

---

## Breaking the Visual Stimulus Into Parts

Norma Graham

---

To perceive is to know "what is where." To perceive visually is to obtain this knowledge through the eyes. At a global level, visual perception may be thought of as a two-fold process: The visual system first

**Norma Graham** is a Professor of Psychology at Columbia University and has research interests in psychophysical, physiological, and computational studies of vision. Address correspondence to Norma Graham, Department of Psychology, Columbia University, New York, NY 10027.

breaks the information that is contained in the visual stimulus into parts; then the visual system puts the information back together again. But why take it apart in the first place? Because the proximal stimulus—the light falling on the retina—bears little direct resemblance to the important aspects of the world that must be perceived, that is, to the distal stimulus. The lack of resemblance between the proximal and distal stimuli makes the task of visual perception inherently difficult. Presumably, the information in the proximal stimulus is analyzed into parts to

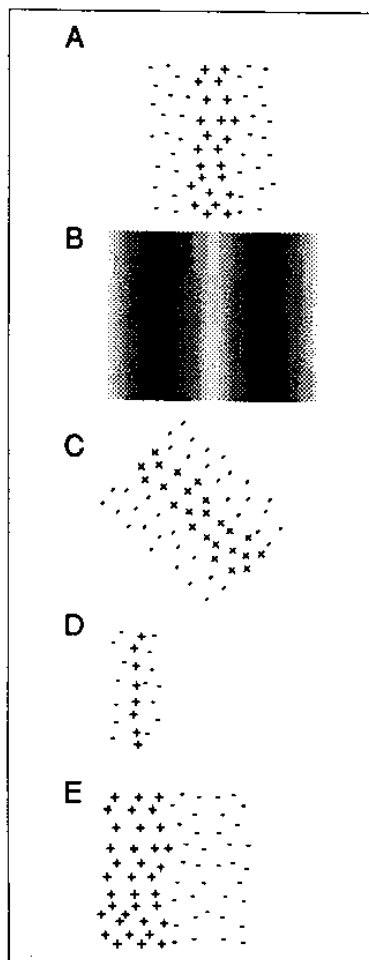
make economical and feasible the reassembly of the information into a representation of the distal stimulus.

This review deals with the elementary parts into which visual information is initially analyzed, in particular, with the elementary parts that are relevant for seeing patterns in space and time. (Color and three-dimensionality are not discussed, but much is known about them as well.) Three or four decades ago, we knew very little about the elementary parts of visual patterns; we still know very little about the processes putting the information from the parts back together again.

Two events were crucial in initiating the successful approach to discovering visual patterns' elementary parts. One was the discovery of the specialized receptive fields of the neurons in the primary visual cortex (area V1), along with the discovery that the parameters of these receptive fields varied in orderly ways from neuron to neuron.<sup>1</sup> The receptive field of a visual neuron is the area on the retina where light patterns can drive its firing (by way of neural pathways that connect the cortical neuron to that small region of the retina). Within the receptive field of a neuron in the visual cortex, there are excitatory and inhibitory subareas. These subareas form adjacent elongated stripes, as shown in Figure 1a. Light falling on an excitatory subarea increases the neurons' firing; light falling in an inhibitory area decreases it.

The second crucial event was the application of Fourier, or linear systems, analysis to the problem of quantitatively defining pattern sensitivity both neurophysiologically and psychophysically.<sup>2</sup> This application introduced the use of sinusoidal gratings ("fuzzy stripes"—see example in Fig. 1b) as elementary stimuli because, according to the theorems underlying Fourier analysis, any visual pattern is equivalent to the sum of many sinusoidal grating stimuli.

These two events led to the sug-



**Fig. 1.** Receptive fields and sinusoidal grating stimuli. (a) Diagram of the receptive field of a neuron in the primary visual cortex. The diagram plots the region of the retina where light spots either excite (+) or inhibit (-) the firing of the cortical neuron. (b) A patch of sinusoidal grating. (The patches used in experiments have gradual edges rather than the abrupt edges shown here.) If this sinusoidal patch is shown so it is centered on the retina on the receptive field in (a), the corresponding neuron will be strongly excited because the white stripes will stimulate the excitatory subregion (yielding a good deal of excitation) and the black stripes will stimulate the inhibitory subregions (yielding very little inhibition). (c)-(e). Diagrams illustrating ways in which the receptive fields of cortical neurons may differ. Differences occur in orientation, width of subregions, and symmetry—odd symmetry in (e) versus even symmetry in (c), (d), and (a).

