Title

Prentice Award Lecture: A Simple Retinal Mechanism That Has Complex and Profound Effects on Perception

Permalink

Journal

Optometry and Vision Science, 62(7)

ISSN

1040-5488

Author

CORNSWEET, TOM N

Publication Date

1985-07-01

DOI

10.1097/00006324-198507000-00001

Copyright Information

0093-7002/85/6207-0427$02.00/0
AMERICAN JOURNAL OF OPTOMETRY & PHYSIOLOGICAL OPTICS

Vol. 62, No. 7, pp. 427–438
Printed in U.S.A.

awards/honors

Prentice Award Lecture: A Simple Retinal Mechanism That Has Complex and Profound Effects on Perception

TOM N. CORNSWEET*
School of Social Sciences, University of California, Irvine, California

In the November 1984 issue of *Scientific American*, Lederman published an article called "The Value of Fundamental Science"; his first sentence says a lot of what I want to say. "One takes up fundamental science out of a sense of pure excitement, out of joy at enhancing human culture, out of awe at the heritage handed down by generations of masters and out of a need to publish first and become famous."[1]

The work I want to tell you about is exciting to me, and now, with this honor, I feel famous.

It is unusual for a professional academy like this one, whose primary goal is and should be the advancement of a profession, to give its highest honor to people doing basic research. That says something special about this academy and its members. I feel highly honored by this award, and I am also very grateful. I suppose all of us like to think that the work we do is important, but not many of us are lucky enough to have somebody else say that they think what we do is worth doing. That is very flattering and gratifying. Thank you.

I would like to tell you about a theory of a very simple retinal mechanism that seems to explain many apparently complicated and unrelated visual phenomena. I think it is interesting because, even if the theory turns out to be wrong in the sense that perhaps there is no such mechanism actually in the retina, it does what all good theories do in that it reveals that what we previously thought were unrelated phenomena can all be manifestations of the same mechanism; and, especially if some of the phenomena seemed complicated and the mechanism is simple, then the world is easier to understand and our thinking is simplified. To me, that step in science, where a new theory seems to take a lot of messy pieces and melt them down to a single clear drop, is the essence of what is esthetically pleasing about science. (The next step, of course, is that your successors, or you if you're lucky, look deeply into the drop and see a whole new set of messy pieces.)

Most of the visual phenomena that I will talk about are probably familiar to you, but I will refresh your memory about them as we go along. They are light and dark adaptation, Weber's law, Ricco's law, brightness constancy, receptive field center-surround antagonism, Mach bands, and changes in acuity with luminance. Some of these phenomena are necessarily related to each other, in that if you have one, you almost have to have the other, and some are not necessarily related, but all of them can be explained by one simple retinal mechanism. And I would like to introduce the mechanism by talking about still another phenomenon that it explains, one that may actually be the fundamental reason why the mechanism evolved, and so may be the reason why all of the other phenomena occur. This fundamental phenomenon doesn't have a name. It is the fact that people are a whole lot better at detecting objects in bright light than you would expect on the basis of their vision in dim light, and vice versa, that is, people do a whole lot better in dim light than you would expect them to do on the basis of how well they do in bright light. Let me explain that.

A basic property of objects in the world is that they reflect fixed proportions of the light falling on them. Suppose your eyes were pointed at an

---

The Prentice Award Lecture, presented at the Annual Meeting of the American Academy of Optometry in St. Louis, Missouri, December, 1984.

Received November 29, 1984.

* Ph.D., Member of Faculty.

object that reflects 50% of the incident light on a background that reflects 40%. In daylight, you would say that you are looking at a gray object on a darker gray background, and you wouldn't have any trouble seeing it—that is, detecting that the object was there rather than not there. Now think about what happens when you look at the same scene under dim illumination. Remember that light acts as though it consists of individual packets of energy, quanta, and that emission and absorption of each individual quantum is random, so light falling on the object is like rain, and so is the light reflected from the object and its background that enters our pupils and rains on the retina to form the retinal image.

Fig. 1 shows the computed light distribution in the retinal image at some particular instant when you are looking at two gray squares on a dark gray background, under very dim light. There is a square centered in the bottom half of the figure, just barely noticeable, that reflects twice as much light as the background and so would look light gray against the dark background in daylight. There is also one centered in the top half of the figure, which is essentially undetectable, that has a reflectance 10% greater than the background, so in daylight it would also be easily seen and would be very noticeably lighter than the background.

If you were to examine the retinal image of this scene microscopically with a magical light detector and watch just one spot, you would see that, from time to time, at unpredictable times, a quantum will arrive there and be absorbed, and quanta will arrive at neighboring spots at other random times. Fig. 1 is a picture of that, but frozen in time. Looking at the retinal image this way, how can you tell whether or not the object at the top is there? Well, the object reflects about 10% more of the quanta incident on it than the background does. So, if the object were really there and you were to count quanta over some time in the region of the image of the object and also in a region the same size on the background, you would find that the total number of quanta was about 10% bigger for the object than for its background. But because the arrival and absorption of quanta are random, the difference would not always be exactly 10%. Sometimes, by chance, it would be larger and sometimes smaller. In fact, sometimes the number of quanta absorbed in the region of the object would actually be smaller than for the background, even though, on the average, the object reflects more light. So even if your visual system were a perfect detector and counter of quanta, and you said that the object was there whenever you counted more quanta in the location of the object than in the background, you would still be wrong sometimes, just because of the fundamental statistical nature of light.

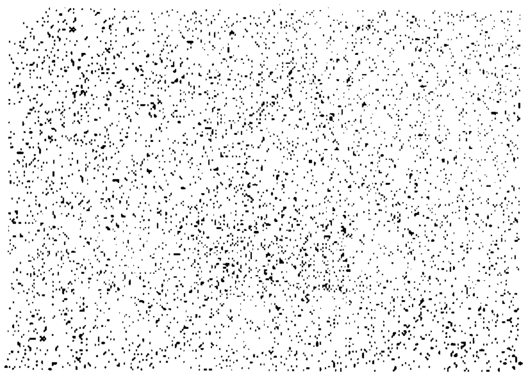

FIG. 1. A representation of the retinal image, at one instant in time, when the scene being viewed consists of a square area centered in the top half that has a reflectance 10% greater than the background and another square area centered in the lower half that has a reflectance twice that of the background. Each black spot represents the absorption of one quantum, and their occurrences follow the appropriate Poisson distributions.

