ORIENTATION-DEFINED BOUNDARIES ARE DETECTED WITH LOW EFFICIENCY

Joshua A. Solomon* < J.A.Solomon@city.ac.uk , +44 20 7040 0192>

Michael J. Morgan < M.J.Morgan@city.ac.uk >

Centre for Applied Vision Research

City, University of London

*Corresponding Author

When compared with other summary statistics (mean size, size variance, orientation variance), visual estimates of average orientation are inefficient. Observers act as if they use information from no more than two or three items. We hypothesised that observers would attain greater sampling efficiency when their task was to perform a texture segmentation rather than a did not require an explicit representation of mean orientation. We tested this hypothesis using a texture-segmentation task. Two arrays of 32 wavelets each were presented; one left and one right of fixation. Orientations in the target array were sampled from wrapped normal distributions having two different means with the same variance. One distribution defined orientations above the horizontal meridian, the other defined orientations below the meridian. All orientations in the other array were defined by a single wrapped normal distribution having the same variance as each of the distributions in the target array. Contrary to our hypothesis, results indicate that observers effectively ignored all but one item from the top and bottom of each array. In fact, we found no change in the threshold difference between the target's two means when all but one item from the top and bottom of each array were removed. We are forced to conclude that the visual system does not compute the average of more than a few orientations, even for texture segmentation.

Introduction

Human observers are thought to be fairly adept at perceptual tasks that require statistical summaries of feature content. In order to quantify this proficiency, we typically use the mathematical concept of efficiency (Fisher, 1925). Given any sample size N, efficiency is the ratio of M to N, where M is the sample size that the ideal observer would need in order to estimate a statistic with the same precision as a human observer.

For example, when presented with two sets, containing ($N = 8$) circles each, human observers can select the set whose circles have the larger average diameter as well as the otherwise-ideal observer that perfectly measures the diameters of $M = 5$ randomly selected circles in each set (Solomon, Morgan, & Chubb, 2011; Gorea, Belkoura, & Solomon, 2014). Thus, we can say that the efficiency of size averaging can be as high as 5/8 or 62.5%. The efficiency for discriminating between sets of circles on the basis of the variance in their diameters is similarly high (Solomon, et al., 2011), as is the efficiency for discriminating between sets of wavelets on the basis of the variance in their orientations (Solomon, 2010).

Efficiency typically falls when the display set-size N is increased. The effective set-size M, on the other hand, may be more resilient to this manipulation. For example, Morgan, Mareschal, Chubb, & Solomon (2011) reported that human observers discriminated between dot patterns with different levels of positional variance as well as an ideal observer that used 5 or 6 randomly selected dots, regardless whether each dot-patterns contained 11 or 121 dots. Solomon (2010) explicitly tested the null hypothesis that each observer had a maximum effective set-size M_{max}, such that $M = \min\{M_{max}, N\}$ for $N = 1$, 2, 4, and 8, when trying to

discriminate between sets of wavelets on the basis of their average orientation or orientation variance.

Solomon (2010) was unable to reject this null hypothesis, and the maximum effective set-size for orientation averaging was surprisingly low. Maximum-likelihood estimates of M_{max} varied (across observers) between values of 1 and 3. A subsequent review of the literature (Solomon, May, & Tyler, 2016) suggests that low values such as these are fairly typical in tasks that require observers to estimate average orientation.

We do not know what prevents human observers from attaining higher values of M_{max}, but we had an idea. All previous experiments on orientation averaging required observers to estimate the average orientation in an array of oriented stimuli, like Gabor patterns or line segments, and compare that average with something else, like another average or the vertical meridian. We hypothesized that effective set-sizes might be greater if the task did not require observers to form explicit estimates of average orientation. Instead, we designed the task described below, in which observers looked for a boundary between arrays having different average orientations. Contrary to our hypothesis, observers did no better than an otherwise-ideal observer that ignored all but one item from each array. In a second experiment, we found no change in performance when all but one item was removed from each array.

Methods

This experiment was conducted in compliance of the Declaration of Helsinki, where applicable. It was approved by City University London's Senate Ethics panel, in conjunction with the EPSRC project ''The Efficiency of Visual Statistics'' (see Acknowledgment). Data were collected from the two authors plus a third

psychophysicist, JF, who gave his informed consent to participate. He was selectively recruited on the basis of his high efficiency for gaze-averaging (Florey, Clifford, Dakin, & Mareschal, 2016).