gestion that different neurons—or, in the more physiologically neutral language preferred by psychophysicists, different analyzers (sometimes called detectors, channels, units, or pathways)—were selectively sensitive to different aspects of spatial and temporal patterns. The dimensions along which the selectivity occurs are the dimensions that arise naturally in discussions of Fourier analysis—for example, the orientations of the stripes, their spatial frequencies (widths), and their temporal frequencies (how rapidly a pattern oscillates or drifts with time). Thus, the fuzzy stripes that form the elementary stimuli in Fourier analysis were just what was required to define the selective sensitivity to local light patterns that is conferred on neurons in the primary visual cortex by the pattern of excitatory and inhibitory subregions in their receptive fields. Using these patterns, psychophysical and electrophysiological work has fleshed out the hypothesis that a fundamental stage of visual pattern processing is a stage of multiple analyzers, acting in parallel. This stage breaks the proximal visual stimulus down into parts because each analyzer responds only to a limited range of values along several dimensions, such as spatial frequency and orientation, as I describe in further detail in the following section. (Although the parts responded to by these low-level analyzers are like sinusoidal patches and do serve as elementary units of visual processing, they clearly do not represent elementary units in the finished percept. Visual scenes do not appear to be composed of many small striped patches!)

These original events inspired an enormous amount of research, both neurophysiological and psychophysical.<sup>3</sup> The reports of these studies were scattered over the pages of many different journals, and many of the researchers themselves had little idea of how coherent the whole story was becoming. In fact, when in

1980 I started writing a survey<sup>4</sup> of this vast and dispersed literature, I planned to describe briefly the overall framework of multiple-analyzer models and then go on to point out at length all the problems with the overall picture (the discrepancies between different sets of results, the holes in existing models, etc.). Instead, I was surprised to discover that almost all of the difficulties and inconsistencies were in psychophysical studies done with high-contrast patterns (suprathreshold patterns). If one considered only the psychophysical studies using patterns of near-threshold contrast (contrast so low that the pattern is imperfectly discriminable from a steady blank field of the same space-average luminance), there were still hundreds of studies employing a wide variety of psychophysical tasks, and the conclusions from different investigators using different methods agreed remarkably well. (A reason for this difference between near-threshold and suprathreshold results is considered briefly in the following section.)

#### THE MULTIPLE ANALYZER MODEL

The receptive fields of analyzers could differ in the size of their subregions and, therefore, in preferred spatial frequency. (The largest size of subregion shown in Fig. 1 and, therefore, the lowest preferred spatial frequency is in Fig. 1e; the smallest size and, therefore, the highest preferred spatial frequency is in Fig. 1d.) The receptive fields could differ as well in orientation (vertical in Fig. 1a, d, and e; oblique in Fig. 1c). In general, they could differ in a number of other properties as well, for example, in symmetry (e.g., even-symmetric fields, as in Fig. 1a, vs. odd-symmetric fields, as in Fig. 1e), spatial position (the location in the visual field of the receptive field's center), and spatial extent (how long

each subregion is relative to its width; how many subregions there are). Similarly, although harder to diagram, the temporal responses of neurons, or analyzers, could differ in many properties.

Table 1 lists the dimensions necessary to specify the spatiotemporal receptive field of an analyzer. Alternatively, these dimensions can be described as dimensions necessary to characterize a particular visual stimulus—a sinusoidal patch like that in Figure 1b (which may, however, be sinusoidal in space, time, or both).

Adequately explaining why these dimensions might be basic dimensions for pattern vision is beyond the scope here. Briefly, however, as mentioned earlier, any visual stimulus is equivalent to the sum of many sinusoidal-patch stimuli. Thus, if we knew how a visual system responded to sinusoidal patches, we might hope to be able to compute its response to any stimulus at all. This would be strictly possible if the system were what is known as a linear system, that is, if the response to a sum of two stimuli were exactly the sum of the responses to each stimulus alone. We can hope that this is approximately true for the visual system we actually have, or, at least, for some of its subparts. As it turns out, the multiple analyzers are usually assumed to be linear, but strict linearity is not crucial for the conclusions here.

To connect the outputs of these multiple analyzers to the observer's responses in psychophysical experiments, a very simple decision rule is typically used. The kind of rule obviously depends on the kind of experiment because the form of the observer's responses varies from one type of experiment to the next. For detection-threshold experiments (in which the task of the observer is simply to say whether a nonblank pattern of near-threshold contrast has been presented or not), the most common rule is probably the *maxi-*

*mum-output rule*, namely, the observer says, "Yes, a pattern is present," if and only if the largest of all the analyzers' outputs is greater than some criterion.

