Breaking camouflage: Binocular disparity reduces contrast masking in natural images

Susan G. Wardle   School of Psychology, The University of Sydney, Australia

John Cass   School of Psychology, University of Western Sydney, Australia

Kevin R. Brooks   Department of Psychology, Macquarie University, Australia

David Alais   School of Psychology, The University of Sydney, Australia

Visual overlay masking is typically studied with a mask and target located at the same depth plane. Masking is reduced when binocular disparity separates the target from the mask (G. Moraglia & B. Schneider, 1990). We replicate this finding for a broadband target masked by natural images and find the greatest masking (threshold elevation) when target and mask occupy the same depth plane. Masking was reduced equally whether the target appeared at a crossed or an uncrossed disparity. We measure the tuning of masking and determine the extent of the benefit afforded by disparity. Threshold elevation decreases monotonically with increasing disparity until ±8 arcmin. Two underlying components to the masking are evident; one accounts for around two-thirds of the masking and is independent of disparity. The second component is disparity-dependent and results in additional masking when there is zero disparity. Importantly, the reduction in masking with disparity cannot be explained by interocular decorrelation; we use a single-interval orientation discrimination task to exclude this possibility. We conclude that when the target and mask are presented at different depths they activate distinct populations of disparity-tuned neurons, resulting in less masking of the target.

Keywords: depth, contrast sensitivity, masking, binocular vision

Introduction

Stereopsis—the process of perceiving depth from the differences between two monocular half-images—is one of the many depth cues that contribute to the visual system's ability to segment objects in a cluttered visual scene. Stereopsis produces a compelling sensation of depth even in the absence of other depth and segmentation cues, e.g., random-dot stereograms (Julesz, 1971). As stereopsis provides quantitative depth information, it is reasonable to expect that it may assist in situations where image segmentation has been deliberately impoverished, for example, when noise is added to a signal, as in masking experiments. In a classic paradigm known as simultaneous contrast masking, a mask is superimposed on a target stimulus and subsequently measured contrast thresholds for the detection or discrimination of the target are elevated relative to an unmasked baseline condition (e.g., Legge & Foley, 1980).

There has been limited investigation of whether interocular differences and relative disparity are effective in reducing contrast masking. Binocular detection of a horizontal or vertical grating against narrowband or broadband noise is improved when the gratings are out of phase in each eye (Henning & Hertz, 1973, 1977). When the mask and the target are the same mean frequency and orientation, the improvement in detection is limited to spatial frequencies less than 6 c/deg (Henning & Hertz, 1973). Similarly, detection thresholds are lowered for a Gabor target presented at a different disparity than a Gaussian noise mask, compared to when the Gabor and mask are at the same disparity (Moraglia & Schneider, 1990). In this case, there was no reported difference between crossed and uncrossed disparities in the amount of stereo release from masking. The size of the effect reported by Moraglia and Schneider (1990) was around 7 dB for small binocular disparities (13.52 arcmin) in either direction.

Henning and Hertz did not control eye vergence, thus it is unclear whether the signal or the noise was out of phase in their experiments. Moraglia and Schneider (1990) controlled vergence by asking subjects to fixate on a binocular fixation square prior to each trial. The stimuli

were displayed for only 90 ms, too brief for vergence eye movements to occur after fixation. However, this method of controlling vergence is not as precise as using nonius lines and would only ensure that the fixation mark was constrained within Panum's area. Moraglia and Schneider (1990) originally found no change in the degree of release from masking with increasing disparity during free viewing. However, when vergence was controlled, the benefit of disparity was only evident at the smallest disparity of 13.52 arcmin. At larger disparities (40.56 arcmin and 67.60 arcmin), performance was comparable to when the target and mask were at the same depth. This suggests that when vergence was not controlled and observers had 1 s to view the stimulus, they were able to adjust eye vergence at the larger disparities and obtain a similar advantage to that evident at 13.52 arcmin. Overall, this suggests that a disparity limit exists for the release from masking, but the tuning of the effect has not yet been measured. The first purpose of the current study is to address this by measuring the extent of masking as a function of crossed and uncrossed disparities between target and mask.

A concern with previous studies on the effect of relative disparity on contrast masking is that the results could be explained by sensitivity to interocular decorrelation. Previous demonstrations of stereo release from masking (Henning & Hertz, 1973, 1977; Moraglia & Schneider, 1990, 1992; Schneider, Moraglia, & Jepson, 1989) have used two-interval forced-choice detection tasks. It is possible that the advantage conferred by the disparity signal may result from the presence of an additional cue in the interval containing the target (disparity) rather than reflecting a relative improvement in luminance contrast sensitivity. If this is the case, the lower detection thresholds measured in the disparity conditions reflect the visual system's known sensitivity to interocular decorrelation (Cormack, Stevenson, & Schor, 1991; Tyler & Julesz, 1978), rather than the ability to perceive the target at lower contrast when it is at relative depth to the mask. The second aim of this study is to rule out this explanation by using a single-interval discrimination task to measure the release from masking with disparity.

The final purpose of this paper is to extend the release from masking to a new set of stimuli—natural images. Few studies have examined masking in natural images. Natural images typically approximate a 1/frequency amplitude spectrum in the spatial domain, i.e., they contain more energy at low than high spatial frequencies (Field, 1987; Tolhurst, Tadmor, & Chao, 1992). Bex, Solomon, and Dakin (2009) measured the masking of spatially narrowband targets embedded in natural image movies. The greatest threshold elevation occurred for detecting low spatial frequency targets. The amount of masking depended partly on the amplitude spectrum of the mask; masking was reduced when the natural images were "whitened" to contain an equal distribution of all spatial frequencies. The visibility of a target was relatively independent of the local RMS contrast in the natural image mask; however, it was affected by local edge density. In another study, maximal suppression of the perceived contrast of a texture patch occurred when it was surrounded by a (non-overlapping) natural image with typical amplitude spectrum characteristics (McDonald & Tadmor, 2006). Together, these results imply that the mechanisms underlying visual masking may be calibrated to natural image statistics.

