Response

Dear Editor,

Ruderman's letter hinges on two points, which we will now demonstrate to be unsupportable.

His first point was based on our use of a limited range of spatial frequencies for estimating power spectra. For instance, we used such a range in our calculation of the power spectra of his occlusion-based images with exponential correlation (Balboa, Tyler, & Grzywacz, 2001; Ruderman, 1997). In this range (from 5 to 250 cycles/image), these images produced spectra consistent with published data on natural images. In other words, the ensemble produced spectra that fell as $f^{-\alpha}$, where f was spatial frequency and α a constant parameter. (Such a fall is often called scaling.) To mount a counterargument, he built one of his occlusion-based images and showed that the spectrum did not fall at low frequencies. His point was that our spectra only appeared to fall as $f^{-\alpha}$ because we plotted them over a limited range of frequencies. He then said, "spectra of natural images... have demonstrated scaling... over their entire frequency range", thus claiming to have refuted our conclusions. We agree with him that occlusion-based images will often be flatter at low frequencies. We even point this out in our paper.

However, we do not agree that spectra of natural images fall as $f^{-\alpha}$ over the entire frequency range. Some authors who claim scaling in natural images present spectra in ranges of spatial frequencies that are even narrower than the one we use (Field, 1987). Other authors who use a larger frequency range report significant curvatures in double-logarithmic plots of their spectra (Dong & Atick, 1995; Tolhurst, Tadmor, & Chao, 1992). These curvatures are similar to those that we predict in our paper.

Moreover, when Ruderman speaks about "the entire frequency range", he is speaking loosely, but he presumably means that one should look at the entire available frequency spectrum. In other words, one should use frequencies as low as 1 cycle/image. The trouble is that, in this case, the size of the image is what determines the lowest spatial frequency that one can measure. What would happen if one could measure spectra from larger images? To understand what would happen in principle, one must assume an infinitely large image, which would allow us to measure power at arbitrarily low frequencies. In this case, it is possible to calculate the integral of the power spectrum from $f = 0$ to F. This integral is the contribution to the variance attributed to frequencies within this interval. This partial variance is

$$\sigma_{0,F}^2 = \int_0^F \mathrm{d}f\, 2\pi f P(f), \tag{1}$$

where $P(f)$ is the rotationally averaged power spectrum. If, as asserted by Ruderman, $P(f) = kf^{-\alpha}$ over the "entire frequency range", and if $\alpha > 2$, as is the case for about half of natural images, (Field, 1987; van der Schaaf & van Hateren, 1996), then

$$\sigma_{0,F}^2 = \lim_{\varepsilon \to 0} \int_\varepsilon^F \mathrm{d}f\, 2\pi f \frac{k}{f^{\alpha}} = \infty. \tag{2}$$

In other words, the variance would be infinite, which is inapplicable to natural images. Hence, if one could measure the power spectrum of a natural image literally over the "entire frequency range", then this spectrum would have to become flatter than $f^{-\alpha}$ at low spatial frequencies. Flattening at low frequencies is the only way to prevent the variance from diverging to infinity. This is the behavior that Ruderman shows in Fig. 2 of his letter and which we discuss in our paper.

However, Ruderman may say that this argument is beside the point, because images are never infinitely large. But this is only a practical matter. In principle, images can be built to encompass arbitrarily large fields of view. For instance, in photographic cameras, images form in a focal plane where the film or the electronic sensors reside. In principle, this plane could be infinite, allowing the entire infinite world to be captured by the image. Of course, this does not happen in practice, because films are finite, and because lenses have small fields of view or distort the images. However, there is no reason why one could not build arbitrarily large films. Furthermore, one can circumvent the lens limitation by using a pinhole camera. Such cameras can form well defined, practically undistorted images across an extremely wide angular field and over a large range of distances (Hecht, 1998). The problem with these cameras is that they require long exposure times. Nevertheless, there is nothing to prevent exposure times from being arbitrarily large and nothing to prevent using different exposure times in different parts of the image if necessary.

Consequently, because images can in principle be arbitrarily large, the issue is not whether natural spectra become flatter at low frequencies, but at what frequency they begin to do so. The only reason that such flattening is not observed in some studies is that they use a constrained range of frequencies, due to the physical limitation of their cameras. Commercially available cameras have notoriously small fields of view. Therefore, there is no hope that one can measure power at very low spatial frequencies with such cameras. To be completely fair, one should publish spectra with frequencies expressed in cycles/deg not cycles/image. In Ruderman's own work, the lowest spatial frequency is about 0.06 cycle/deg (Ruderman & Bialek, 1994). If this corresponds to 1 cycle/image, then his images have fields of view of only about 17°. This is much smaller than the fields of view of biological systems, which thus have access to much lower spatial frequencies than he measured.

Under such limited range, Ruderman's occlusion-based images with exponential correlation do indeed have power spectra that fall as $f^{-\alpha}$. However, he provides no argument against the spectra becoming flatter at the lowest frequencies.