In the past 20 years, it has become increasingly likely that the physiological substrate for the multiple analyzers is cortical area V1 (and perhaps V2). V1 is the primary visual receiving area of the cortex; that is, it is the first place in the cortex that receives information about the visual stimulus, as shown on the left in Figure 2. Over the past 20 years, it has also become increasingly clear that V1 and V2 are only two of the many different areas in the cerebral cortex concerned with vision, and some of the other areas are also shown in Figure 2. Recent estimates suggest that 25% to 40% of human cerebral cortex is concerned with processing the information from visual stimuli, but only about 10% to 20% of that volume is thought to be V1.

The simple decision rule in the psychophysical model is thus a vastly oversimplified representation of 80% or 90% of visual cortex, and a vastly oversimplified representation of all the many visual cortical areas higher than V1 or V2! Nonetheless, a model consisting of multiple analyzers coupled with this kind of simple decision rule is sufficient to explain at a quantitative level the results of near-threshold psychophysical experiments. In response to near-threshold patterns, only a very small minority of the analyzers send above-baseline responses upstream. Perhaps this sparseness of information limits the kinds of processing that higher levels can perform to kinds describable by simple decision rules. It is as if the simplicity of the near-threshold experimental situation has made all the higher levels of visual processing transparent, allowing the properties of the low-level multiple analyzers to shine through.

Simple decision rules might well be inadequate to account for results

**Table 1.** *Dimensions of pattern vision*

Dimension	Multiple analyzers? <sup>a</sup>	Question		
		Labeled outputs?	Probabilistic independence?	Mutual inhibition?
Spatial frequency	Yes, with bandwidth of 0.5–1.5 octaves	Yes	Yes	Perhaps; weak near threshold
Orientation	Yes, with bandwidth of 15–60 deg of rotation	Yes	Yes	Perhaps; weak near threshold
Spatial position (x and y)	Yes; along one dimension the receptive field has an excitatory center and inhibitory flanks and (maybe) excitatory outer flanks; along the other dimension it has only an excitatory center; full receptive field is approximately circular	Yes	Yes	Perhaps; weak near threshold
Spatial extent (width, length)	Probably not near threshold	—	—	—
Spatial phase (symmetry)	Unclear	—	—	—
Temporal frequency	No narrowly tuned analyzers, but 2 or 3 very broadly tuned analyzers	Yes	Unknown	Unknown
Temporal position	Yes, with impulse response that is purely excitatory at high spatial frequencies and biphasic at low spatial frequencies	Yes	Yes	Unknown
Temporal extent (duration)	Unknown	—	—	—
Temporal phase (symmetry)	Unclear	—	—	—
Direction of motion	Yes at high velocities, no at low velocities	Yes	Probably yes	Perhaps (little evidence)
Eye of origin	Yes; may differ in degree of binocularity as well	No	Unclear	Unclear

<sup>a</sup> Bandwidths are given as full width at half peak height. The bandwidths, although typical, generally depend on values on other dimensions, which are noted only when this dependence is dramatic. See Graham<sup>4</sup> (especially chap. 12) for more information.

of almost any suprathreshold experiment, however, because in response to a suprathreshold pattern, a very large number of analyzers will send above-baseline responses upstream, thus opening up possibilities for many different kinds of higher level processing. It is perhaps no wonder that initial attempts to explain the results of suprathreshold experiments, using models that emphasized the multiple-analyzer stage without trying to build sophisticated later stages, were unsatisfactory.

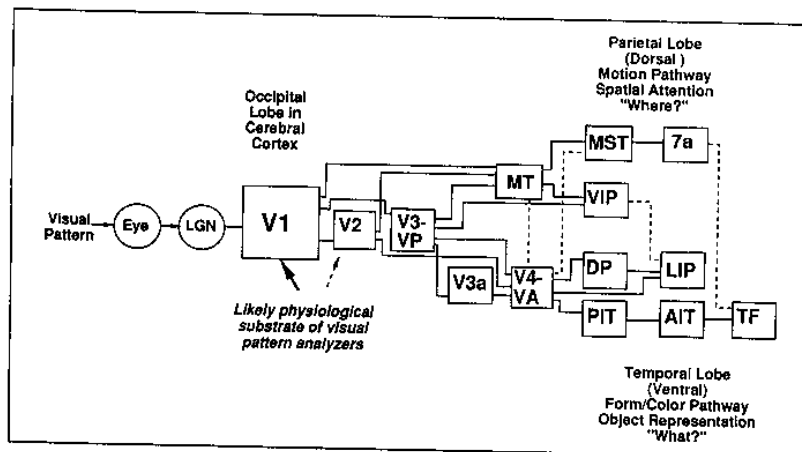
The results of near-threshold studies are summarized in Table 1 in terms of four questions, each of which can be answered by near-threshold experiments. This section briefly describes these questions and