In this paper, we examine whether stereopsis is effective in breaking contrast masking for a broadband target with typical natural image spectra embedded in complex natural image masks during free viewing (Experiment 1A) and controlled fixation (Experiment 1B). To preview the results, we find that masking is greatest when the mask and target are at the same depth plane, replicating Moraglia and Schneider's (1990) result with a new stimulus set consisting of natural images. We then measure the tuning of masking across a range of crossed and uncrossed disparities, which has not yet been reported (Experiment 2). We find narrow tuning of masking as a function of disparity. The benefit of relative disparity for detecting the target occurs at relatively small disparities and asymptotes at larger disparities. Finally, we design a single-interval orientation identification task (Experiment 2) in order to rule out interocular decorrelation as an alternative explanation for the reduction in masking observed with disparity.

Experiments 1A and 1B

The aim of these experiments is to measure the release from masking produced by disparity in natural image stimuli. In Experiment 1A, the natural image mask was fixed at zero disparity and the target was a phase-randomized section of the mask that was either at zero, crossed, or uncrossed disparity. In Experiment 1B, it was the target that was fixed at zero disparity and the mask that was presented at zero, crossed, or uncrossed disparity. Experiment 1A was conducted under free viewing with no fixation; vergence eye movements were controlled in Experiment 1B by nonius lines and a short presentation time.

Methods
Observers

Authors SW and JC participated in both experiments. In each, there were two additional observers who were naive to the experimental aims. In Experiment 1A, one was an experienced psychophysical observer (DAP) and the other had never participated in any psychophysical experiment (SD); two experienced psychophysical observers naive to the hypotheses (DM and AC) participated in Experiment 1B.

Stimuli

Stimuli were generated on an Apple MacPro 4.1 computer using MATLAB (The MathWorks) with functions from the Psychtoolbox (Brainard, 1997; Pelli, 1997). Stimuli were presented on a CRT monitor (Mitsubishi Diamond Pro 2070SB) at 90 or 100 Hz with screen resolution of 1280 × 960 pixels. A bit-stealing algorithm (e.g., Tyler, Chan, Liu, McBride, & Kontsevich, 1992) was applied to obtain 1786 grays for higher contrast resolution. The monitor was gamma-corrected to obtain linear output from each RGB gun.

Natural image stimuli were selected from the Zurich Natural Image database (Einhäuser, Kruse, Hoffmann, & König, 2006) and converted to grayscale images. Images were selected using a custom-made program and according to the criteria that they contained no man-made structures and had a dense spatial structure. The stimuli used for the natural image masks were segments (280 × 280 pixels, 7.27°) of the original large images. Two different sets of 5 images were used for Experiment 1A and Experiment 1B (Figure 1). All images were scaled to have a normalized mean luminance of 0.5 (display range 0–1) and the RMS contrast was set to 0.155. If this scaling produced any pixels outside the monitor display range, the image was not included in the experiment and another was chosen until the criterion was satisfied. To eliminate hard edges, the mask was presented in a circular raised cosine window with a uniform center of 6.75° diameter and an outer transition band of 0.47°. Beyond this, the screen was set to a mean luminance gray. The target was a phase-scrambled version of the central region of the grayscale mask. Phase scrambling was done in Fourier space. The target (100 × 100 pixels) was presented in a circular raised cosine window (center 2.08° diameter, outer band 0.52°) that ramped off to mean luminance. The target measured 2.6° in diameter after windowing.

To assist with stereo fusion, a black square outline (360 × 360 pixels, 9.35° in height and width) surrounded the stimuli in each eye's view and served as a fusion lock. When fused, it appeared as a frame 2 pixels wide in the plane of the screen. The natural image mask was centered in the black square. In target intervals, the target was added to the mask and the total image divided by two, in order to keep luminance values within the monitor display range. For intervals with the mask only, the natural image was added to a field of mean luminance and divided by two so that the only difference between the intervals was the presence of the target.

In Experiment 1A, the mask was fixed at zero disparity (relative to the fusion lock and monitor edges) and the target appeared at −6 arcmin (crossed), +6 arcmin (uncrossed), or zero disparity (i.e., at the same disparity as the mask and frame). In Experiment 1B, the same relative disparities were used; however, the target was fixed at zero disparity and the mask appeared at crossed, uncrossed, or zero disparity. Horizontal disparity was produced by shifting the target in each eye's image to the left or right by 4 pixels before adding it to the mask. Baseline detection thresholds for the target without the mask were obtained for each of the three target disparities in Experiment 1A and for zero disparity in Experiment 1B.

Procedure

Participants rested their head on a chin and forehead rest fixed 57 cm from the screen and viewed the stimuli through a mirror stereoscope. The room was dark with the walls painted black and the only illumination came from the monitor screen. Prior to starting the experiment, observers completed a short training program in which they were shown stimuli similar to the mask and target at suprathreshold contrast and required to classify their
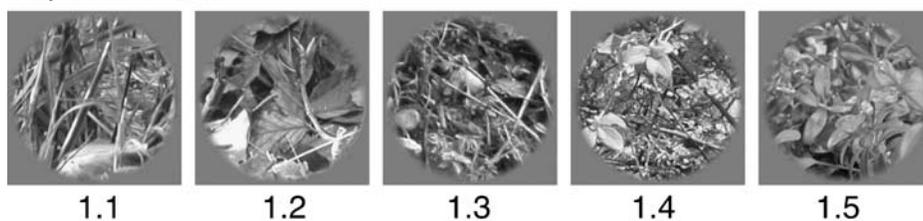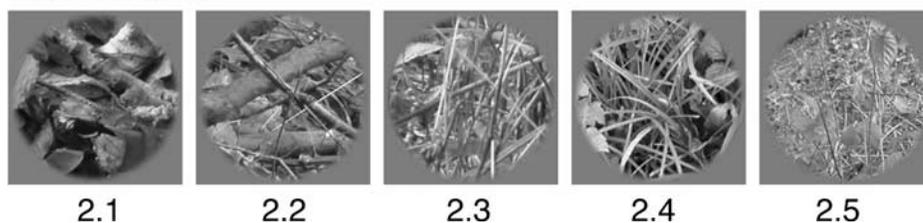

Figure 1. Natural image masks used in Experiments 1A and 1B; the image numbers are arbitrary labels. The masks shown here have not been standardized for luminance and contrast.

relative depth order in a series of trials. This served as training for the experimental conditions and ensured that observers had adequate stereo vision for the task. Performance was near ceiling for all observers tested.