Stimuli were generated and responses were collected on a MacBook Pro computer, with the brightness turned all the way up, so that background luminance was 20 cd/m^2. When luminance and/or contrast is an experimental variable, it is important to correct display nonlinearities. Neither luminance nor contrast is a variable in our experiments, consequently it was not necessary to correct for the MacBook's native gamma function. The Psychtoolbox (Brainard, 1997; Pelli, 1997) was used for stimulus generation. Psychophysica (Watson & Solomon, 1997) was used for data analysis. Both codes are available upon request. Head positions were not restrained, but observers were asked to maintain a comfortable viewing distance (~0.65 m) for the duration of the experiment.

On each trial of Experiment 1, observers were presented with four arrays, containing N=16 wavelets each. Two arrays appeared on the left side of fixation. Together, they formed a donut-shaped configuration. The other two arrays formed a donut-shaped configuration on the right side of fixation. (See Figure 1.) All four arrays appeared simultaenously, and remained visible for 1.7 s.

Donut centers were positioned on the horizontal meridian, 4.8 degrees of visual angle away from a small fixation spot, which remained visible throughout the experiment. The visual angle between the center of each wavelet and the center of its donut was 1.5, 2.4, 3.2, or 3.6 degrees.

Experiment 2 was virtually identical to Experiment 1, except that each array contained only N=1 wavelet. That is, there were 4 wavelets on each trial. Their centers were positioned 4.8 degrees of visual angle left and right of the vertical

meridian and 2.8 degrees above and below the horizontal meridian. These four wavelets, therefore, occupied positions corresponding to the centers of the four arrays in Experiment 1.

Each wavelet in both experiments was specified by the Gabor function: the product of a 1-dimensional sinusoid and a 2-dimensional Gaussian blob. The sinusoid had a spatial frequency of 2.4 cycles per degree and the blob had a space constant (i.e. the Gaussian standard deviation) of 0.31 degrees of visual angle. As the specific luminance profile of each wavelet was not critical to this study, no attempt was made to correct for the Mac's native gamma function ($\gamma = 2.35$). Consequently, our wavelets were not true Gabor patterns. Each was presented at maximum contrast. Its spatial phase was selected independently from a uniform distribution over all 2π radians.

Figure 1. Sample stimulus from Experiment 1. Observers reported whether the orientation-defined texture boundary was left or right of fixation. This is a particularly easy trial!

On each trial, we randomly selected the wavelets left or right of the vertical meridian to be the "target." Wavelets on the other side of the vertical meridian were therefore the "nontarget." Orientations in the target were independently sampled from

two wrapped normal distributions. One distribution defined orientations above the horizontal meridian (i.e. in the top half of the target), the other distribution defined orientations below the meridian (i.e. in the bottom half of the target). The two distributions had different expected values (μ_{TT} and μ_{TB} for the top and bottom halves of the target, respectively), but the same standard deviation, σ. Orientations in the both halves of the nontarget (indexed by NT and NB) were defined by a single wrapped normal distribution having the same standard deviation. The expected values of all three wrapped normal distributions were randomly selected from a uniform distribution over all π radians. However, the expected values of the target distributions were correlated, as described in the next paragraph.

Observers were instructed to report whether the target was right or left of fixation. They were also asked to maintain fixation at the center of the laptop screen, but we did not enforce compliance with this additional request. For reasons described in the Modeling section, we used a minimum of two standard deviations (σ) with each observer. Both of these standard deviations were fairly small, so that the wrapped normal distributions were well-approximated by normal distributions.

Standard deviations in used in Experiment 1 were, for JAS, 0, 4°, 8°, and 16°; for MJM, 0, 6°, and 8°; and for JF, 0 and 8°. Standard deviations used in Experiment 2 were, for JAS, 0 and 16°; for MJM, 0 and 8°; and for JF, 0 and 8°. For each observer, each experiment, and each standard deviation, we used at least 4 66-trial QUEST staircases (Watson & Pelli, 1983) to obtain independent, maximum-likelihood estimates of the "threshold" angle, $\Delta\mu=|\mu_{TT}-\mu_{TB}|$, between the expected values of the two distributions defining the target, required for 81%-correct responses.