their answers. The logic by which one can answer these questions about multiple analyzers on the basis of psychophysical experiments (some of which is referred to as multidimensional signal detection theory) is itself an area in which a good deal of progress has been made over the past two decades.<sup>4,5</sup>

#### Are There Multiple Analyzers on a Given Dimension?

Experimental results in which two stimuli having values close together on some dimension (e.g., two grating patches of very similar orientations) interact more than stimuli having far-apart values (e.g., two grating patches of perpendicular orientations) are taken as evidence for mul-

multiple analyzers on that dimension. The kind of interaction demonstrated depends on the kind of experiment in question. In an adaptation experiment, for example, an observer might adapt to a vertical grating by looking at it for a period of some minutes while moving his or her eyes in order to prevent conventional afterimages. Then the observer would be tested with gratings of a number of different orientations. Typically, the detection thresholds for test gratings similar in orientation to the adapting grating (e.g., somewhat tilted off vertical) would be elevated after adaptation, whereas the detection thresholds for test gratings very different in orientation (e.g., horizontal) would not be affected.



**Fig. 2.** A simplified diagram of the visual cortical connections in the macaque monkey with labeling to reflect the current idea that there are at least two major streams of information processing. The dashed lines represent interconnections between the two streams. The lines without arrows indicate connections that are reciprocal. The letters in the boxes label the visual areas according to the several conventions currently in use. Two labels (7a and TF) are derived from old architectural maps of the cortex. Another set of labels (V1, V2, V3, V3a, V4) is based on beliefs about the order of processing within the cortex. The remaining labels are based on geographical location of the area (in the species in which originally discovered, which may be rather different from the macaque): DP = dorsal prelunate, MT = middle temporal, MST = medial superior temporal, VP = ventral posterior, VA = ventral anterior, PIT = posterior inferior temporal, and AIT = anterior inferior temporal. LGN = lateral geniculate nucleus. The visual system of the human is thought to be similar to that of the macaque. The presumed physiological substrate of the multiple analyzers in the psychophysical model is cortical area V1. Adapted from Movshon.<sup>7</sup>

This value-selective behavior is explained by assuming (a) that analyzers sensitive to different orientations exist and (b) that those analyzers that are sensitive to the adapting orientation were fatigued or inhibited in some manner by the adaptation period.

In summation experiments, the degree of interaction between two values is the degree to which the detectability of a compound stimulus containing both values (e.g., a superposition of two orientations) exceeds the detectability of each component. For example, a compound pattern composed of two gratings of very similar orientations (which probably both stimulate much the same analyzers) is much more detectable than a compound composed of gratings of perpendicular orientations (which probably stimulate different analyzers).

Even this latter compound con-

taining perpendicular orientations is somewhat more detectable than its components, however, an effect that is usually explained as *probability summation*. To understand probability summation, it may help to think about tossing coins, with each coin representing a set of analyzers and getting a head on the coin representing the analyzers' detection of the stimulus. For the compound stimulus containing two far-apart orientations, two sets of analyzers each have a chance to detect the compound (those sensitive to one component and those sensitive to the other), which is like tossing two coins in order to get at least one head (a probability of .75 if the coin is fair). But for a stimulus containing only one of the orientations, only one set of analyzers has a chance to detect it, which is like tossing one coin in order to get a head (a probability of .50). Thus, if there is prob-

ability summation among analyzers, a compound should be somewhat more likely to be detected than either of its components even though the components are not stimulating the same analyzers. Probability summation is an example of behavior considered by multidimensional signal detection theory, behavior that needs careful consideration before psychophysical experiments can be interpreted properly.