The experiment was a temporal two-alternative forced-choice procedure. Participants indicated whether the target was in the first or second interval using the keyboard. Feedback was given by changing the fusion lock for 500 ms to white (correct response) or dark gray (incorrect response). For baseline conditions, the mask was not presented, and a small black dot appeared above the fixation frame to signal each interval, and was not present in between intervals. In Experiment 1A, fixation was not controlled; each interval was 1 s and the inter-stimulus interval (ISI) was 500 ms. In Experiment 1B, binocular fixation (and hence vergence) was controlled by nonius lines and a shorter presentation time of 200 ms. This duration was a compromise between being long enough to perform the task effectively but short enough to prevent completion of a vergence eye movement (McKee, Bravo, Taylor, & Legge, 1994). Before each interval in a trial, observers viewed two sets of nonius lines (top pair presented to right eye and bottom pair presented to left eye) and pressed a key to initiate the interval when the lines appeared to be aligned.

Contrast detection thresholds were estimated using the "QUEST" adaptive staircase algorithm (Watson & Pelli, 1983). Each experimental run included two interleaved staircases of 25 trials (50 trials per threshold estimate). For each participant and each condition, responses from all runs were pooled and fitted with a cumulative Gaussian psychometric function to establish the target contrast at detection threshold. Threshold was defined by a criterion of 75% correct. There were two practice trials at the start of each staircase that were not included in the threshold estimate.

Results and discussion

Contrast detection thresholds were calculated separately for each observer and condition by pooling the data from all runs of the condition and estimating the value of the cumulative Gaussian psychometric function fitted to the data at 75% correct. Threshold elevation relative to baseline is calculated in decibels (dB), as shown in the following equation:

$$\text{threshold elevation (dB)} = 20 \log_{10} \frac{M}{B}, \qquad (1)$$

where M is the detection threshold for the masked target and B is the baseline detection threshold for the target against a background of mean luminance.

In both experiments, and across all disparities and images, detection thresholds were elevated when the mask was present relative to baseline detection of the target on a background of mean luminance (Figure 2). Threshold

elevation (averaged across images and observers) was greater when the target and mask were at the same disparity than when the target was in front (Experiment 1A: $t_{(4)} = 12.118$, $p < 0.001$, Experiment 1B: $t_{(4)} = 7.593$, $p < 0.01$) or behind (Experiment 1A: $t_{(4)} = 12.532$, $p < 0.001$, Experiment 1B: $t_{(4)} = 4.682$, $p < 0.01$) the mask.[1] Averaged over observers, this pattern of results occurred for each of the 10 natural image masks in the two experiments (see Figures 2b and 2d)[2]. The data for individual observers (not shown) also followed this pattern. The sign of the disparity did not modulate the effect, as there was no significant difference in threshold elevation for crossed versus uncrossed disparities of the target (Experiment 1A, $t_{(4)} = 1.739$, $p > 0.05$) or the mask (Experiment 1B, $t_{(4)} = 2.078$, $p > 0.05$). Thus, relative depth, defined by binocular disparity, between the natural image mask and natural-image-like target did provide a partial reduction in masking from natural images, regardless of whether the target appeared in front or behind the mask.

The magnitude of the release from masking under stereoscopic conditions (averaged across observers and images) was 9.0 dB (target in front) and 7.7 dB (target behind) in Experiment 1A. For Experiment 1B, it was 5.9 dB (target in front) and 4.4 dB (target behind). These values are the difference in threshold elevation (averaged across observers and images) between crossed or uncrossed disparity conditions and zero disparity. The greater variability between thresholds for different images and the smaller effect in Experiment 1B may be because fixation (and thus absolute disparity) was controlled and the stimulus presentation time was much shorter in that experiment, making it a more difficult task. Baseline thresholds for detecting the target alone were higher in Experiment 1B (see Figures 2a and 2c), suggesting that the shorter stimulus duration may have decreased performance. It has long been known that performance on various stereoscopic tasks improves as durations increase through the range used here (e.g., McKee, Levi, & Bowne, 1990; Ogle & Weil, 1958; Watt, 1987), predicting more effective depth segmentation for longer durations. However, the replication of the effect under these stricter conditions demonstrates that the reduced masking in stereo conditions could not be explained by vergence eye movements alone in Experiment 1A.

Experiment 2

In this experiment, we derive a tuning curve for the release from masking with stereopsis, extending the range of disparities between the target (fixed at zero disparity) and the mask to ±24 arcmin. The second issue we address is whether stereo unmasking in detection experiments reflects a genuine release from masking. An alternative explanation for the better performance when there is
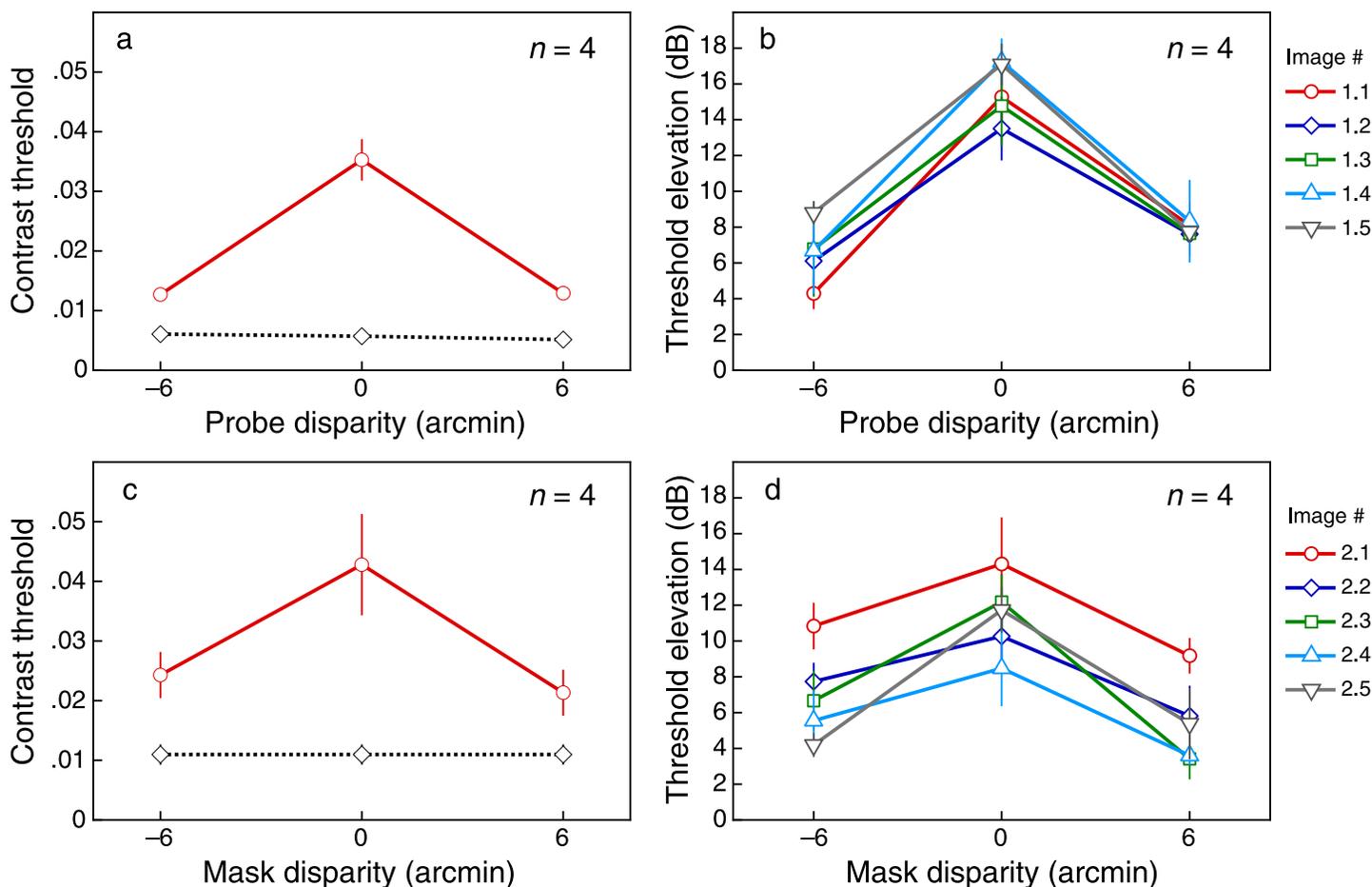
Figure 2. Data averaged over all observers in (top) Experiment 1A and (bottom) Experiment 1B. (a, c) Mean contrast detection thresholds averaged over 5 natural image masks. Baseline detection thresholds are in black; thresholds for the superimposed mask and target are in red. (b, d) Mean threshold elevation in dB plotted separately for each image mask. Error bars represent ±1 *SE*.