Now, because of the particular kind of randomness that governs quanta, the likelihood of being wrong in the situation I just described gets bigger when the average number of quanta counted gets smaller. So, for instance, the gray object will be detectable essentially 100% of the time under daylight conditions where the average number of quanta absorbed by the retina is very large. However, under dim illumination even a perfect quantum counter will often be wrong, and we would say that we cannot detect the object reliably, in that we cannot be sure whether it is there or not.

There are two ways that we could change things so that the object would be more reliably detectable under dim light, and both involve counting larger numbers of quanta, which is equivalent in some ways to increasing the illumination on the object. One thing we could do would be to count quanta over a longer time, and the other would be to count over a larger area. Those are two of the things we do when we want to take a photograph under dim illumination, where exactly the same statistical problems show up. (Opening up the iris diaphragm improves things by actually increasing the number of quanta per second arriving at the film. That is exactly equivalent to increasing the scene illumination, so it is not relevant here.) Increasing the exposure duration enough for the film to catch enough quanta will let us take an adequate picture, but at the cost of blurring objects that are moving. The other alternative is to use faster film. What that really does is to use film that has bigger light-sensitive grains, and so we are really increasing the area over which quanta are collected. The cost there is spatial resolution, that is, fast film gives pictures that look grainy and lose fine detail, whereas film that shows fine detail has very small grains, which don't catch enough quanta to make a useful picture in dim light.

There is a great deal of evidence in the vision literature showing that in dim light, the human visual system probably both increases the time during which it adds up the effects of quanta and also increases the area over which it adds them.[2] We say that both temporal and spatial summation increase under low light levels. As a consequence, we can reliably detect objects in dim light, but our spatial and temporal resolution are worse than in bright light.

The temporal effects are obviously very important, but I am not going to talk about them here. I will just talk about changes in spatial summation. Picture what is happening this way. The retina consists of an array of densely packed receptors. Suppose that each of them could signal the amount of light it receives, independently of all the others. Then, when an image falls on the retina, a neural representation of the image will emerge from the receptors, and the fineness of the neural image will equal the fineness of the receptor packing. That system would work well enough so long as the intensities of light in the retinal image are great enough that each receptor can catch a large number of quanta quickly enough. But if the light level gets so low that each receptor catches a quantum only once in a while, then the system will have great difficulty in reliably detecting objects, for the statistical reasons I talked about before. What the retina evidently does under those conditions is to change the way it acts so that instead of each receptor independently signaling its quantum catch, the outputs of large groups of receptors are all added together, that is, summated, so that each group, in effect, acts like one large receptor that catches more quanta per second. Then the system can reliably detect objects under much dimmer light levels.

A retina organized like that would not be able to respond to fine details, because its receptors are, in effect, too big. But at low light levels, if the receptors were small enough to resolve fine details, they would not be able to anyway because of the "noisy" statistical nature of light. So, in dim light the large summation areas help in detection of large objects and don't make things any worse with respect to seeing fine detail. However, if the light level goes up so that the resolution of fine detail becomes physically possible, then if the summation areas did not get smaller, they would prevent us from seeing as well as the physical nature of light would allow us to, and their large size is no longer needed for detecting large objects. So the ideal system is one in which, when the light level is low, large groups of receptors add their signals together, and as the light increases, the sizes of these groups get smaller, in correspondence with the statistical properties of light. There is much psychophysical evidence to support the idea that the human visual system does make an adjustment of this type so that the areas of spatial summation grow when the light level is low and shrink when it is high.[2]

What I have said so far is a very general description of a process, but it turns out that when we start thinking specifically and explicitly about the process, there are some surprises. So let me start by describing a simple theory or model of how this kind of change in spatial summation might happen, and then I will show you how a visual system with that mechanism would actually behave.

Fig. 2 is a diagram of the model. This is a slice through a simplified retina. The line of detectors or receptors represents, of course, a slice through a whole two-dimensional surface packed with receptors. An image, that is, some light distribution, falls on the receptor array. Each receptor sends a signal into a network, where it spreads out, and there is a corresponding array of output channels. Each output channel sends out a sig-

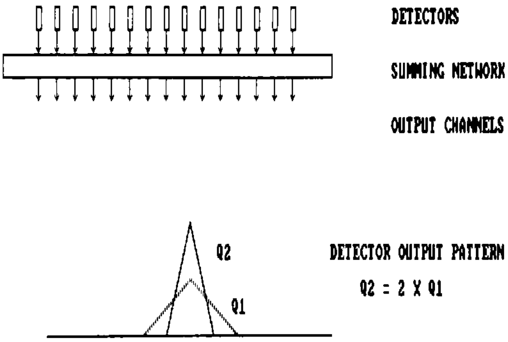

FIG. 2. A schematic representation of the IDS model. This is a slice through a two-dimensional surface. The input is a light distribution (the retinal image) falling on an array of detectors or receptors. Each receptor spreads its output in the summation network according to a spread function, such as the one shown at the bottom, that changes with the input intensity. For higher intensities (e.g., Q2), the spread function is higher in the center but covers a smaller area in such a way that its volume is constant. Then an array of output channels signals the sums of the spread functions in its location, as illustrated in Fig. 4.

nal that is just the sum of all the signals that arrive in its location in the summation network.

Now the special property of the model has to do with the way that the signal from each receptor spreads in the summation network. One possible shape for this spread is shown in the lower half of Fig. 2. Specifically, we assume that each receptor sends out a signal with the following properties. First, the strength of the signal at its center is proportional to the intensity of light falling on the receptor. Second, the signal spreads out over an area that is bigger when the intensity of illumination is weaker, but in such a way that the total volume under the spread is constant.

