THOUGHT
A JOURNAL OF PHILOSOPHY
THE NORTHERN INSTITUTE OF PHILOSOPHY

Thought ISSN 2161-2234

ORIGINAL ARTICLE

Seeing and Windows of Integration

Ned Block

New York University

DOI:10.1002/tht.62

I am grateful to Bradley Richards and J. H. Taylor for their thoughtful critiques and for the chance to clarify and re-think the main line of argument in "The Grain of Vision and the Grain of Attention" (Block 2013).

My article concerned peripheral vision. Much of normal vision is peripheral. The fovea is the central area of the retina that is needed for fine discrimination, for example in reading. It subtends only 2° of visual angle, a bit more than the width of the thumbnail at arm's length, so if you are looking at something the size of a hand at arm's length, on any one fixation most of it is seen nonfoveally. All perception degrades in acuity in the periphery because cone cells in the retina decrease with eccentricity. But *object-perception* in the periphery suffers from *more* than a degradation in acuity; it suffers from crowding. Crowding is a phenomenon of peripheral vision in which "things . . . lose the quality of form . . . without losing crispness . . ." (Lettvin 1976). It is widely agreed that the explanation of crowding is that vision involves assigning features to objects; but there are minimal "windows of integration" within which the visual system cannot determine which features are to be bound to which objects. (Binding is the process by which the visual registration of a blue square and a red circle involves blueness being attributed to the square instead of the circle.) The windows of integration grow larger farther from the fixation point—that is, where the eyes are pointing. See Figure 1 for an indication of the size and shape of these windows. In the periphery, the windows can be large enough so that normally two or more objects are in the same window.

I argued that there is conscious object-seeing without object-attention in a phenomenon that I called identity-crowding. Identity-crowding is the special case of crowding in which the crowded items are all the same so that there is no issue of determining which features are bound to which objects. See Figure 2 for an example.

As Taylor and Richards note, my argument was that in identity-crowding, the subject consciously detects the crowded object (in the sense of distinguishing consciously between presence and absence), consciously differentiates the object from the background, consciously discriminates the crowded object from other objects and consciously identifies the crowded object, so it is difficult to see a rationale for denying that

Correspondence to: E-mail: Ned.block@nyu.edu

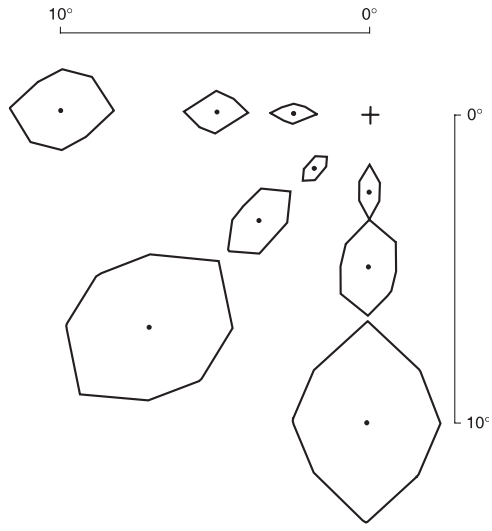

Figure 1: Fixating at the '+', the size and shape of windows of integration as they change with eccentricity are indicated by the roughly oval figures. What is not shown in this diagram is that the windows are overlapping. Reprinted with permission from Pelli and Tillman (2008): Nature Neuroscience.

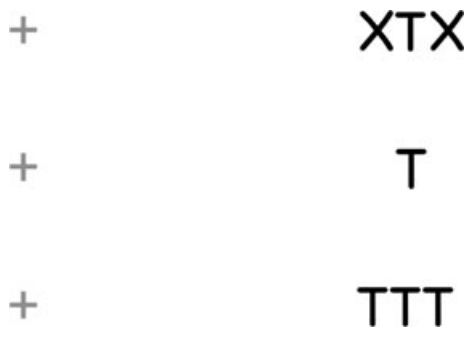

Figure 2: If you fixate on the '+' on the left, you can see the 'T' in the middle row perfectly well since you can attentionally select it. But it is very hard to make out the 'T' in the top row. The issue between me and Taylor and Richards is whether one can consciously see the middle 'T' in the bottom row. This figure is a variant of one in (Block 2013) but with crowded 'T's instead of crowded 'A's.