relative disparity between the target and mask is that a cue other than stereo depth can be used to discriminate between the two intervals in the trial. There are two candidates for alternative cues that are discussed below.

First, interocular decorrelation could be used to discriminate between the intervals in the detection task. When the mask and target are both at zero disparity, the cross-correlation between the images presented to the left and right eyes is perfect and uniform across the image. However, when the target is shifted in each eye's image relative to the mask, it produces a dip in the cross-correlation function that corresponds to the spatial location of the target. The human visual system is exceptionally sensitive to decorrelation (Cormack et al., 1991; Tyler & Julesz, 1978). Thus, in the detection task, subjects would only need to detect a decorrelation in one of the intervals in order to produce lower thresholds in the conditions where the mask and target appeared separated in depth.

Second, some observers in the detection task reported that when the target was not visible, it sometimes

appeared that there was some depth in the mask in one interval. This means that detection of stereo depth in the stimulus did not necessarily correspond to detection of the target. If observers were basing their detection on one of these factors, their improved performance at non-zero disparities would not reflect a genuine ability to perceive the target at lower contrast when it is presented in depth, and detection thresholds would be artificially lowered in the stereo conditions. As all previous work on stereo unmasking has used a detection task (Henning & Hertz, 1973, 1977; Moraglia & Schneider, 1990, 1992; Schneider et al., 1989), we designed a discrimination task in which task performance is decoupled from the stereo information. The target was a 1 c/deg Gabor, tilted either ±10° from vertical, and observers were required to identify whether it was tilted clockwise or anticlockwise from vertical in a single interval. Thus, the detection of interocular decorrelation or stereo depth would in no way assist with the threshold task.

Pilot experiments revealed that orientations present in the natural image masks interfered with orientation

discrimination of the target. Consequently, we used a random noise mask that was filtered to contain the $1/f$ spectrum typical of natural images (Field, 1987; Tolhurst et al., 1992) but that contained equal spatial frequency energy at all orientations. The mask therefore contained no coherent orientation information that could interfere with the task of judging target orientation.

Methods

Observers

Three authors (SW, JC, DA) and one observer naive to the experimental aims (AC) participated.

Stimuli

The stimulus size, cosine windows, and general setup were identical to those in Experiment 1 except for the following details. The target was a 1 c/deg Gabor that was oriented ±10° from vertical in each monocular half-image. The mask was $1/f$ noise, filtered from white noise in Fourier space. A unique $1/f$ mask was generated for every trial in the experiment. The mask was presented at one of nine disparities in each run [−24, −18, −12, −6, 0, 6, 12, 18, 24 arcmin]. The target was always at zero arcmin (fixation plane).

Procedure

Thresholds were measured to determine the minimum contrast required to correctly discriminate the orientation of the target at 75% correct. The task was a single-interval binary choice; participants were required to indicate the orientation of the target (tilted clockwise or anticlockwise) on every trial. Unlike Experiment 1, the target was present on every trial. Fixation was controlled using nonius lines before each trial and a 200-ms stimulus duration. Two interleaved QUEST (Watson & Pelli, 1983) staircases with 30 trials each were used in each run of a condition; two additional practice trials from the start of each QUEST were not included in the threshold estimate. Each observer completed at least two runs of each disparity condition, resulting in at least 120 trials per threshold estimate (Figure 3).

Results and discussion

Contrast thresholds for orientation discrimination were calculated separately for each mask disparity and the baseline condition by pooling the data from all runs of the condition and estimating the 75% correct-discrimination value of the cumulative Gaussian psychometric function
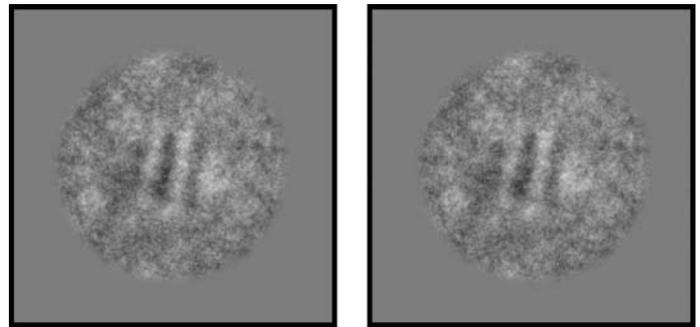

Figure 3. Example of the stimuli used in Experiment 2. The $1/f$ noise mask and 1 c/deg Gabor target (oriented ±10° from vertical) are summed together. The images in this figure can be free-fused; convergent fusion will display the target in front of the mask.

fitted to the pooled data. These thresholds are plotted separately for each observer in Figure 4b.