The two plots in the bottom of Fig. 2 are for a spread that is cone shaped. (Of course, it looks like a triangle in this one-dimensional slice through it.) When the intensity is higher, the peak of the cone is higher but its width is smaller so that the total volume under it stays the same. All that I am going to say about the consequences of this model is essentially the same regardless of the particular shape of this spread. It can be conical, or exponential, or Gaussian, or anything, so long as the height at the center is proportional to the intensity of the image there and it spreads farther when the intensity is lower so that its volume is constant. That is all there is to the model. It is one explicit form of a mechanism that would show larger areas of spatial summation at low light levels. It is helpful in talking about the model to give it a name, and I will call it the intensity-dependent spread, or IDS model.

Now let me talk about what the model does, that is, how vision would behave if this mechanism were part of it. And here I want to mention that I worked out most of the results I will talk about in collaboration with Jack Yellott, who is also at the University of California at Irvine, and Jack and I have written a fairly extensive mathematical treatment of most of what I will say here.[3] First, it is probably no surprise that the ability of the system to detect objects varies with light level in the way it should so that there is the right trade-off between spatial resolution and the averaging of quantal fluctuations. In other words, at low light levels, the detectability of large objects is good because there is much adding up of quanta, and at high light levels, where there are so many quanta that quantal statistics do not limit how well a perfect system could see, the summation areas shrink so that they don't blur the neural image.

Now let's look at some of the other things the IDS model does. They are more surprising. First, suppose the eye is pointed at a large uniformly lighted blank field, so that the retinal image has the same intensity everywhere. Then, obviously, the "neural image" or output pattern will simply be the same everywhere. Now suppose a single bright point is superimposed on the otherwise uniform field. The output pattern of the IDS system for that stimulus is plotted at the top in Fig. 3. The output where the bright point falls is higher than the background. That is no surprise. But the output right next to the bright point is actually lower than the background. Before the bright point was added, the output pattern was the same everywhere. It would just be plotted as a straight horizontal line in Fig. 3. Then, when the bright point is added, the output at the center goes up, but the output just off center actually gets smaller. So adding light to the retina makes some of its outputs get smaller. I will explain how that happens in a minute, but let me talk about something else first.

Another way of showing the output pattern for this stimulus is to indicate the neural activity in a view looking down at the retina, as shown at the bottom in Fig. 3. At the center, the activity increases, and so it is given a plus sign, and in a surrounding ring, the activity decreases, so it is given a minus sign. This kind of plot is the familiar receptive field plot first used by Kuffler, in 1953, to describe the activity patterns in cat

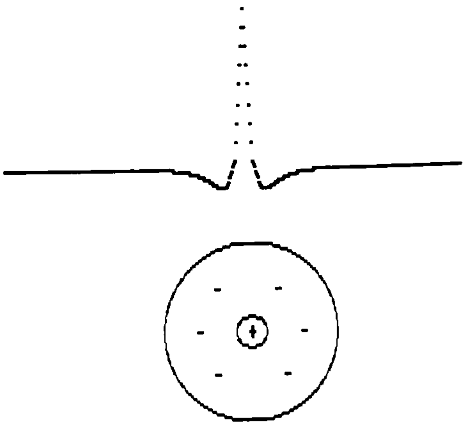

FIG. 3. The output pattern for the IDS model when the input is a bright point on a less bright background. The upper curve is a section through the two-dimensional array and the lower sketch is a two-dimensional view, usually called a receptive field map. The central region is given a plus sign because the activity in it increases when the bright point is added to an otherwise uniform field. In the surrounding ring, the activity decreases, so it is given a minus sign. Thus the IDS model manifests what is called center-surround antagonism, although no inhibition is involved. (The shape of the spread function in this and Figs. 5, 6, and 10 is taken to be exponential. The straight portions of the upper plot are linear interpolations between detectors, included only to improve the appearance of the plot.)

retinal ganglion cells.[4] When a receptive field is arranged like this one, it is said to show center-surround antagonism, in that the activity at the center seems to oppose the activity in the surround. This phenomenon, where adding light to some part of a retina causes the activity in other parts to be reduced, has been seen in every mammalian retina that has been studied and it is very natural to say that the activity at the bright spot must be inhibiting the activity around it, and so this phenomenon is said to demonstrate the action of laterally spreading inhibition. In fact, this kind of so-called center-surround antagonism is usually used as evidence to demonstrate and prove the existence of lateral inhibition in the retina. But the IDS model does exactly the same thing, and it has no inhibition or subtraction processes in it. It involves only summation—only the adding up of activity.

Fig. 4 is an attempt to explain what is happening in an IDS system when it is shown a bright spot on a less bright background. The input pattern is shown at the top. Each receptor spreads its signal, and the spreads are plotted below the receptors here as triangles. Notice that all of the spreads are the same except for the one receptor that is more brightly lighted, and its spread is higher at its center but narrower. Each output channel just adds up all of the activity that gets to it, and the result of that adding up for each output channel is plotted as a point below it. Look, first, at the output channel marked "A." It gets a signal from receptor 2, right above it, and also gets some signals from receptors 1 and 3 on each side of it and adds them all up. The output channel just to the right of channel A gets the same amount of signals, so its output is the same, the background level.

Now look at the channel lebeled "B." Before the bright point was added to the background, it would also have had the same input as all the others. But when the bright point was added, the spread from the receptor directly under the bright point, receptor 6, shrank so it no longer adds to the output of channel B, the total input to channel B is less, and so its output drops below the background level. So the output of channel B is reduced not because it is inhibited by the action of the bright spot, but because it is adding up less activity than it would have if the bright spot were not there. So the change in spread of summation in the IDS model does something that looks just like the effect of inhibition, but there is no negativity or subtraction going on. It is only a change in summation. (Of course, you can play a semantic game and say that the shrinking of the spread in the IDS model is a subtraction of a kind, but I don't think that is useful. If you do want to call that a subtraction, it is certainly a different kind of subtraction, involving very different mechanisms, than the subtraction implied by the standard view of inhibition.)