one sees the crowded object. Taylor says my argument appeals to the information the subject has but this is a misleading summary of my argument, since what my argument appeals to is not simply the information but also that it is consciously and visually appreciated. To avoid the constant listing of these four abilities I will just speak of the identity-crowding abilities, where that term is meant to specify that the abilities are conscious. Although I did not sufficiently emphasize this in the article, the conscious identification involves binding properties to the crowded object.

The form of both Taylor's and Richards' critiques is to criticize my reasons for thinking that we consciously see the identity-crowded items and to give an alternative account of the identity-crowding abilities. Both Taylor and Richards grant that there is no object-based attention to the crowded items and so I will not discuss that issue further here. They dispute whether we consciously see the items we cannot attend to, focusing on the argument from the identity-crowding abilities to conscious seeing. Both argue that the conscious phenomenology that I take as the phenomenology of seeing is importantly cognitive rather than perceptual phenomenology. Taylor argues for an account of the identity-crowding abilities in terms of inference from identifying the flankers and from visual appreciation of clutter, congruity and uniformity to the identity of the middle item. As we will see, there might be some truth in Taylor's inferential hypothesis but that truth is not incompatible with what I am claiming.

Richards' argument is more obscure: he argues that the identity-crowding abilities are fundamentally due to unconscious perception, and that unconscious perception somehow surfaces in the conscious phenomenal state. He doesn't say how this surfacing is supposed to take place, nor does he mention any other case in which we have unprompted conscious judgments that are a result of surfacing of unconscious perception. (I say "unprompted" because blindsight and various forms of unconscious priming can surface in prompted judgments where, for example, the subject is asked to guess whether he saw an 'X' or an 'O'.) He says "I offer a better explanation of identity-crowding reports and experiences: unconscious perception indirectly contributes to the overall phenomenal state of the subject through nonperceptual phenomenology associated with judgments." And he mentions that the nonperceptual phenomenology might include mental imagery. I guess the idea is that you unconsciously see the middle 'T' in Figure 2 and as a direct result you judge that there is a 'T' in the middle and maybe form a mental image of a 'T' in the middle.

In contrast to Taylor and Richards, I think the explanation of the identity-crowding abilities is sufficiently perceptual to involve consciously seeing identity-crowded objects. I gave an argument against the inference view in the paper in terms of asymmetries of identification (Petrov and Popple 2007). For example, subjects were 72% correct in identifying the sequence '\/\' in crowded vision but only 53% correct in identifying its mirror image, '/\/'. I said that such asymmetries can only be explained perceptually. The explanation is not known but Petrov and Popple hypothesize that it derives from the fact that optic flow that results from locomotion is almost always directed outwards, that is from the nose toward the ears, because one tends to move forward rather than backward. Still, Taylor is right that these asymmetries show only that the abilities I appealed to must be at least partly perceptual. To be clear: I do not deny that part of the explanation of subjects' abilities are as Taylor and Richards claim. Rather my point is that the explanation at least in some cases of the subjects' abilities involves conscious perception of identity-crowded objects.

More detail about how features are bound to objects in identity-crowding will be of use. Consider pictures below from "Metamers of the ventral stream" (Freeman and Simoncelli 2011).

Figure 3: When fixating on the crosshairs, the pictures in each pair are indistinguishable from the original and from each other. The pictures are chosen to illustrate the role of crowding in camouflage; the "Where's Waldo" character (red-striped shirt) and the soldier in Afghanistan can be seen only when fixated. You won't find the bottom pair very useful if you are reading this in black and white rather than color. Reprinted with permission from Freeman and Simoncelli (2011): Nature Neuroscience. See also Figure 3 of Rosenholtz, Huang, and Ehinger (2012).