A Gaussian function was fitted (using the Levenberg–Marquardt algorithm) to each observer's masking data as a function of mask disparity. This function is described by Equation 2, where α is the horizontal asymptote, a is the height of the peak (in decibels above α), b is the position of the center of the peak, and c is the width of the peak (Equation 2). Three parameters were used for the fit; a, c, and α; the center of the Gaussian was fixed at $b = 0$ as maximal threshold elevation was expected when the mask was at zero disparity:

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}} + \alpha. \qquad (2)$$

The contrast thresholds as a function of mask disparity were well described by a Gaussian fit for each of the observers ($\chi^2 = 6.2$–24.1, $Q = 0.520$–0.001; see Figure 4b). The data were averaged across observers for each disparity and plotted as threshold elevation in dB (Equation 1). A Gaussian function was also fitted to the average data ($\chi^2 = 1.75$, $Q = 0.972$; see Figure 4a). Confidence intervals were obtained for the parameter estimates from this overall fit using Monte Carlo simulations.

A substantial amount of contrast masking occurred even at the greatest depth separation tested. This suggests that at least two underlying components contribute to the masking. The first is independent of the disparity of the mask and is described by the elevation of the Gaussian fit, parameter α. This was 13.5 dB for the overall model (95% CI: [12.9, 14.1] dB). The second is the magnitude of masking that was dependent on the disparity of the mask, quantified by a, the estimated peak of the Gaussian. This was 5.0 dB in the overall model (95% CI: [3.8, 6.1] dB). Finally, the parameter c is the standard deviation, which for the overall model was 3.9 arcmin (95% CI: [1.3, 6.7] arcmin). The bandwidth estimate based on this standard deviation is 4.6 arcmin at half-width and half-amplitude.
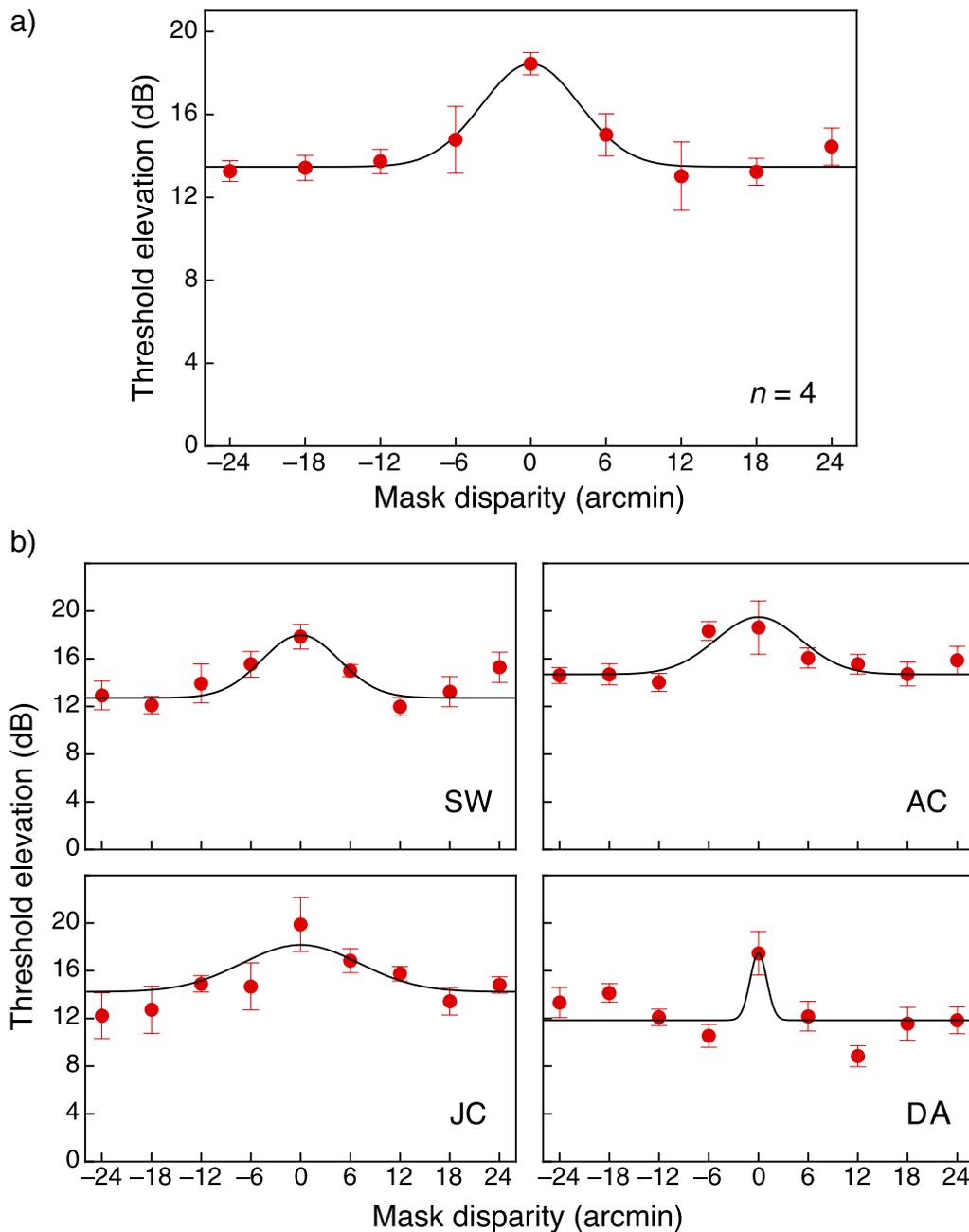
Figure 4. (a) The data points show mean threshold elevation as a function of mask disparity averaged over all observers in Experiment 2. Error bars represent ±1 *SE*. The continuous black line is the best fitting Gaussian function, as described in Equation 2. (b) Threshold elevation for individual observers. Error bars represent standard deviations derived from bootstrapping, and the continuous line is the best fit of Equation 2.

We can then calculate the disparity at which a curve at this bandwidth ceases to be significantly above baseline by computing the values of the fitted function at the upper 95% confidence limit for α. This indicates that for our data and the disparity range tested, threshold elevation declines for disparities up to ±8 arcmin, after which it is not significantly different from the baseline level of masking.