Modeling

On each trial, the *ideal observer* perfectly measures the orientation of each wavelet in the display. The ideal observer then calculates the sample mean orientations ($\bar{\theta}_{\mathrm{TT}}, \bar{\theta}_{\mathrm{TB}}, \bar{\theta}_{\mathrm{NT}}$, and $\bar{\theta}_{\mathrm{NB}}$) in each half-donut. The ideal observer responds correctly if and only if $\left|\bar{\theta}_{\mathrm{TT}} - \bar{\theta}_{\mathrm{TB}}\right| > \left|\bar{\theta}_{\mathrm{NT}} - \bar{\theta}_{\mathrm{NB}}\right|$. Thus, the ideal observer's probability correct can be written:

$$P_{\mathrm{ideal}}(C) = P\left[\frac{\left(\bar{\theta}_{\mathrm{TT}} - \bar{\theta}_{\mathrm{TB}}\right)^2}{\left(\bar{\theta}_{\mathrm{NT}} - \bar{\theta}_{\mathrm{NB}}\right)^2} > 1\right]. \quad (1)$$

Recall that each orientation was independently selected from a distribution that was well-approximated by the normal distribution. Consequently, the differences between sample means are even-better approximated by independent normal distributions, $\bar{\theta}_{\mathrm{TT}} - \bar{\theta}_{\mathrm{TB}} \sim \mathcal{N}(\Delta\mu, 2\sigma^2/N)$ and $\bar{\theta}_{\mathrm{NT}} - \bar{\theta}_{\mathrm{NB}} \sim \mathcal{N}(0, 2\sigma^2/N)$, and

$$P_{\mathrm{ideal}}(C) = 1 - F(1), \quad (2)$$

where F is the cumulative distribution function of the non-central F-ratio distribution, with 1 degree of freedom in the numerator and denominator and non-centrality parameter $\lambda = N(\Delta\mu)^2/(2\sigma^2)$. The ideal observer's threshold is the value of $\Delta\mu$ at which $P_{\mathrm{ideal}}(C) = 0.81$.

In order to establish the efficiency of human performance, it is necessary to compare it with the inefficient observer (IO). The IO is ideal, except it uses only a fraction (M/N) of the available information. Thus it responds correctly if and only if $\left|\bar{\theta}^*_{\mathrm{TT}} - \bar{\theta}^*_{\mathrm{TB}}\right| > \left|\bar{\theta}^*_{\mathrm{NT}} - \bar{\theta}^*_{\mathrm{NB}}\right|$, where $\bar{\theta}^*_{\mathrm{TT}} - \bar{\theta}^*_{\mathrm{TB}} \sim \mathcal{N}[\Delta\mu, 2\sigma^2_{\mathrm{E}}/M]$ and $\bar{\theta}^*_{\mathrm{NT}} - \bar{\theta}^*_{\mathrm{NB}} \sim \mathcal{N}[0, 2\sigma^2_{\mathrm{E}}/M]$. Its probability correct can be written:

$$P_{\mathrm{IO}}(C) = P\left[\frac{\left(\bar{\theta}^*_{\mathrm{TT}} - \bar{\theta}^*_{\mathrm{TB}}\right)^2}{\left(\bar{\theta}^*_{\mathrm{NT}} - \bar{\theta}^*_{\mathrm{NB}}\right)^2} > 1\right] \quad (3)$$

$$= 1 - F^*(1) \ , \tag{4}$$

where F^* is the cumulative distribution function of the non-central F-ratio distribution, with 1 degree of freedom in the numerator and denominator and non-centrality parameter $\lambda^* = M(\Delta\mu)^2/(2\sigma^2)$, and its threshold is the value of $\Delta\mu$ at which $P_{\mathrm{IO}}(C) = 0.81$. It should be noted that the IO is ideal when $M = N$.

The purpose of our experiment was to see whether efficiencies would increase when explicit representations of average orientation were not required. For establishing efficiency, the IO provides a sufficient baseline against which human performance can be compared (see Fig. 2 for examples). Consequently, it isn't strictly necessary to consider more complicated models of performance. However, we know that the IO is a poor model for human behavior because, whatever the value M, its threshold will be proportional to the standard deviation of external noise (σ). Consequently, it cannot successfully account for human performance in the absence of external noise.