Metamers are pairs of items that are physically different but look the same. You will notice that in each of the pictures in Figure 3 there are "crosshairs", two pairs of collinear line segments poking out of the photos. Follow those lines to where they cross in the picture. If you fixate at the intersection of these lines, you will not be able to tell the difference between the two pictures in each pair and the picture from which they were constructed. These pictures are distorted by a variant of a texture analysis and synthesis algorithm due to Portilla and Simoncelli (2000). In the pictures of Figure 3, the algorithm starts with a "normal" picture that is then "texturized" in the periphery in a way that looks weird if you fixate on it but has been shown to look the same in the periphery as the original picture (Freeman and Simoncelli 2011).

Figure 4: The top row contains the original pictures and the bottom row contains the versions that have been texturized by the algorithm (Portilla and Simoncelli 2000). Each top item looks roughly the same as the one below it in peripheral vision. Reprinted with permission from Portilla and Simoncelli (2000): Springer.

What is ''texturizing''? I will give a brief description but see Appendix A in (Balas, Nakano, and Rosenholtz 2009) for a real explanation. It will be helpful to understand texturizing by looking at some different examples. The pictures of Figure 3 texturize relative to a fixation point indicated by the crosshairs, but one can also texturize a *whole image* assuming the image is entirely inside one window of integration (i.e., with a fixation point understood to be far enough away so that the whole picture is in the window). Examples are presented below in Figure 4. The crosshairs are not shown but they can be assumed to be far to one side. The top pictures are the originals and the bottom ones are the texturized versions, ''mongrels'' in Ruth Rosenholtz's terminology (Balas et al. 2009; Rosenholtz, Huang, and Ehinger 2012). Each top picture should look very roughly the same in the periphery as the mongrel below it. I say ''very roughly'' because vision in the periphery makes use of multiple overlapping windows of integration, whereas the figure assumes a single window of integration, and because the algorithm used has not been tweaked for to produce metamers as have the versions used in Figures 3 and 6.

The algorithms were designed to take a digitized picture and random noise as input and transform them into another picture that is judged by human observers to have the same texture as the original. The algorithms operate in accord with constraints that preserve the relevant local statistics. Constraints are motivated by known properties of the visual system but also by trial and error. When applications of the algorithm yielded a picture with a distinguishable texture from the original, the authors tweaked the algorithm to eliminate distinguishable results. It turned out that variants of these

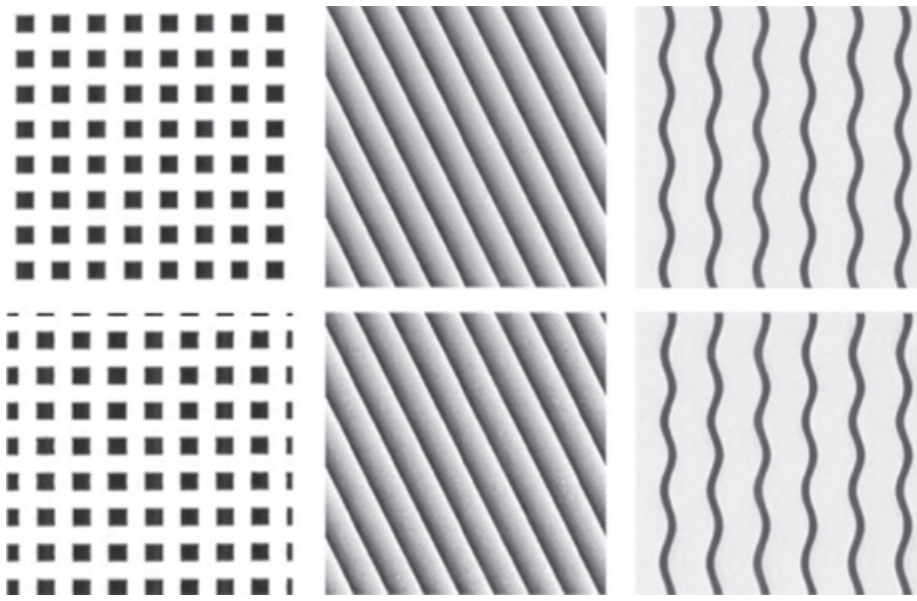

Figure 5: As with Figure 4, the top row contains the original pictures and the bottom row contains the versions that have been texturized by the algorithm. Each top and bottom pair will look the same in peripheral vision modulo a shift in phase (i.e., what part of the repeating pattern is at the edges). Reprinted with permission from Portilla and Simoncelli (2000): Springer.

algorithms reproduce the properties of crowding in peripheral vision (Balas, Nakano, and Rosenholtz 2009; Freeman and Simoncelli 2011).