A number of other phenomena follow directly from the result I just discussed. Fig. 5 shows an old familiar one. The intensity profile is shown at the bottom and the output pattern for an IDS model is plotted above it. This is the famous

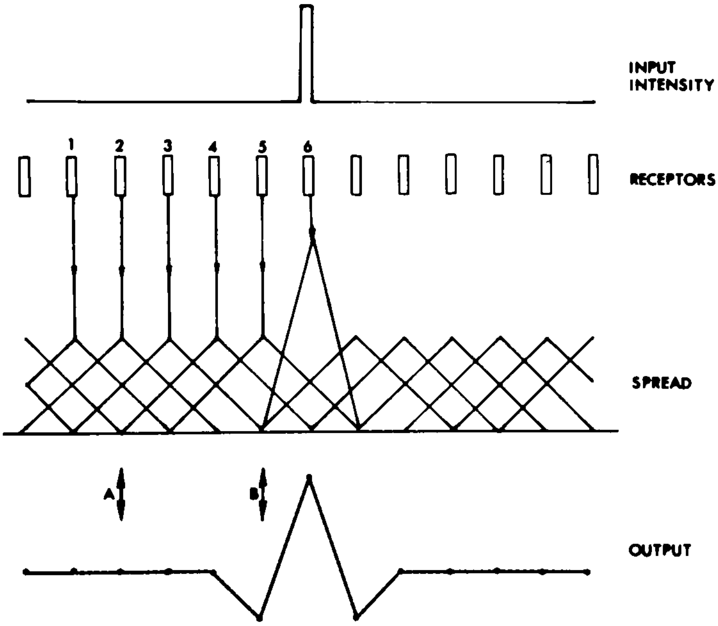

FIG. 4. A diagrammatic description of how the addition of a bright spot can reduce the activity of some output channels. See text for an explanation.

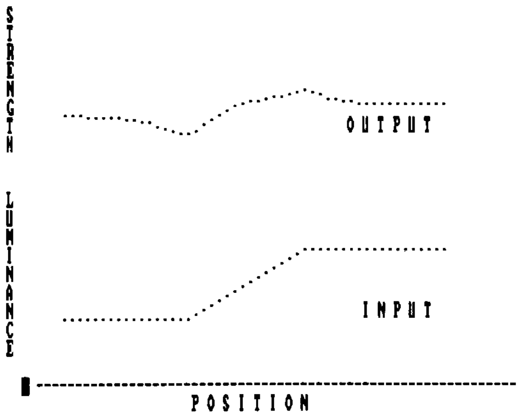

FIG. 5. The lower curve is a plot of luminance vs. position for the classical Mach band stimulus pattern. The upper curve is a plot of the corresponding output pattern for the IDS model.

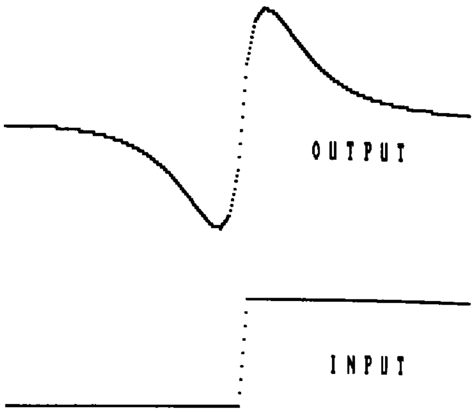

FIG. 6. The input pattern when the stimulus is a sharply focused edge is shown at the bottom, and the corresponding output pattern is plotted at the top. This kind of result is usually called edge enhancement.

Mach band stimulus pattern, and if the brightness that a person sees at each point in the pattern depends on the output of an IDS system, the pattern ought to show a dark region where the luminance first starts to increase and a less prominent bright one where it levels off again, which is exactly what happens in human vision. The light and dark regions are called Mach bands after Ernst Mach, who first described them in the 1860s.[5]

Fig. 6 shows the IDS output when the edge in Fig. 5 is steepened until it is vertical. In other words, it is the intensity profile for a sharp edge or step. The IDS model does what in the engineering literature is called edge enhancement. I will come back to this kind of picture later.

Fig. 7 is a plot of the contrast sensitivity function for the IDS model. It shows how the sensitivity of the system, plotted on the vertical axis, is related to the fineness of the stimulus pattern, plotted as spatial frequency on the horizontal axis. The little patches of grating below the horizontal axis indicate the relative fineness of the stimuli. When this function is measured psychophysically, the results show that the curve for the human visual system is just like this one.

Let me refresh you, briefly, on what this curve means. Look at the middle first. The sensitivity is high there. In other words, patterns that are of medium fineness can be easily seen, or can be seen when they have very low contrast. Now if you move toward the right, the patterns get finer and the sensitivity falls, until, at the right-hand end of the curve, the patterns are so fine that they cannot be resolved no matter how high their contrast is. That right end of the curve, then, is of particular practical interest because it represents the limit of visual acuity, and I will come back to that later.

Now go back to the middle of the curve and start moving toward the left. The patterns get coarser, changes in luminance become more gradual, and sensitivity falls there, too. It is hard to see patterns that are either too fine or too broad.

This kind of plot actually contains a lot of information, and because of that, it is being used more and more in a clinical setting, as a general description and evaluation of vision.[6] For instance, it is being applied to the diagnosis of optic nerve pathologies and to define the nature of the visual loss in amblyopia. When it is used like that, any losses that show up are usually given a physiological interpretation. For instance, the fact that in a normal eye, sensitivity falls going toward the left part of the curve, toward low spatial frequencies, has always been attributed to lateral inhibition. Therefore, when some pathology causes changes in that part of the curve, the pathology is usually interpreted as one that causes changes in the inhibition mechanism. The fact that an IDS system produces the same curve without inhibition suggests that some of those interpretations might need to be reexamined.

