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, 14(5), 869–876.
- Tjan, B. S. (2009). Three essential ingredients of crowding. *Journal of Vision*, 9(8), 988.
- To, M. P. S., Gilchrist, I. D., Troscianko, T., Kho, J., & Tolhurst, D. J. (2009). Perception of differences in peripherally-viewed natural image stimuli: Why is peripheral viewing so much worse than foveal? *ACM Transactions on Applied Perception*, 6(26), 1–9.
- To, M., Lovell, P. G., Troscianko, T., & Tolhurst, D. J. (2008). Summation of perceptual cues in natural visual scenes. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 275(1649), 2299–2308.
- To, M. P. S., Lovell, P. G., Troscianko, T., & Tolhurst, D. J. (2010). Perception of suprathreshold naturalistic changes in colored natural images. *Journal of Vision*, 10(4), 11–22.
- Tolhurst, D. J. (1989). Cortical magnification and peripheral innervation density. In J. J. Kulikowski, C. M. Dickinson, & I. J. Murray (Eds.), *Seeing contour and colour* (pp. 671–677). Oxford: Pergamon Press.
- Tolhurst, D. J., & Ling, L. (1988). Magnification factors and the organization of the human striate cortex. *Human Neurobiology*, 6(4), 247–254.
- Tolhurst, D. J., & Thompson, I. D. (1981). On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 213(1191), 183–199.
- Tolhurst, D. J., To, M. P. S., Chirumuta, M., Troscianko, T., Chua, P. Y., & Lovell, P. G. (2010). Magnitude of perceived change in natural images may be linearly proportional to differences in neuronal firing rates. *Seeing and Perceiving*, 23, 349–372.
- Torrilba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4), 766–786.

- van den Berg, R., Roerdink, J. B., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision*, 7(2), 1–11 [14].
- van den Berg, R., Roerdink, J. B., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Computational Biology*, 6(1), e1000646.
- van Essen, D. C., Newsome, W. T., & Maunsell, J. H. (1984). The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. *Vision Research*, 24(5), 429–448.
- Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A – Optics Image Science and Vision*, 14(9), 2379–2391.
- Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, 82(3), 184–199.
- Wright, M. J., & Johnston, A. (1983). Spatiotemporal contrast sensitivity and visual field locus. *Vision Research*, 23(10), 983–989.
- Yu, H. H., Verma, R., Yang, Y., Tibballs, H. A., Lui, L. L., Reser, D. H., et al. (2010). Spatial and temporal frequency tuning in striate cortex: Functional uniformity and specializations related to receptive field eccentricity. *European Journal of Neuroscience*, 31(6), 1043–1062.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115, 787–835.