It will be instructive to consider what the algorithms do to certain repeating patterns. The result is depicted in Figure 5.

I hope you are surprised by Figure 5 since the texturized patterns look pretty much the same as the originals on which they are based—in marked contrast to Figure 4 where the texturized pictures are distorted versions of the original. Why is there no distortion for certain regular patterns? The texture analysis and synthesis algorithms were originally designed to capture textures by analyzing and synthesizing local statistics. For some regular patterns, the local statistics just yield the original pattern! One way to think about the matter (Rosenholtz, Huang, and Ehinger 2012) is that the texturization process is like the compression schemes used to reduce the size of a music file or a photograph (e.g., an MP3 or JPEG) so that it will be useable by an iPhone but still sound or look nearly as good as the original. These compression schemes are "lossy" but not so lossy as to ruin the percept of the photo or music at least for a nonexpert consumer. In the case of a sufficiently simple and regular repeating pattern, however, the compressed version is the same as the original!

Ruth Rosenholtz kindly agreed to apply her version of the Portilla-Simoncelli algorithm to a display of objects more like the ones used in crowding. Rosenholtz's algorithm (Balas, Nakano, and Rosenholtz 2009; Rosenholtz, Huang, and Ehinger 2012) starts by blurring the original image so as to simulate the decrease in acuity in peripheral

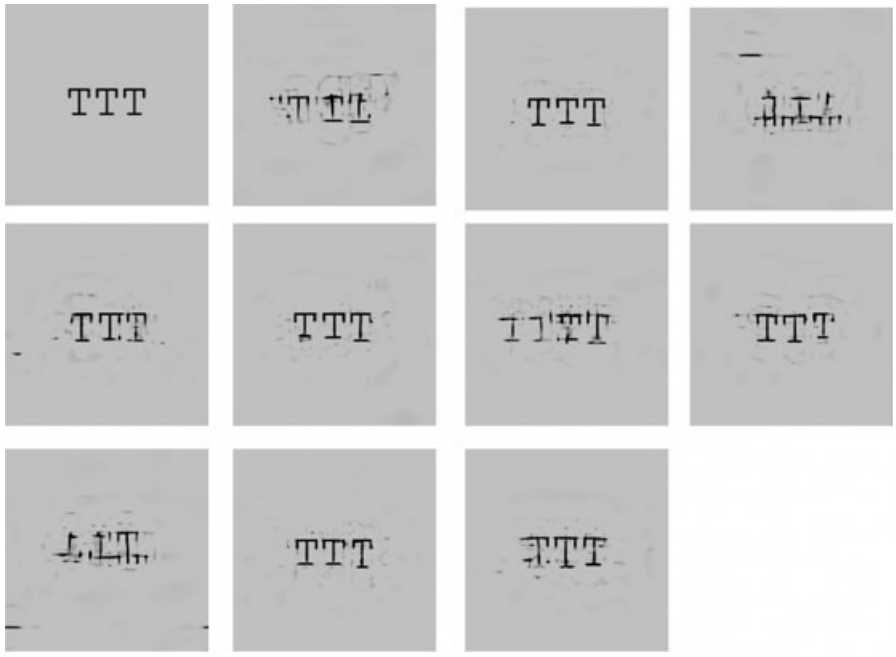

Figure 6: The top left triple of 'T's has been distorted by different applications of the mongrelization algorithm using different random "seeds". Thanks to Ruth Rosenholtz.