Center-surround antagonism, Mach bands, edge enhacement, and this kind of contrast sensitivity function can all be explained either by an IDS model or by a lateral inhibition model. In fact, if you compute the contrast sensitivity curve for an IDS model having some particular shape of spread function, and then you compute the corresponding curve for a lateral inhibitory model with the same shape of spread function, the two contrast sensitivity curves are identical.

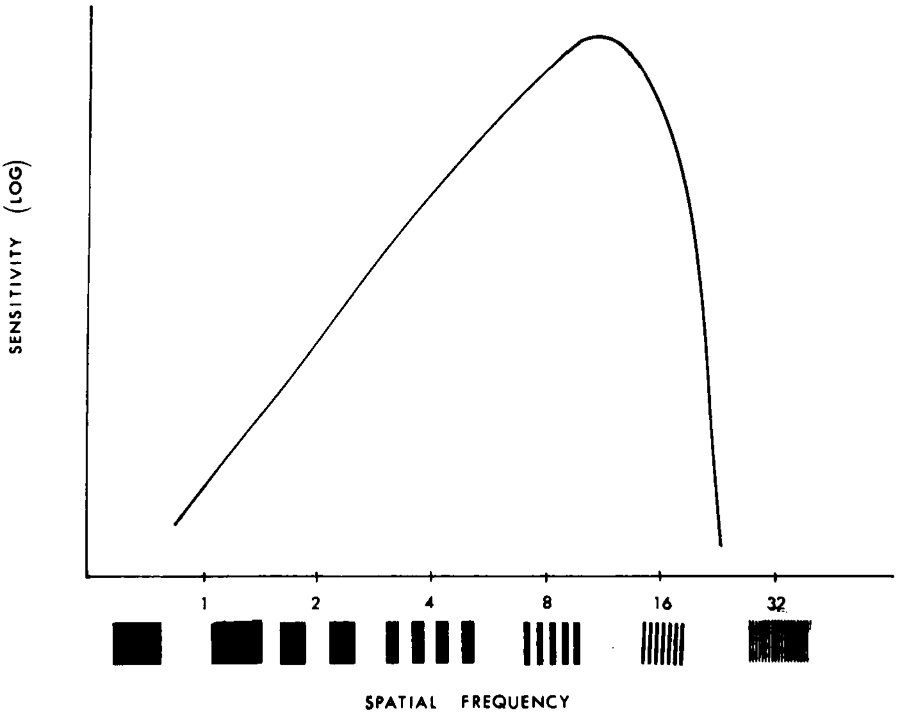

FIG. 7. The contrast sensitivity function for the IDS model. The pieces of grating along the horizontal axis are meant to indicate the approximate relative finenesses of the spatial frequencies plotted. When such a function is measured psychophysically, of course, the stimuli have sinusoidal luminance profiles. (The spread function in this and Figs. 8 and 9 is taken to be Gaussian.)

So any phenomenon that is represented by the contrast sensitivity function, and there are many, can be equally well explained by either the IDS or the lateral inhibition model.

Now I want to talk about a series of phenomena that are explained by the IDS model as I have described it, but which can be explained by the lateral inhibition model only if we add things to it or modify it significantly. One strong difference between a lateral inhibition model and the IDS model is that the inhibition model is linear and the IDS model is inherently nonlinear. One of the things that means is that if you double the intensity of a spot of light, the lateral inhibition model predicts that the output will be doubled and the IDS model predicts that it will be less than doubled. That nonlinearity has a lot of very important consequences, and I will talk about some of them now.

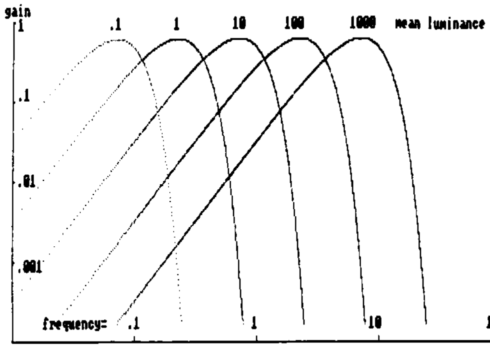

FIG. 8. The contrast sensitivity function for the IDS model shifts toward higher frequencies when the mean luminance of the test stimuli is increased. When the height of the peak of the spread function is assumed to increase linearly with illuminance, as for the computations underlying this figure, the contrast sensitivity function shifts in direct proportion to the square root of the mean luminance.

Fig. 8 is the contrast sensitivity function again, for the IDS model, but this time it is plotted for several different average illumination levels. In other words, suppose you measure sensitivity for several different spatial frequencies when the targets are illuminated at some level, and you plot the results. They would look like one of these curves, say the one at the left. If you then increase the illumination on the targets tenfold and repeat the measures, you would get the next curve over, and so on. For a visual system with intensity-dependent spatial summation, when the average illumination increases, the whole curve shifts toward higher spatial frequencies, and when the horizontal axis is plotted as the logarithm of frequency, the shape of the curve does not change at all, it just translates in a perfectly orderly way. In fact, it shifts in proportion to the square root of the intensity of the illumination. That shift is a consequence of the nonlinearity of the IDS model. A lateral inhibition model predicts that the curve will be absolutely unchanged at different illumination levels.

Van Ness and Bouman[7] and Kelly[8] showed that the curve for the human visual system shifts in the same direction as the IDS model in Fig. 8, but it doesn't behave in so orderly a way. Therefore, the IDS model, as I have described it so far, is not a complete explanation for the actual psychophysical results. However, the effects are clearly in the right direction, and when certain realistic constraints are added, like the fact that there is a finite number of receptors and they are of finite size, the results come even closer to those found for human vision. One of the important consequences of the shift in the contrast sensitivity function with average illuminance is that the right-hand end of the curve, where visual acuity is represented, shifts to higher frequencies as the illumination increases, so acuity increases with the intensity of illumination on the test chart. If the curve did not shift this way, there would be no need to control the illumination in a visual acuity test. But in fact, human visual acuity increases in proportion to the square root of the illumination level, exactly as the IDS model does.[9]

Now I will talk about a related but more esoteric phenomenon, Ricco's law. Ricco's law describes the detectability of small spots. It says that, so long as a spot is smaller than some critical size, it will be just detectable if it contains a certain number of quanta, regardless of how big the spot is. So if a small spot is just detectable and then it is made twice as big but its intensity per unit area is cut in half, it will again be just detectable. To put it more generally, so long as a spot is smaller than some critical size, its detectability depends only on the total number of quanta in its retinal image. Fig. 9 shows that the IDS model obeys Ricco's law in a way that is very similar to the way the human visual system does.