In identification experiments, the degree of interaction between two values is the degree to which stimuli of those two values are confusable (e.g., a slightly off-vertical line being confused with a vertical line), as measured in any of several experimental paradigms. The confusion presumably occurs because both values stimulate the same analyzers. Here, I consider only the identification of patterns that are themselves near threshold, that is, imperfectly discriminable from a blank field.

In uncertainty experiments, an observer's performance is measured both when the observer is uncertain about which stimulus will be presented and when he or she is certain. If, for example, two stimuli are very far apart in orientation and an observer does not know which of the two will be presented on a given trial, the observer cannot detect either stimulus as well as he or she can when certain as to the orientation. This decrement due to uncertainty is presumed to occur because the observer has to monitor two different sets of analyzers; in contrast, when the observer is certain which stimulus will be presented, he or she can monitor just one set. If, however, the two stimuli are quite similar in orientation, ignorance of which will be presented does not deleteriously affect performance, presumably because the observer is always monitoring just the one set of analyzers sensitive to both of the orientations. (Uncertainty effects of the magnitude found in these near-threshold experiments do not appear to be due

to limitations in attention capacity, however, but rather due to independent noisiness in the analyzers, another example of the behavior considered by multidimensional signal detection theory.) In uncertainty experiments, in short, interaction between values occurs when uncertainty does *not* cause a decrement in performance.

In the column titled "Multiple Analyzers?" Table 1 indicates the dimensions on which value-selective behavior (more interaction for close-together values than for far-apart values) has been found in the above kinds of experiments. This column also gives the approximate bandwidth of the analyzers—that is, the range of values over which a single analyzer is sensitive. Two notes of caution: Bandwidth estimates may well be slightly revised in the future because the complete interpretation of any of these kinds of experiments is still in some doubt (e.g., is value-selective adaptation due to fatigue of the analyzers or to inhibition among analyzers?). Also, to answer unambiguously the question of multiple analyzers along any one dimension, definitional issues involving two or more dimensions are sometimes involved.<sup>6</sup>

One conclusion about the existence of multiple analyzers along different dimensions is worth special mention. Although spatial frequency and temporal frequency are in many ways formally identical, the value selectivity found on the spatial frequency dimension is much more pronounced than that on the temporal frequency dimension; in other words, each analyzer must respond to a much narrower range of values on the spatial frequency dimension than on the temporal frequency dimension.

#### **Are the Outputs of the Multiple Analyzers Labeled?**

To have a *labeled* output means that the higher stages of processing keep track of which output comes

from which analyzer. For example, if the analyzers' outputs are labeled, the observer might be assumed to identify which of several stimuli was presented on a trial by finding out which analyzer had the biggest output on that trial. Although value-specific behavior in near-threshold summation and adaptation experiments can be explained without assuming that the multiple analyzers' outputs are labeled, explanation of value-specific behavior in uncertainty and identification experiments seems to require labeled outputs (although the labeling may be fuzzy or otherwise imperfect).

Notice that on all but one of the dimensions in Table 1, where multiple analyzers exist, those multiple analyzers are labeled. The one exception is the eye-of-origin dimension. For this dimension, near-threshold adaptation and summation experiments show value-selective behavior, but near-threshold identification and uncertainty experiments do not. The interpretation of these results is as follows: Although some analyzers respond better to one eye than the other, the relevant higher stages of visual processing do not keep good track of which analyzer responded better to which eye. Thus, these higher stages cannot identify the eye of origin of a monocular stimulus, nor can they monitor only those analyzers more sensitive to a particular eye (as they would have to in order to cause an uncertainty effect on the eye dimension). This evidence of lack of labeled outputs in near-threshold identification and uncertainty tasks may seem particularly surprising because in stereopsis computations, some stages of visual processing must be keeping track, to some extent at least, of which analyzer responds best to which eye.

#### **Are the Multiple Analyzers Probabilistically Independent?**

Are the outputs of different analyzers variable? If so, is the variability

in the output of different analyzers probabilistically independent, so that the random variability in the output of analyzers sensitive to one range of values does not correlate with the random variability in analyzers sensitive to a different range of values? Or, in other words, are the noise sources in different analyzers independent? Probabilistic independence among the outputs of multiple analyzers can show its effects in near-threshold summation experiments (as probability summation for compounds containing far-apart values), in uncertainty experiments, and in some forms of identification experiments. For all dimensions with multiple analyzers, experimental results suggest the analyzers are probabilistically independent (although the interpretation is complicated on the temporal frequency dimension by the breadth of tuning of individual analyzers and on the eye dimension by the same factor and also by lack of labeling).