vision. The blurred image plus a random seed will produce a different "mongrel" each time. These mongrels look about the same as one another in peripheral vision. What is the significance of the fact that with different random seeds you get different mongrels? Rosenholtz's thought is that "The visual system may well serve up a number of "images," which are samples from a process with the measured statistics; akin to our mongrels. This may be the explanation for the shifting and dynamic percept many observers experience when attending to their peripheral vision" (Balas et al. 2009, p. 5). This may just be making a virtue of necessity since the algorithm requires a random seed and so inevitably different applications will look different—but only in focal vision. The prediction is that the different outputs will look roughly the same in peripheral vision, the dynamic percept resulting from small differences. The result is in Figure 6 in which the triple of 'T's in the upper left corner has been subjected to ten different texturizations/mongrelizations.

I would say that six of the ten mongrels are recognizable as triples of upright 'T's and another two as triples of 'T's in which some are inverted. The middle 'T' is recognizable in 7 or 8 of them as an upright 'T'. These figures suggest that the visual system can sometimes bind features to some objects in peripheral vision so long as the objects are all the same. It is this ability to bind—part and parcel of identification—that was scanted in my original article.

One interesting consequence of the mongrelization algorithm is that it predicts certain kinds of errors. Some of the 'T's in the mongrels in Figure 6 and one of the 'A's in the

Figure 7: Subjects have a hard time distinguishing between (a) and (b) if fixating on the '+'. From (Balas, Nakano, and Rosenholtz 2009). Note however, that (b) is more peripherally presented than (a). Reprinted with permission from Association for Research in Vision and Ophthalmology.

mongrel of Figure 7 are inverted and Rosenholtz and her colleagues have confirmed in pilot data (not a full experiment) that subjects have a hard time distinguishing (a) and (b) in Figure 7, though the reader should note that (b) is presented more peripherally than (a). The inversion effect may be slightly larger for 'A's than for 'T's and of course will be absent for 'I' which no doubt explains why the crowding case of individual line segments in the original article is the most convincing case of object-seeing without object-attention. These results may also explain the tendency subjects have in the famous Sperling experiment to confuse inverted letters in uncued rows with upright letters (Kouider et al. 2010).

The points made so far fall short of showing that one actually sees the middle 'T' as opposed to a merely multiply-T-ish texture. Every percept is constituted by a "perceptual attributive"—that represents an attribute—and a singular element—that represents an individual (Burge 2010). However, I have not yet given any argument that our perception of the middle 'T' involves such a singular element. Crowding is a phenomenon of object-perception: there is little or no crowding for textures (Ikeda, Watanabe, and Cavanagh 2013) but whether that fact counts for or against my point depends on whether identity-crowding is a genuine case of crowding. However, I did present an argument against an unorganized "bag of features" model in the original article (Block 2013, p. 180). I cannot reproduce that argument here, but to summarize, in a cued change-detection paradigm, crowded letters could be locationally cued, and performance was just as good with crowded letters as with uncrowded letters. Features must be bound at least to a location in order to be locationally cued and the fact that performance did not depend on whether the letters are crowded or not suggests the processes are the same, and the uncrowded case is almost certainly object-perception. To summarize, there is reason to believe that the singular element required for seeing obtains in identity-crowding but the issue deserves further consideration.

Let me return to the critiques. Taylor alleges that our identity-crowding abilities (to consciously detect, differentiate, discriminate and identify identity-crowded items) are due to not to perception of the identity-crowded items but to inference. The mongrels suggest that though there is room for inference to play some role, there is substantial perceptual binding of properties to objects in some cases of identity-crowding. It should be said that Taylor presents no actual evidence that identification of the middle 'T' is due

to inference. However, the main conceptual point is that a limited role for inference is compatible with our seeing the identity-crowded items.

Richards alleges that subjects' identification abilities are the result of unconscious perception surfacing in judgments. He says:

> "In identity-crowding cases, unlike ordinary crowding cases, there is enough consistency in the unconscious information to permit correct judgments, and these judgments have an associated phenomenal character. I will assume that the character is itself sensory (i.e., in this case, mental images or the like), but I do not exclude the possibility of non-sensory phenomenology associated with propositional attitudes. . . . .the best explanation is sensory phenomenology of textures and an accompanying non-sensory phenomenology associated with judgments that are more successful when the stimuli are uniform."