Fig. 9 is complicated, and it is not appropriate to explain it fully here, but generally it shows that Ricco's law holds up to some critical size (where the curves break away from the diagonal line), and furthermore, that the critical size is not constant but depends on the background luminance. (The caption to Fig. 9 contains a partial explanation.) The human visual system behaves in essentially this same way.[10]

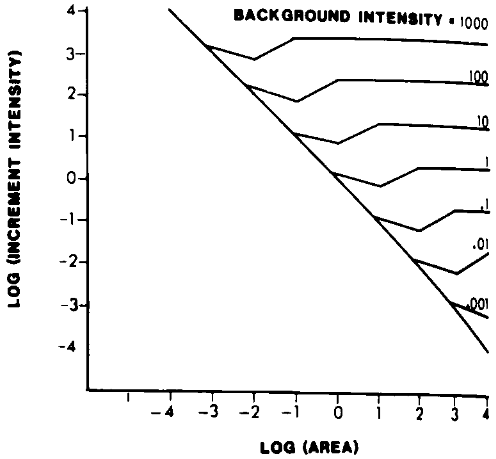

FIG. 9. The IDS model obeys Ricco's law. The visual system is presented with a spot of light of adjustable size and luminance on a background of adjustable luminance, and the luminance of the test spot is adjusted until it just produces an output of a particular amplitude. (In the corresponding psychophysical experiment, the luminance of the test spot is adjusted until it is at the subject's detection threshold.) Every point on these curves represents a test spot of some luminance and size that just produces this criterion response. For the lowest background luminance (0.001 arbitrary units), the intensity of the test spot (the increment) that is required for its detection is inversely proportional to the area of the test spot up to an area of about 3 log units. This means that, for all areas smaller than that, the product of the area and luminance is constant, that is, the total energy in the test spot at threshold is constant. That is Ricco's law. At higher background luminances, Ricco's law begins to fail for smaller test spots. The human visual system exhibits very similar behavior.

Fig. 10 shows what may be the most striking result of the IDS theory. Fig. 6 plotted the response of the model when the input is an edge. The IDS model shows Mach bands and edge enhancement. Fig. 10 plots the responses to a series of edges. Look first at the plot in the lower left corner. This is an intensity profile through a uniformly lighted disk or stripe of light with a luminance of 2 on a background of luminance 1. The plot above it is the corresponding output pattern for the IDS system. In the rest of the plots in the lower row, the luminances of the stripe and its background are increased, but the ratio of the luminances of the stripe and its background is the same, 2:1, for all the stripes. As the luminances rise, the differences in luminance across the edges of the stimulus stripes increase, and the responses, in the upper row, change in that they get narrower, but the amplitudes of the responses are all identical. In fact, we can prove that, for the IDS model, the amplitude depends only on the ratio of the in-

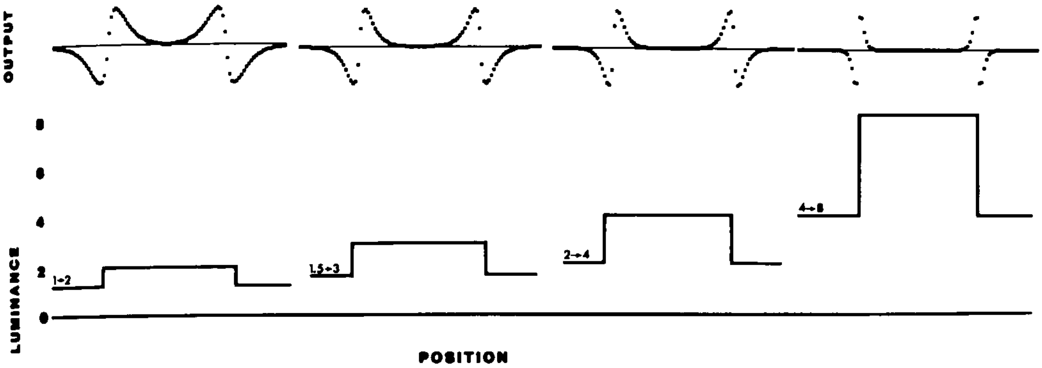

FIG. 10. Along the bottom row are a set of luminance profiles for spots or stripes, and the corresponding output patterns are plotted above them. The luminances of the various patterns are different, but the ratios of the luminances of all the stripes to their backgrounds is the same, 2:1. The output patterns get narrower for the stimuli of higher luminance, but the peak and trough amplitudes are all identical.

tensities on the two sides of an edge and not at all on the actual levels themselves.

Now what does that mean? It means that a visual system with IDS would obey that most venerable of psychophysical laws, Weber's law. Weber's law, when applied to luminance, says that the detectability of a test spot depends only on the ratio of the luminance of the spot to the luminance of the background, that is, Delta I over I is constant at threshold. For an IDS system, the only difference between a field without a test spot, that is, a uniform field, and a field with a test spot, is those peaks and troughs in the output that you see corresponding to the edges of the stripes in Fig. 10. So if the rest of the human system is to decide whether or not a test spot is present, and all it has to base its judgment on is those peaks and troughs, then it would say that a test spot was present if the peaks or troughs were just big enough to detect. And because the heights of the peaks and troughs depend only on the ratio of the two intensities, all these patterns would be equally detectable, so an IDS system would have to obey Weber's law.