That good fits were established with symmetrical functions also confirms (for discrimination) the finding in Experiments 1A and 1B that there is no difference in crossed compared to uncrossed disparity in the extent of the release from masking. The target is not easier to discriminate when it is in front of the mask compared to when it is behind.

General discussion

We used simultaneous overlay[3] masking to examine the effect of binocular disparity on contrast detection thresholds for a broadband target with typical (1/*f*) natural image spectra embedded in complex natural image masks. For all ten natural image masks, threshold elevation was observed at all disparities (including zero). However, improvements in detection threshold were observed when there was relative depth between the target and mask. This was the case under conditions of free viewing (Experiment 1A) and when vergence was controlled with fixation at zero disparity (Experiment 1B). In agreement with Moraglia and Schneider (1990), there was no difference in the extent of the release from masking between equivalent crossed and uncrossed disparities. We used ±6 arcmin, about half the smallest disparity (13.52 arcmin) used by Moraglia and Schneider (1990). They reported a release from masking of around 7 dB in magnitude under optimal conditions, which is comparable with our range of 4.4–9.0 dB.

Observers were able to discriminate the orientation of a Gabor target embedded in 1/*f* noise at lower contrasts when there was relative depth between the target and mask (Experiment 2). This is important because in this case, task performance was independent of whether stereo information was present, unlike all previous investigations. This confirms that the target can genuinely be perceived at a lower contrast when there is relative depth separating it from the mask, and that detection thresholds were not artificially lowered by the presence of an additional cue (e.g., disparity in the mask) in the target interval of a two-interval, 2AFC trial. In addition, it rules out the possibility that observers were detecting interocular decorrelation in the target interval in the discrimination tasks and implies that it was disparity that lowered thresholds.

We quantified the tuning of the masking in Experiment 2, using a broader range of disparities (±24 arcmin). A Gaussian was fitted to the tuning curves and parameters were estimated to quantify the effect of stereo release from masking. The extent of the advantage for discriminating a target when it is in depth (for a Gabor target in 1/*f* noise) is 5 dB, as quantified in Experiment 2 by the parameter *a* in the Gaussian model (Equation 2). This is less than the disparity-independent masking, which was estimated at 13.5 dB compared to baseline. The maximum level of masking occurs when there is no relative disparity between mask and target and declines monotonically for disparities up to ±8 arcmin. For our observers, increasing the disparity beyond 8 arcmin does not result in any further reduction in masking (within the range tested), and the effect is comparable for crossed and uncrossed disparities.

To tie together the results of Experiments 1 and 2, we can compare threshold elevation at the disparity values used in both experiments, in the range −6 to +6 arcmin. Threshold elevation at each disparity is similar in magnitude for Experiments 1A and 1B (see Figures 2a and 2c), which used similar stimuli and a 2AFC detection task. Threshold elevation at the same disparity values in Experiment 2 is greater, and at some disparities is nearly double the elevation in Experiment 1 (see Figure 4a). The greater threshold elevation overall in Experiment 2 may be due to the more demanding single-interval discrimination task. However, if the difference between threshold elevation at zero disparity and at ±6 arcmin is considered, the magnitude of the release from masking in Experiment 1B and Experiment 2 is less than in Experiment 1A. This may be because Experiments 1B and 2 used the same short stimulus duration (200 ms) and vergence control with nonius lines, whereas Experiment 1A used a longer duration of 1 s and no vergence control. Moraglia and Schneider (1990) also reported a greater reduction in masking with disparity when the stimulus duration was 1 s and vergence was not controlled.

Functional implications

The reduction in masking observed with binocular disparity between target and mask indicates that disparity is a useful cue for image segmentation. In transparent motion, it has been observed that introducing binocular disparity between two transparent planes of random dots translating in opposite directions reduces inhibition in area MT (Bradley, Qian, & Andersen, 1995) and increases the number of surfaces that can be simultaneously detected (Greenwood & Edwards, 2006). We suggest that the stereo release from contrast masking we observe may be related. Threshold elevation resulting from masking may be conceptualized as a reduction in the signal-to-noise ratio associated with a target-relevant response. Introducing relative disparity between the target and masking stimulus may, therefore, serve to improve signal-to-noise response by effectively parsing signal and noise into different "disparity channels." The question of whether this parsing enhances the signal (by disinhibition, for example; Bradley et al., 1995) or effectively reduces target-relevant noise is not determinable from the current experiments.

A related finding is that increasing the degree of disparity between opposing motion signals leads to greater segregation of overlapping motion planes (Qian, Andersen, & Adelson, 1994). Qian et al. generated random-dot displays that consisted of two planes of motion in opposite directions. Each dot was paired with a dot moving in the opposite direction to produce locally balanced motion. These paired dot displays typically appear as motion noise; however, if stereo depth is added to differentiate the two planes, they appear as two transparent planes of motion. Observers reported that the clarity of transparent

motion in these patterns increases with increasing disparity up to at least 27 arcmin. However, we found no equivalent trend in stereo unmasking for a similar range of disparities in Experiment 2. The improvement in target detection observed with relative disparity between target and mask leveled out at small disparities (after 8 arcmin). The reason for this difference is not clear. In metacontrast masking, where the target and mask do not overlap spatially, the sign and magnitude of the disparity does modulate the effect. Lehmkuhle and Fox (1980) reported that gap position in a Landolt C target defined by stereoscopic contours was more easily identified when in front of a (non-overlapping) annulus mask than when on the same depth plane. The degree of masking of the C by the annulus declined monotonically as a function of depth separation, and there was an asymmetry as masking increased when the C was behind the mask.

It has been suggested that stereopsis might have evolved to assist in breaking camouflage, perhaps most notably by Julesz (1971). McKee, Watamaniuk, Harris, Smallman, and Taylor (1997) investigated the detection of a moving dot target amidst "Brownian motion" noise dots, in a visual search paradigm. They found that when the (randomly changing direction) noise dots are in the same depth plane as the (linear trajectory) signal dot, performance was poor. As the plane of noise dots was separated further from the plane of the signal dot, performance improved up to 6 arcmin, after which it declined again toward "same plane" levels. As in the present experiments, detection was equally good regardless of whether the target was in front or behind the noise plane. McKee et al. found that the addition of disparity between a static target and static plane of noise dots also improved detection performance and to a greater extent than for moving stimuli. Stereopsis has also been demonstrated to break camouflage when the target in one eye is not visible (Brooks & Gillam, 2006a, 2006b, 2007; Cook & Gillam, 2004; Gillam, Blackburn, & Nakayama, 1999). Here, the presence of an unmatched "target" feature in one eye allows us to "infer" a matching target in the other eye even though it is entirely camouflaged against its background (i.e., absolutely no visible features). Under such circumstances, an appropriate perceived depth emerges in the (monocular, "half-camouflaged") target.