Of course unconscious information processing plays a role in all perception. But the mongrelization algorithms are designed to reproduce the conscious qualities of peripheral vision. For example, the metamers discussed above in connection with Figure 3 are cases in which a distorted image *looks the same* as the original so long as one fixates at the crosshairs. The algorithms designed by Simoncelli and Portilla were constructed on the basis of the experimenters' judgments about which results appeared the same in texture: "The only true test of texture synthesis system is whether the results appear (to a human observer) to be "the same" as the original" (Simoncelli and Portilla 1998). So there is reason to think that the mongrelization process is telling us about conscious vision.

As I mentioned, both Taylor and Richards criticize my reasons for thinking that we consciously see the identity-crowded items and try to give an alternative account of the identity-crowding abilities. Richards' critique of my account is that I am wrong to suppose we consciously see the identity-crowded items because perception in the periphery is mainly unconscious. Richards' reasoning is mistaken because whether or not there is more unconscious than conscious information processing in the periphery is irrelevant to the issue at hand, which is whether there is *enough* conscious information processing to ground seeing the objects in the periphery. The fact that one consciously sees something is not impugned by the finding that one also unconsciously sees it. This fact is perhaps obscured by pragmatic principles. Calling unconscious visual processing "unconsciously seeing something" suggests that there is no conscious seeing, but unconscious visual processing can and does underpin conscious seeing.

Richards says "The key point here is that given the nature of peripheral processing in general and the phenomenology of the experience there is a bias in favor of the unconscious processing explanation of the successful identification in identity-crowding." His claim about the nature of peripheral processing in general is based on the "two visual systems" work of Goodale and Milner: There is a conscious "ventral" stream that is mainly concerned with attaining an accurate conscious representation of the world that can be used to plan action and an unconscious "dorsal" stream that is in charge of moment-to-moment guiding of motor movement.

Richards says: "Further, the ventral stream associated with conscious experience primarily makes use of foveal information and ignores low-resolution information from the periphery (Milner and Goodale 2008, p. 783). What information processing there is in the periphery is typically dorsal stream and unconscious." However, what Milner and Goodale actually say on p. 783 is not that the ventral stream "ignores" peripheral information, but that "The ventral stream exploits the high resolution and wavelength selectivity that characterize processing in the fovea, and is much less interested in the low-resolution information from the periphery" (Milner and Goodale 2008). There is a difference between "ignores" and "much less interested". You can see for yourself that conscious vision does not "ignore" the periphery by looking at the pictures in Figure 2. When you fixate at the intersection of one of the pairs of lines you have a conscious percept of the periphery, though a "mongrelized" one.

There have been myths about peripheral vision in philosophy that exaggerate the difference between foveal and peripheral vision. Robert Van Gulick, citing Daniel Dennett (1991) and the alleged "fact that the retina lacks cones at the periphery" says "We have the firm belief that our phenomenal experience of the entire visual field is colored; that is how it seems to us. But if we hold fixation to the front and hold a marker whose color we do not know at arm's length to the side, we cannot discern its color. If it is gradually moved toward the front, we cannot see its color until it is far toward the center of our field of vision" (Van Gulick 2007, p. 529). Dennett often says this sort of thing, for example: " . . . most people—"naive subjects" in the standard jargon—suppose that their color vision extends all the way to the periphery of their visual fields" (D. Dennett 2005, p. 41). However, what determines color perception in the periphery of the visual field most directly is the cortical areas that are the neural basis of conscious perception. In the case of color, let us suppose that the relevant area is V4. V4 is approximately "retinotopic" (Henriksson et al. 2012), that is it is organized much the way the retina is organized. So rather than appealing to a supposed lack of peripheral color receptors on the retina, what we should look at is the neurons in V4 that represent the periphery. The visual system has many sophisticated mechanisms for "filling in", for integrating information from successive fixations, and for "memory color" that can supply color information to experience that does not exist on the retina at any one moment. Compare the experience of fixating on the crosshairs in one of the top pictures of Figure 2 with one of the bottom ones. If you are reading this article in color, you will have no difficulty in perceiving the peripheral color at the bottom as compared with the top.