Equivalent-noise models (Nagaraja, 1964; Pelli, 1990; Dakin, 2001) can. In addition to inefficiency, these models posit an internal noise, which decreases the fidelity with which stimuli are represented in the visual system. With an appropriate distribution, the addition of external noise can mimic the effects of internal noise. For that reason, such external noise is known as ‘‘equivalent’’ noise when its variance matches that of the internal noise. The effect of equivalent noise on performance is negligible when external noise has a much greater standard deviation. "High-noise" (Pelli & Farell, 1990) conditions such as this are required for estimates of efficiency that are uncontaminated by the imprecision with which individual stimuli are represented.

As in most applications of signal-detection theory (Green & Swets, 1966), we assume that internal noise has a normal distribution. Within the visual system, the orientation of each item is perturbed by an independent sample of this internal noise. Consequently, if the variance of internal noise is denoted σ_{E}^2, then noisy, inefficient (but otherwise ideal) human observers will get more-widely distributed values ($\bar{\theta}_{\mathrm{TT}}^{**}, \bar{\theta}_{\mathrm{TB}}^{**}, \bar{\theta}_{\mathrm{NT}}^{**}$, and $\bar{\theta}_{\mathrm{NB}}^{**}$), when they try to calculate the sample means of each half-donut, such that $\bar{\theta}_{\mathrm{TT}}^{**} - \bar{\theta}_{\mathrm{TB}}^{**} \sim \mathcal{N}[\Delta\mu, 2(\sigma^2 + \sigma_{\mathrm{E}}^2)/M]$ and $\bar{\theta}_{\mathrm{NT}}^{**} - \bar{\theta}_{\mathrm{NB}}^{**} \sim \mathcal{N}[0, 2(\sigma^2 + \sigma_{\mathrm{E}}^2)/M]$.

Given these calculations, the noisy, inefficient observer's (NIO) decision process parallels that of the IO and ideal observer. Its probability correct can be written:

$$P_{\mathrm{NIO}}(C) = P\left[\frac{\left(\bar{\theta}_{\mathrm{TT}}^{**} - \bar{\theta}_{\mathrm{TB}}^{**}\right)^2}{\left(\bar{\theta}_{\mathrm{NT}}^{**} - \bar{\theta}_{\mathrm{NB}}^{**}\right)^2} > 1\right] \tag{5}$$

$$= 1 - F^{**}(1)\ , \tag{6}$$

where F^{**} is the cumulative distribution function of the non-central *F*-ratio distribution, with 1 degree of freedom in the numerator and denominator and non-centrality parameter $\lambda^{**} = M(\Delta\mu)^2/[2(\sigma^2 + \sigma_{\mathrm{E}}^2)]$, and its threshold is the value of $\Delta\mu$ at which $P_{\mathrm{NIO}}(C) = 0.81$.

Results

Each independently measured estimate of threshold appears as a small dot in Figure 2. Larger symbols illustrate maximum-likelihood estimates of these thresholds, based on all the responses in each condition. Blue and amber symbols, illustrating data from Experiments 1 and 2, respectively, have been slightly nudged right and left,

respectively, for greater legibility. In general, we found that thresholds increased with the standard deviation of wavelet orientations, and thresholds collected in Experiment 2 (with N=1 wavelet per array) were similar to thresholds collected in Experiment 1 (with N=16 wavelets per array). This similarity is inconsistent with our hypothesis. Specifically, it does not support the idea that texture boundaries can be detected by averaging orientation content across multiple elements.

Further evidence against our hypothesis was obtained by fitting the IO and NIO models to the thresholds depicted in Figure 2. Both models fit the data from JAS and MJM best when its calculations were based on just M = 1 wavelet pattern from each side of the potential texture boundaries. Observer JF was only slightly more efficient. For this subject, the NIO fit best when using M = 1.3 wavelets from each side of the potential texture boundaries.[1]

[1] Non-integer values such as this can be interpreted in two ways. One possibility is that they reflect a mixture of effective set sizes. For example, on some trials the observer might have used just one randomly selected element from each side of the potential texture boundaries, whereas on other trials the average of two or more elements was computed. In this case, M would reflect the root-mean-square of the mixture of effective set sizes. Alternatively, non-integer values of M might reflect an unequal weighting of two or more texture elements in calculations of the average orientation on each side of a potential texture boundary.