Physiological mechanisms

The observation that masking is reduced when depth from binocular disparity separates the mask and target leads to speculation that at least some kinds of masking may occur at a late stage in visual processing, after matching for stereopsis has occurred. There is psychophysical evidence that this is likely to be the case for luminance-modulated (McKee et al., 1994) and contrast-modulated (Harris & Willis, 2001) dichoptic masking. McKee et al. (1994) found that a high-contrast bar in one

eye masked a test bar presented at corresponding retinal locations in the other eye. However, thresholds were dramatically reduced when a second high-contrast bar was added next to the test bar, providing a stereo match for the dichoptic mask and specifying a disparity-defined depth away from the plane of the screen. Thus, the stereo matching appeared to cause a partial release from dichoptic masking, suggesting that the site of masking is later in visual processing than the site of stereo correspondence. Harris and Willis (2001) reported a similar reduction in dichoptic masking following stereo matching. When two sinusoidal components of a contrast-modulated mask of slightly different spatial frequency were presented separately to each eye, the mask appeared as a single surface slanted in depth and masking was greatly reduced compared to when the mask components were presented to the same eye.

Numerous psychophysical studies indicate that overlay masking is the result of two processes. Within-orientation-channel masking is tightly bound to the orientation of the target stimulus (Baker & Meese, 2007; Cass, 2010; Cass, Stuit, Bex, & Alais, 2009; Foley, 1994). The other component, cross-orientation masking, is far more broadly tuned for orientation and exerts a multiplicative (divisive) effect with respect to mask contrast (Baker & Meese, 2007; Cass, 2010; Foley, 1994; Legge & Foley, 1980). Interestingly, cross-orientation masking effects are observed in the outputs of single neurons located in primary visual cortices of various mammals, including cat (DeAngelis, Robson, Ohzawa, & Freeman, 1992), primate (Carandini, Heeger, & Movshon, 1997), and at population levels of analysis in humans (Burr & Morrone, 1987). The mechanisms of cross-orientation masking remain controversial, with some evidence favoring precortical or thalamo-cortical mechanisms (Freeman, Durand, Kiper, & Carandini, 2002; Li, Thompson, Duong, Peterson, & Freeman, 2006; Li, Peterson, Thompson, Duong, & Freeman, 2005; Priebe & Ferster, 2006), and others, an intra-cortical locus (Allison, Smith, & Bonds, 2001; Heeger, 1992; MacEvoy, Tucker, & Fitzpatrick, 2009; Morrone, Burr, & Maffei, 1982). It is of interest to determine whether the release from masking afforded by binocular disparity is specifically related to either the orientation-tuned or the cross-orientation components.

The current study is unable to distinguish at what stage in visual processing the release from masking with disparity occurs, specifically whether it occurs before or after binocular processing. We use the term "release from masking" throughout the paper as a description of the behavioral data and do not imply a mechanistic interpretation. One possible explanation for the increase in masking when the target and mask occupy the same depth plane is that there is a greater overlap between the populations of disparity-tuned neurons that respond to the target and mask. Recent psychophysical evidence indicates that orientation-specific masking produces similar results monoptically and dichoptically, suggesting it

may be mediated either at the site of, or following, binocular combination (i.e., intra-cortically; Cass et al., 2009). However, other evidence supports an early site for cross-orientation masking, before binocular combination (Baker, Meese, & Summers, 2007; Meese & Baker, 2009).

Summary and conclusions

We have confirmed that a broadband target can be detected at a lower contrast when it is presented in relative depth to a natural image mask than if it occupies the same depth plane. Importantly, we are able to rule out interocular decorrelation as an alternative explanation for the effect by using a discrimination task in which the orientation of the target must be identified. Thus, stereo depth provides a partial release from contrast masking and is a useful cue for image segmentation. This is an example of stereopsis breaking camouflage in a static scene. Within the range of disparities tested, the level of masking decreased monotonically from the peak at zero relative disparity until 8 arcmin of crossed or uncrossed disparity. The results indicate that a small amount of relative stereo depth between a mask and target will assist in segmentation and that there is not generally an increasing benefit from greater disparity.

Acknowledgments

Commercial relationships: none.
Corresponding author: Ms. Susan G. Wardle.
Email: swardle@psych.usyd.edu.au.
Address: Brennan MacCallum Building (A18), Sydney, NSW 2006, Australia.

Footnotes

[1]All *t*-tests reported here are pairwise comparisons using the Bonferroni correction to control error rate at 0.05.

[2]To check whether spectral slope was related to the degree of masking, we calculated the slope for each natural image mask and correlated the slopes with the amount of masking at zero disparity. Slopes ranged from −1.3 to −1.7, but there was no relationship between slope and the degree of masking.

[3]This is conventional terminology in the masking literature; it does not imply any particular depth order.

References

Allison, J. D., Smith, K. R., & Bonds, A. B. (2001). Temporal-frequency tuning of cross-orientation suppression in the cat striate cortex. *Visual Neuroscience, 18,* 941–948.

Baker, D. H., & Meese, T. S. (2007). Binocular contrast interactions: Dichoptic masking is not a single process. *Vision Research, 47,* 3096–3107.

Baker, D. H., Meese, T. S., & Summers, R. J. (2007). Psychophysical evidence for two routes to suppression before binocular summation of signals in human vision. *Neuroscience, 146,* 435–448.

Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision, 9*(10):1, 1–19, http://www.journalofvision.org/content/9/10/1, doi:10.1167/9.10.1. [PubMed] [Article]

Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature, 373,* 609–611.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Brooks, K., & Gillam, B. (2006a). Quantitative perceived depth from sequential monocular decamouflage. *Vision Research, 46,* 605–613.