The lateral inhibition model makes a very different prediction. It, or any other linear model, says that the detectability would have to depend on the difference between the luminances on the two sides of the edge, not on their ratio, so the stripes at the right in Fig. 10 would be easiest to detect and the ones at the left, the hardest. In other words, Weber's law would not hold. To fix that problem, people as far back as Fechner have suggested that the receptors may contain some mechanism such that their signal is not directly proportional to the light falling on them, but proportional to the logarithm of the light. A mechanism like that would give Weber's law, and if the lateral inhibition theory is modified by assuming that the receptors do a log transformation, then it also predicts that the heights of all the peaks and troughs in Fig. 10 would be equal. However, that still is not enough of a modification to the inhibition theory to explain several of the other phenomena that are correctly predicted by the IDS theory, like the shift in the contrast sensitivity curve shown in Fig. 8.

Weber's law is important in the science of vision, but a closely related phenomenon, brightness constancy, strongly affects what we see all the time. Think again about the retinal images represented in Fig. 10. Each curve represents an edge, the luminances change from one edge to another, and the ratios of the luminances on the two sides of the edges are all the same. What that set of input patterns represents, in terms of the real world, is this: Suppose you look at an object with some reflectance, say it reflects 50% of the light falling on it, and the object is on a background that reflects less, say half as much—25% of the incident light. (The stripe in Fig. 11 reflects about 50% of the incident light and its background about 25% so if you view Fig. 11 from far enough away that you can't see the fine stripes that it is made of, it is an example of such an object.) The retinal image of some region at the edge of the object will be a step between luminances with a ratio of 2:1. Now suppose you are looking at the object in sunlight and then the sun goes behind a cloud, and so the light incident on the object and its background drops by a factor of, say, 100. The intensities of light reflected to the eye will then also drop by 100-fold, and so the illuminances on the two sides of the edge will both drop by 100-fold, but, and this is the crucial thing, they will still be in the ratio of 2:1. So in normal seeing, when the illumination level changes, the illuminances

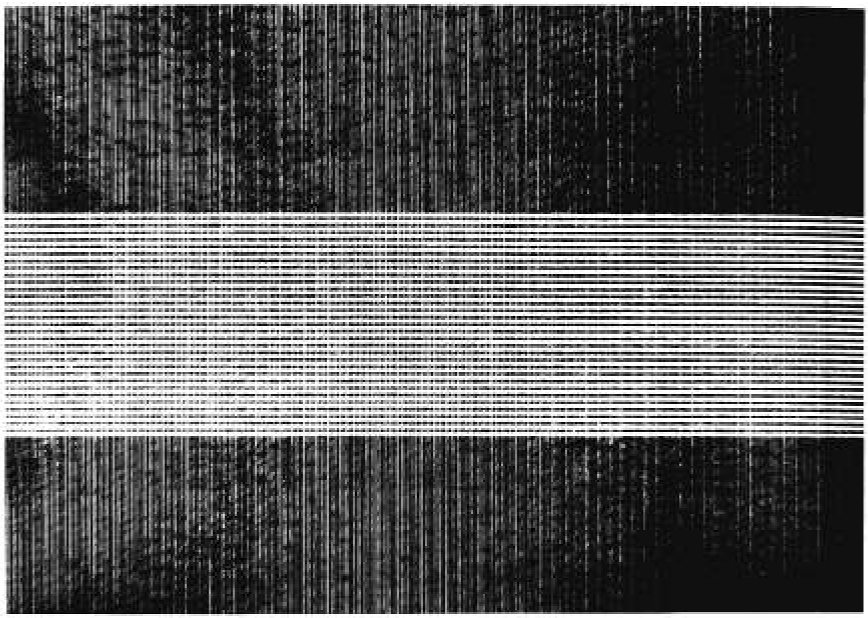

FIG. 11. The average reflectance of the center stripe is twice that of its background. If you view this figure from far enough away (or with a poor enough spectacle correction) that the fine lines are not visible, and then change its illumination, the intensity profiles of its retinal images will correspond to those plotted in Fig. 10.

on the retina also change, but their ratios don't change. The various input patterns in Fig. 10, then, represent an object with a reflectance twice as great as its background, under four different illumination levels.

If our visual systems actually responded to the intensities of the retinal images, we would see dramatic changes in the scene every time the lighting changed, but we don't see anything like that. When the sun goes behind a cloud, the objects in the world don't look much different, even though their retinal images have changed drastically. That perceptual phenomenon is called brightness constancy, and it powerfully affects what we see all the time. It is as if we ignore the intensities of retinal images and somehow see the reflectances of the objects directly, or as though we somehow disregard the illumination level.

Fig. 10 shows that an IDS visual system accomplishes exactly that. The amplitude of its response does not depend at all on the actual intensities in the image, but only on the ratios of intensities at edges, and those ratios don't change when the illumination changes. They are properties of the objects themselves. So if the rest of the visual system judges the brightnesses of objects on the basis of the amplitudes of the responses of an IDS retina, it will show perfect brightness constancy.

I want to emphasize, here, that all I have just said about Weber's law and brightness constancy applies only to large targets with sharply focused edges. If a target is very small, the IDS model predicts that it will not obey Weber's law. That is evident in something I have already talked about; the fact that acuity increases when the illumination increases is a violation of Weber's law. If Weber's law held for acuity targets, then they would be equally detectable regardless of illumination, because the ratio of the reflectance of the target and its background does not change when the illumination changes. So the IDS model predicts that Weber's law should break down for very small objects and the psychophysical results agree.

The IDS model also predicts that Weber's law should break down for targets that are defocused or have blurred rather than sharp edges, and I have not yet been able to find out whether anybody has done that experiment. Similarly, the IDS model predicts that brightness constancy should not hold for very small or for blurred objects, and I don't think anybody has tested that yet.