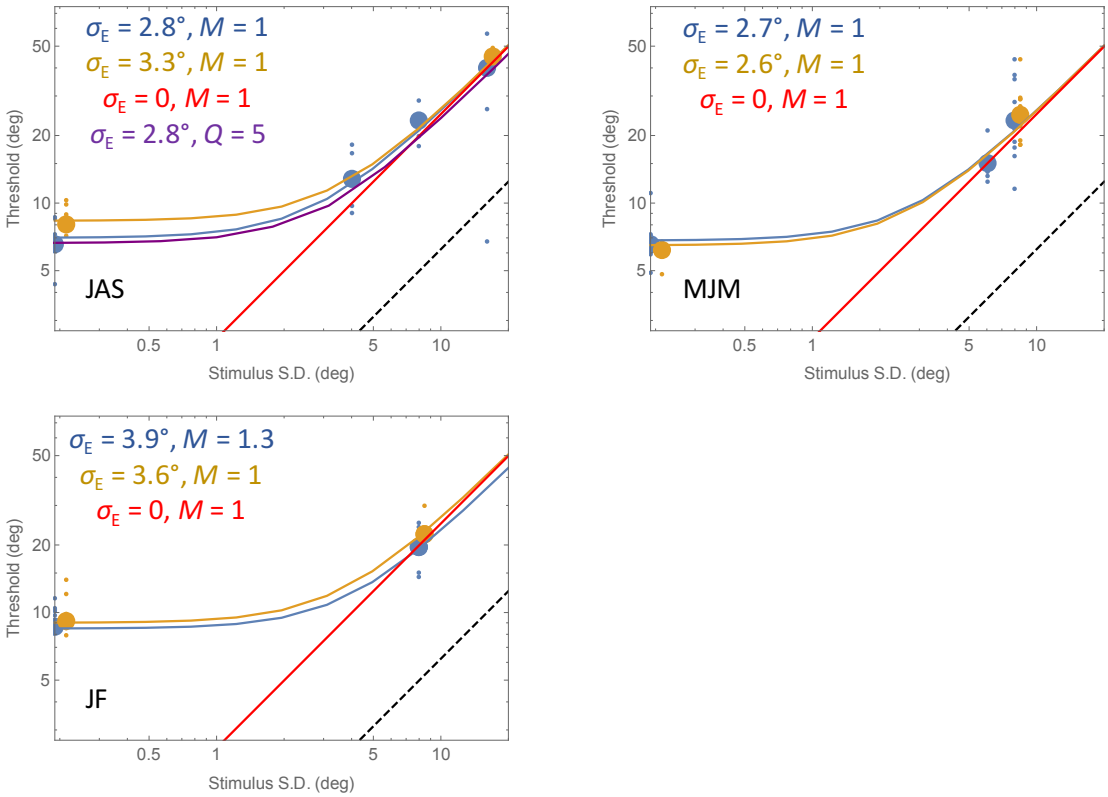

Figure 2 (online version; blue, amber, red, and purple can be replaced by black, gray, dotted and dot-dashed for print). Results of Experiments 1 and 2. Threshold is the difference between the expected orientations in the top and bottom halves of the target "donut" for 81% response accuracy in our two-alternative, forced-choice task. (See Figure 1.) Non-target and target orientations had the same standard deviation (”Stimulus S.D.”) on each trial. Each small blue symbol illustrates the result of a single block of 66 trials with two 32-wavelet donuts. Some of these have been obscured by the large blue symbols, which illustrate the single, maximimum-likelihood estimate of threshold for each Stimulus S.D. Solid blue curves illustrate the performance of the noisy, inefficient (but otherwise-ideal) observer, whose decisions are based on *M* randomly selected wavelets from the top and bottom of each donut. Best-fitting values of the equivalent input noise and *M* are given in each panel. Amber symbols and curves illustrate analogously derived thresholds from the trials in which all but one wavelet was removed from the top and bottom of each donut. Black dashed lines illustrate the performance of the (noiseless) ideal observer and red lines illustrate the performance of the (noiseless) inefficient observer, whose (otherwise ideal) decisions are based on 1 randomly selected wavelet from the top and bottom of each donut. The purple curve illustrates the performance of an observer whose decisions are based on the orientation variances in 5-wavelet samples from each donut. Prior to computing variances, the orientation of each wavelet in each sample was perturbed by an equivalent input noise identical to that inferred using the noisy, inefficient observer.