Further it is just a myth that the retina lacks color receptors in the periphery. Hue discrimination at 50° eccentricity is *as good as in the fovea* if the size of the stimulus is increased enough. And there is even some color sensitivity out to 80°–90° (Mullen 1992).

To sum up, Taylor and Richards are right to press on my argument for seeing identity-crowded objects on the basis of our ability to consciously detect, differentiate, discriminate and identify identity-crowded items. What I said in my paper (Block 2013) made a less than convincing case. The points made above, however, fill some of the gaps by showing that the pooling processes in peripheral vision treat repeating items of the sort involved in identity-crowding differently from nonrepeating items. This is no doubt

the source of our ability to consciously see the identity-crowded objects despite lack of object-based attention to them.

References

Balas, B., L. Nakano, and R. Rosenholtz. "A Summary Statistic Representation in Peripheral Vision Explains Visual Crowding." *Journal of Vision* **9**.12 (2009): 1–18.

Block, N. "The Grains of Vision and Attention." *Thought* **1** (2013): 170–84.

Burge, T. *Origins of Objectivity*. Oxford: Oxford University Press, 2010.

Dennett, D. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge MA: MIT Press, 2005.

Dennett, D. C. *Consciousness Explained*. Boston: Little Brown, 1991.

Freeman, J. and E. Simoncelli. "Metamers of the Visual Stream." *Nature Neuroscience* **14**.9 (2011): 1195–201.

Henriksson, L., J. Karvonen, N. Salminen-Vaparanta, H. Railo, and S. Vanni. "Retinotopic Maps, Spatial Tuning, and Locations of Human Visual Areas in Surface Coordinates Characterized with Multifocal and Blocked fMRI Designs." *PLoS ONE* **7**.5 (2012).

Ikeda, H., K. Watanabe, and P. Cavanagh. "Crowding of Biological Motion Stimuli." *Journal of Vision* **13**.4 (2013): 1–6.

Kouider, S., V. de Gardelle, J. Sackur, and E. Dupoux. "How Rich is Consciousness? The Partial Awareness Hypothesis." *Trends in Cognitive Sciences* **14** (2010): 301–7.

Lettvin, J. Y. "On Seeing Sidelong." *The Sciences* **16**.4 (1976): 10–20.

Milner, A. D. and M. A. Goodale. "Two Visual Systems Re-Viewed." *Neuropsychologia* **46** (2008): 774–85.

Mullen, K. T. "Colour Vision as a Post-Receptoral Specialization of the Central Visual Field." *Vision Research* **31**.1 (1992): 119–30.

Pelli, D. and K. Tillman. "The Uncrowded Window of Object Recognition." *Nature Neuroscience* **11**.10 (2008): 1129–35.

Petrov, Y. and A. V. Popple. "Crowding is Directed to the Fovea and Preserves Only Feature Contrast." *Journal of Vision* **7**.2 (2007): 1–9.

Portilla, J. and E. Simoncelli. "A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients." *International Journal of Computer Vision* **40**.1 (2000): 49–71.

Rosenholtz, R., J. Huang, and K. Ehinger. "Rethinking the Role of Top-down Attention in Vision: Effects Attributable to a Lossy Representation in Peripheral Vision." *Frontiers in Psychology* **3**.1 (2012): 1–15.

Simoncelli, E. and J. Portilla. "Texture Characterization via Joint Statistics of Wavelet Coefficient Magnitudes," in *Proceedings of Fifth International Conference on Image Processing, I*. 1998, 1–5.

Van Gulick, R. "What If Phenomenal Consciousness Admits of Degrees?" *Behavioral and Brain Sciences* **30**.5/6 (2007): 528–9.