#### **Is There Mutual Inhibition Among the Analyzers?**

When stimuli that are far apart on the dimension of interest produce an effect that is opposite to the effect found when stimuli are close together, mutual inhibition among analyzers is a possible interpretation. This kind of evidence for inhibition has occasionally been reported in adaptation, summation, and identification experiments. Inhibition is also a possible interpretation if adapting to a compound stimulus produces less effect than adapting to one of the components by itself. For both kinds of evidence, however, reasonable alternative explanations not requiring inhibition have also been put forth. Consequently, the psychophysical evidence for mutual inhibition from near-threshold experiments is not compelling. Yet such inhibition is found in cortical physiology, and appears to be useful in the explanation of suprathreshold

results. Thus, "perhaps" is listed in Table 1 for mutual inhibition to indicate dimensions for which evidence consistent with inhibition has been observed in near-threshold psychophysical experiments.

### CONCLUSION

The many hundreds of near-threshold psychophysical studies of pattern vision published in the past three decades form an impressive and compelling body of evidence for a model in which a fundamental process is the breaking down of the visual stimulus by a set of multiple analyzers, acting in parallel, with different ranges of sensitivity along different dimensions. Considered together, these studies and the neurophysiological studies of the same period suggest that the physiological substrate of the multiple analyzers is area V1 (the lowest level of cortical visual processing) and perhaps area V2.

As is consistent with this presumed physiological substrate, these multiple analyzers are apparently at a relatively low level in the full stream of visual processing (although coming after a number of other processes, e.g., light adaptation). It seems clear that much complicated computation intervenes between the analyzers' outputs and observers' perceptions.

In the past three decades, we have learned much about how our visual systems analyze the proximal visual stimulus into parts. A major challenge for the future is to find out how the parts that result from this analysis are "put back together" into a perception that generally corresponds very well to the distal stimulus—the arrangement of objects that the perceiver must know about and interact with in order to survive, the "what is where." In trying to take this step forward, we can build on the precise quantitative knowledge

about the multiple analyzers that we have gained over the past three decades from both physiology and psychophysics, particularly from near-threshold psychophysics.

### Notes

1. D.H. Hubel and T.N. Wiesel, Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex, *Journal of Physiology*, 160, 106–123 (1962); D.H. Hubel and T.N. Wiesel, Functional architecture of macaque monkey visual cortex, *Proceedings of the Royal Society (London), Series B*, 198, 1–59 (1977).
2. F.W. Campbell and J.C. Robson, Application of Fourier analysis to the visibility of gratings, *Journal of Physiology*, 197, 551–566 (1968).
3. Early, seminal psychophysical studies include C. Blakemore and F.W. Campbell, On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images, *Journal of Physiology*, 203, 237–260 (1969); A. Pantle and R. Sekuler, Velocity-sensitive elements in human vision: Initial psychophysical evidence, *Vision Research*, 8, 445–450 (1968); M.B. Sachs, J. Nachmias, and J.C. Robson, Spatial-

frequency channels in human vision, *Journal of the Optical Society of America*, 61, 1176–1186 (1971); J.P. Thomas, Model of the function of receptive fields in human vision, *Psychological Review*, 77, 121–134 (1970).

4. N. Graham, *Visual Pattern Analyzers* (Oxford University Press, New York, 1989).

5. See, e.g., N. Graham, P. Kramer, and D. Yager, Signal-detection models for multidimensional stimuli: Probability distributions and combination rules, *Journal of Mathematical Psychology*, 31, 366–409 (1987); S.A. Klein, Double-judgment psychophysics: Problems and solutions, *Journal of the Optical Society of America A*, 2, 1560–1585 (1985); N.A. Macmillan and C.D. Creelman, *Detection Theory: A User's Guide* (Cambridge University Press, New York, 1990); J. Nachmias, On the psychometric function for contrast detection, *Vision Research*, 21, 215–223 (1981); D. Pelli, Uncertainty explains many aspects of visual contrast detection and discrimination, *Journal of the Optical Society of America A*, 2, 1508–1531 (1985); J.P. Thomas, Detection and identification: How are they related? *Journal of the Optical Society of America A*, 2, 1457–1467 (1985).

6. See, e.g., Graham, note 4.

7. A. Movshon, Visual processing of moving images, in *Images and Understanding*, H. Barlow, C. Blakemore, and M. Weston-Smith, Eds. (Cambridge University Press, Cambridge, England, 1990).