Discussion

When attempting to detect a boundary between two texture arrays, differing in their average orientations, our best observer performed as poorly as an otherwise-ideal detector, whose calculations were based on fewer than 2 elements per texture array. The other two observers performed as if they had ignored all but one texture element in each array. Indeed, their performances were largely unaffected when all but one texture element was removed from each array. We consider these data interesting in their failure to support our hypothesis that visual system really is capable of computing average orientation. We must conclude that the visual system does not compute the average of more than a few orientations, even for texture. This is in addition to its well-documented low efficiencies (e.g. Solomon, et al., 2016) in tasks that require explicit estimates of that average.

We must stress that an effective set-size of 1 does not imply use of a single element within each array. For example, consider an observer whose decisions are based on the sample variances of Q randomly selected wavelets from each 32-wavelet donut. Monte Carlo simulations indicate that this observer would have an effective set-size of 1 (i.e. efficiency would be 1/16) when $Q = 5$, a value well within conventional estimates for the effective set-size in tasks where the ideal strategy requires the computation of orientation variance (Morgan et al., 2008; Solomon, 2010). Moreover, the performance of this "variance discriminator" is virtually identical to that of the NIO (with an effective set-size of 1 and comparable equivalent noise), regardless of external noise. This can be seen by comparing the blue and purple curves in Fig. 2.

Standard “back-pocket” models of texture segregation (Chubb & Landy, 1991)

typically hypothesize "second-order" mechanisms that are selective for regional (i.e. relatively large-scale) variation in the (rectified) output of more localized "first-order" mechanisms that share a preference for the same stimulus orientation. If the wavelets in our stimuli were well-matched to the receptive fields of first-order mechanisms, then intuition would suggest that second-order mechanisms would respond to input stemming from a relatively large pool of our wavelets. Therefore, we consider our very small values of effective set-size to be counter-intuitive. Either our intuition about back-pocket-model efficiency is wrong, or observers use something other than large-scale second-order filters, when detecting orientation-defined texture boundaries.

Acknowledgment

This research was sponsored by the Engineering and Physical Sciences Research Council (Grant EP/H033955/1).

References

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

Chubb, C., & Landy, M. S. (1991). Orthogonal distribution analysis: A new approach to the study of texture perception. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 291-301). Cambridge, MA: MIT Press.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A – Optics Image Science and Vision, 18*(5), 1016–1026.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, UK and London: Oliver and Boyd.

Florey, J., Clifford, C. W. G., Dakin, S., & Marschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports, 6,* 32210. doi:10.1038/srep32210.

Gorea, A., Belkoura, S., & Solomon, J. A. (2014) Summary statistics for size over space and time. *Journal of Vision, 14*(9):22, 1-14. doi: 10.1167/14.9.22.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Morgan, M. J., Mareschal, I., Chubb, C. & Solomon, J. A. (2012) Perceived pattern regularity computed as a summary statistic: Implications for camouflage. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 279,* 2754-2760.

Nagaraja, N. S. (1964). Effect of luminance noise on contrast thresholds. *Journal of the Optical Society of America, 54,* 950–955.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. B. Blakemore (Ed.), *Vision: Coding and efficiency.* Cambridge: Cambridge University Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10,* 437-42.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A, 16,* 647-653.

Solomon, J. A. (2010) Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision, 10*(14):19, 1-16. doi: 10.1167/10.14.19.

Solomon, J. A., May, K. A., & Tyler, C. W. (2016) Inefficiency of orientation averaging: evidence for hybrid serial/parallel temporal integration. *Journal of Vision, 16*(1):13, 1–7, doi:10.1167/16.1.13.

Solomon, J. A., Morgan, M. & Chubb, C. (2011) Efficiencies for the statistics of size discrimination. *Journal of Vision, 11*(12):13, 1-11. doi: 10.1167/11.12.13.

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

Watson, A. B. & Solomon, J. A. (1997) Psychophysica: Mathematica notebooks for psychophysical experiments. *Spatial Vision, 10*, 447-466.