Ruderman's second point relates to Eq. (4) of our appendix. That equation considers a one-dimensional (1D) cut of an occlusion-based image. Such an image consists of M regions of uniform intensity such that the jth region spans $x_j \leqslant x \leqslant x_{j+1}$ and its intensity is $\hat{I}_j$. Eq. (4) of our appendix is the power spectrum of this cut, namely,

$$\left|\widetilde{I}(f)\right|^2 = \frac{1}{(2\pi f)^2}\left|\sum_{j=1}^{M}\hat{I}_j\big(e^{i2\pi fx_{j+1}} - e^{i2\pi fx_j}\big)\right|^2. \tag{3}$$

According to this equation, as the spatial frequency increases, the power spectrum tends to fluctuate (the parenthetical term on the right hand side), but its envelope falls in inverse proportion to the square of the frequency (the $1/f^2$ term). The only exception occurs at very low frequencies, that is, when $2\pi f \ll 1/(x_{j+1} - x_j)$ for all j. For those frequencies, one can approximate the term inside the parenthesis with the first term of the Taylor series, yielding

$$\left|\widetilde{I}(f)\right|^2 \approx \left|\sum_{j=1}^{M}\hat{I}_j\big(x_{j+1} - x_j\big)e^{i2\pi fx_j}\right|^2. \tag{4}$$

In other words, for frequencies so low that the corresponding periods are longer than the longest step in the 1D cut, the envelope of the power spectrum tends to remain constant with frequency. Ruderman's criticism begins with the observation that a two-dimensional spectra that falls as $f^{-\alpha}$ has 1D cuts with spectrum that fall as $f^{-(\alpha+1)}$. He then concludes, "the approximation

presented by Balboa et al. does not match the spectral behavior seen in the natural image ensembles". However, there are two problems with this conclusion: First, we demonstrated above that it is impossible for natural images to have an $f^{-\alpha}$ spectrum over the "entire frequency range". Therefore, there is no reason to believe that spectra of 1D cuts fall as $f^{-(\alpha+1)}$ over the "entire frequency range". We argued based on Eq. (2) that a flattening should occur at low frequencies, which is what Eq. (4) expresses. Second and more disturbing for Ruderman's arguments, Eq. (3) is not an approximation. Rather, it is the general power spectrum of a 1D cut of an occlusion-based image. And the approximation of this equation by Eq. (4) is unavoidable when $f \rightarrow 0$. Consequently, the power spectrum of a 1D cut of most finite occlusion-based image is flat at low frequencies and fall as frequency squared at high frequencies. But this is essentially the first half of Ruderman's own model (Ruderman, 1997). When he states, "natural images... are collages of regions corresponding to statistically independent objects", he means that these regions have relatively homogeneous properties, such as practically constant intensities. Hence, by criticizing the conclusions obtained with Eq. (4) in our appendix, he is criticizing the first half of his model.

In summary, both of Ruderman's arguments against our paper are contestable. Despite what he says, we showed that when the frequency range of power spectra is broad, they must become flatter than $f^{-\alpha}$ at low frequencies. In addition, we showed that Eq. (4) of our appendix, which Ruderman criticizes, applies to his own model. Fortunately, his criticism is not valid, as one cannot assume scaling over the entire frequency range.

We agree with Ruderman when he states, "discovering which properties give rise to... statistical regularities [of the natural environment] is of great importance... for understanding the design of visual systems (Simoncelli & Olshausen, 2001)." Much of our recent work has been in this direction (Balboa & Grzywacz, 2000a,b,c; Grzywacz & Balboa, 2002). The question here is not whether natural statistics are important, but what are the most relevant ones. In our paper, we suggested that much of power-spectra scaling could be explained by taking into account the power spectrum of occlusion edges. We also suggested that the importance of the distribution of object sizes, as postulated by Ruderman, is an empirical issue that remains to be rigorously demonstrated.

Sincerely,

Acknowledgements

References

Balboa, R. M., & Grzywacz, N. M. (2000a). The minimal local-asperity hypothesis of early retinal lateral inhibition. *Neural Computation, 12*, 1485–1517.

Balboa, R. M., & Grzywacz, N. M. (2000b). Occlusions and their relationship with the distribution of contrast in natural images. *Vision Research, 40*, 2661–2669.

Balboa, R. M., & Grzywacz, N. M. (2000c). The role of early retinal lateral inhibition: more than maximizing luminance information. *Visual Neuroscience, 17*, 77–90.

Balboa, R. M., Tyler, C. W., & Grzywacz, N. M. (2001). Occlusions contribute to scaling in natural images. *Vision Research, 41*, 955–964.

Dong, D., & Atick, J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems, 6*, 345–358.

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, 4*, 2379–2394.

Grzywacz, N. M., & Balboa, R. M. (2002). A Bayesian framework for sensory adaptation. *Neural Computation, 14*, 543–559.

Hecht, E. (1998). *Optics* (3rd ed.). Reading, MA: Addison-Wesley.

Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research, 37*, 3385–3389.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters, 73*, 814–817.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193–1216.

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

van der Schaaf, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research, 36*, 2759–2770.

Norberto M. Grzywacz
Department of Biomedical Engineering and Neuroscience
Graduate Program
University of Southern California
University Park, OHE 500
Los Angeles, CA, USA
E-mail address: nmg@bmsr.usc.edu

Rosario M. Balboa
Departamento de Biotecnología
Universidad de Alicante
Alicante, Spain

Christopher W. Tyler
Kettlewell Institute Eye Research Institute
San Francisco, CA, USA