Brooks, K. R., & Gillam, B. J. (2006b). The swinging doors of perception: Stereomotion without binocular matching. *Journal of Vision, 6*(7):2 685–695, http://www.journalofvision.org/content/6/7/2, doi:10.1167/6.7.2. [PubMed] [Article]

Brooks, K. R., & Gillam, B. J. (2007). Stereomotion perception for a monocularly camouflaged stimulus. *Journal of Vision, 7*(13):1, 1–14, http://www.journalofvision.org/content/7/13/1, doi:10.1167/7.13.1. [PubMed] [Article]

Burr, D., & Morrone, M. C. (1987). Inhibitory interactions in the human-vision system revealed in pattern-evoked potentials. *The Journal of Physiology, 389,* 1–21.

Carandini, M., Heeger, D., & Movshon, J. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17,* 8621–8644.

Cass, J. (2010). Mutual effects of orientation and contrast within and between the eyes: From summation to suppression [Abstract]. *Journal of Vision, 10*(7):1374,

1374a, http://www.journalofvision.org/content/10/7/1374, doi:10.1167/10.7.1374.

Cass, J., Stuit, S., Bex, P., & Alais, D. (2009). Orientation bandwidths are invariant across spatiotemporal frequency after isotropic components are removed. *Journal of Vision, 9*(12):17, 1–14, http://www.journalofvision.org/content/9/12/17, doi:10.1167/9.12.17. [PubMed] [Article]

Cook, M., & Gillam, B. (2004). Depth of monocular elements in a binocular scene: The conditions for da Vinci stereopsis. *Journal Experimental Psychology: Human Perception and Performance, 30,* 92–103.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research, 31,* 2195–2207.

DeAngelis, G. C., Robson, J. G., Ohzawa, I., & Freeman, R. D. (1992). Organization of suppression in receptive-fields of neurons in cat visual-cortex. *Journal of Neurophysiology, 68,* 144–163.

Einhäuser, W., Kruse, W., Hoffmann, K. P., & König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Research, 46,* 1194–1209.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4,* 2379–2394.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms—Masking experiments require a new model. *Journal of the Optical Society of America A, 11,* 1710–1719.

Freeman, T. C. B., Durand, S., Kiper, D. C., & Carandini, M. (2002). Suppression without inhibition in visual cortex. *Neuron, 35,* 759–771.

Gillam, B., Blackburn, S., & Nakayama, K. (1999). Stereopsis based on monocular gaps: Metrical encoding of depth and slant without matching contours. *Vision Research, 39,* 493–502.

Greenwood, J. A., & Edwards, M. (2006). Pushing the limits of transparent-motion detection with binocular disparity. *Vision Research, 46,* 2615–2624.

Harris, J. M., & Willis, A. (2001). A binocular site for contrast-modulated masking. *Vision Research, 41,* 873–881.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9,* 181–197.

Henning, G. B., & Hertz, B. G. (1973). Binocular masking level differences in sinusoidal grating detection. *Vision Research, 13,* 2455–2463.

Henning, G. B., & Hertz, B. G. (1977). Influence of bandwidth and temporal properties of spatial noise on binocular masking-level differences. *Vision Research, 17,* 399–408.

Julesz, B. (1971). *Foundations of cyclopean perception.* Chicago: University of Chicago Press.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70,* 1458–1471.

Lehmkuhle, S., & Fox, R. (1980). Effect of depth separation on metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 605–621.

Li, B., Thompson, J. K., Duong, T., Peterson, M. R., & Freeman, R. D. (2006). Origins of cross-orientation suppression in the visual cortex. *Journal of Neurophysiology, 96,* 1755–1764.

Li, B. W., Peterson, M. R., Thompson, J. K., Duong, T., & Freeman, R. D. (2005). Cross-orientation suppression: Monoptic and dichoptic mechanisms are different. *Journal of Neurophysiology, 94,* 1645–1650.

MacEvoy, S. P., Tucker, T. R., & Fitzpatrick, D. (2009). A precise form of divisive suppression supports population coding in the primary visual cortex. *Nature Neuroscience, 12,* 637–645.

McDonald, J. S., & Tadmor, Y. (2006). The perceived contrast of texture patches embedded in natural images. *Vision Research, 46,* 3098–3104.

McKee, S., Levi, D., & Bowne, S. (1990). The imprecision of stereopsis. *Vision Research, 30,* 1763–1779.

McKee, S. P., Bravo, M. J., Taylor, D. G., & Legge, G. E. (1994). Stereo matching precedes dichoptic masking. *Vision Research, 34,* 1047–1060.

McKee, S. P., Watamaniuk, S. N., Harris, J. M., Smallman, H. S., & Taylor, D. G. (1997). Is stereopsis effective in breaking camouflage for moving targets? *Vision Research, 37,* 2047–2055.

Meese, T. S., & Baker, D. H. (2009). Cross-orientation masking is speed invariant between ocular pathways but speed dependent within them. *Journal of Vision, 9,* 1–15, http://www.journalofvision.org/content/9/5/2, doi:10.1167/9.5.2. [PubMed] [Article]

Moraglia, G., & Schneider, B. (1990). Effects of direction and magnitude of horizontal disparities on binocular unmasking. *Perception, 19,* 581–593.

Moraglia, G., & Schneider, B. (1992). On binocular unmasking of signals in noise—Further tests of the summation hypothesis. *Vision Research, 32,* 375–385.

Morrone, M. C., Burr, D. C., & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. 1. Neurophysiological evidence. *Proceedings of the Royal Society of London B: Biological Sciences, 216,* 335–354.

Ogle, K., & Weil, M. (1958). Stereoscopic vision and the duration of the stimulus. *Archives of Ophthalmology, 59,* 4–17.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Priebe, N. J., & Ferster, D. (2006). Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nature Neuroscience, 9,* 552–561.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals 1. Psychophysics. *Journal of Neuroscience, 14,* 7357–7366.

Schneider, B., Moraglia, G., & Jepson, A. (1989). Binocular unmasking—An analog to binaural unmasking. *Science, 243,* 1479–1481.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics, 12,* 229–232.

Tyler, C. W., Chan, H., Liu, L., McBride, B., & Kontsevich, L. (1992). Bit stealing: How to get 1786 or more gray levels from an 8-bit color monitor. 351–364.

Tyler, C. W., & Julesz, B. (1978). Binocular cross-correlation in time and space. *Vision Research, 18,* 101–105.

Watson, A. B., & Pelli, D. G. (1983). Quest—A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120.

Watt, R. (1987). Scanning from coarse to fine spatial scales in the human visual-system after the onset of a stimulus. *Journal of the Optical Society of America A, 4,* 2006–2021.