There is only one more set of phenomena that I want to talk about, namely light and dark adaptation. Those terms are used in two subtly different ways in visual science. Often, they are used to refer to the fact that visual sensitivity changes over time after the light level has changed. For instance, if we are suddenly placed in a very dimly lighted room, we can see almost nothing at first, and then we get more and more sensitive as time passes. The way that change in sensitivity happens over time is called dark adaptation. The other way that the term is used to refer not to the time course of these changes in sensitivity, but to the process that goes on in the visual system that causes the changes. The IDS theory describes a process that would cause the kinds of changes in sensitivity that are called light and dark adaptation, as I will explain in a minute. But the theory, so far, does not say anything about the way the process proceeds over time, so when I say "adaptation," what I will be referring to is the process that causes changes in sensitivity, and I will not say anything about how fast or slow the process might be.

The result of the adaptation process is that, when we are in dim light, we are more sensitive than in bright light. What that really means is this. If we are looking at a brightly lighted area and a spot of light is added to it, we can only detect the spot if its intensity is pretty high, whereas if we are looking at a dimly lighted area, we can detect a much dimmer added spot. Any visual system that obeys Weber's law will do exactly that, and so any theory that explains the process underlying Weber's law will also necessarily explain at least a major part of the processes of light and dark adaptation. In fact, Weber's law really defines just how much light and dark adaptation actually occur. And in that sense, the IDS model explains light and dark adaptation.

Let me put it another way. People working in this area, especially physiologists, often refer to adaptation as a process in which the gain of the visual system is changed. Changing the gain of a system is equivalent to multiplying or dividing the responses by some factor. One familiar example is the volume control on a hi-fi system. When you turn up the volume, you multiply all the loudnesses by some factor. Another example is the pupil in the eye. When it dilates, it has the effect of multiplying all the intensities in the retinal image by some factor, and that is a change in gain. In fact, the action of the iris in the eye is an example of a light- and dark-adaptation mechanism, although one that has a very limited range.

The IDS mechanism has an equivalent effect. If the light level illuminating a scene is doubled, all of the intensities in the retinal image are doubled. If the pupil were then to shrink to one-half of its former area, it would bring the retinal image intensities back to where they were before, and so the response of the retina would be exactly the same as if the scene illumination had not changed. Or if the pupil did not change, but some mechanism turned down the volume control of the neural circuits in the retina so that they were all one-half as sensitive, it would have the same result, and that is what physiologists suggest goes on in the retina during light adaptation. That notion has always bothered me because it seems very hard to think of a neural mechanism that does multiplication or division, and it is even harder to think of one that does it in a way that depends on the retinal illumination level, so I was surprised to realize that the IDS mechanism, even though it does not have anything in it that seems to be doing anything other than changing the spread of signals and adding them up, does the same thing as a process that would actually multiply or divide. In other words, one way to describe the effect of the IDS mechanism is to say that it changes the retinal gain in inverse proportion to the illuminance. It causes the kinds of changes we usually call light and dark adaptation.

There is one complication here that I have to mention. Dark adaptation is usually measured by putting the subject in total darkness and then measuring how intense a spot of light has to be, against this totally dark background, in order to be seen. That case, where the background has zero light in it, is undefined in the IDS model. That is, the model just does not make any prediction then. But there is an idea that has been around in vision for a long time, and expanded and studied by people like Barlow, that, if added to the IDS model, also lets it handle stimuli on a dark background.[11] That idea is called "dark light." Electrophysiological studies show that, when the retina is healthy, its ganglion cells are always active, even in total darkness. They may fire faster in the light, but they still fire even in darkness. As far as the brain is concerned, this activity in the dark is not qualitatively any different from the activity in light, and so it probably has the same effects as activity due to light, and it has been called by this weird name, dark light.

If we assume that this activity affects the retinal circuits the same way as the activity generated by light, then the IDS model can deal with targets presented on dark backgrounds, because, although the light in the background might be zero, the activity is not.

Barlow and some of his co-workers pointed out that this dark light would have to interfere with the detection of a target shown on a dark background, and suggested that maybe the reason that sensitivity increases over time in the dark is because the amount of dark light decreases with time in the dark.[11] If we incorporate that idea into the IDS model, that is, if we assume that dark light affects the retina just like real light does and that, during dark adaptation, dark light declines in strength, then the IDS model predicts exactly what happens during time in the dark. In fact, any model that includes dark light as equivalent to real light and that predicts Weber's law will necessarily explain dark adaptation.

To sum up, I have talked about several apparently separate and complicated perceptual phenomena—Mach bands, center-surround antagonism, the improvement in acuity as luminance increases, Weber's law, brightness constancy, Ricco's law, and light and dark adaptation. It turns out that all of those phenomena would necessarily occur if the retina contains one extremely simple mechanism, such that each receptor spreads its output over a region that is inversely related to the intensity of light falling on it. In that sense, the IDS theory is very satisfying, because it simplifies our understanding and view of the world.

Thank you again for this honor. I deeply appreciate it.

REFERENCES

1. Lederman LM. The value of fundamental science. Sci Am 1984;251:40–7.
2. Barlow HB. Temporal and spatial summation in human vision at different background intensities. J Physiol (Lond) 1958;141:337–50.
3. Cornsweet TN, Yellott J. Intensity-dependent spatial summation. J Opt Soc Am 1985, submitted for publication.
4. Kuffler SW. Discharge patterns and functional organization of mammalian retina. J Neurophysiol 1953;16:37–68.
5. Ratliff F. Mach Bands: Quantitative Studies on Neural Networks in the Retina. San Francisco: Holden-Day, 1965.
6. Arden GB. Recent developments in clinical contrast sensitivity testing. In: Breinin GM, Siegel IM, eds. Advances in Diagnostic Visual Optics. Berlin: Springer-Verlag, 1983.
7. Van Ness FL, Bouman MA. Spatial modulation transfer in the human eye. J Opt Soc Am 1967;57:401–6.
8. Kelly DH. Adaptation effects on spatio-temporal sine-wave thresholds. Vision Res 1972;12:89–101.
9. Shlaer S. The relation betwen visual acuity and illumination. J Gen Physiol 1937;21:165–88.
10. Blackwell HR. Contrast thresholds of the human eye. J Opt Soc Am 1946;36:624–42.
11. Barlow HB. Dark and light adaptation: psychophysics. In: Jameson D, Hurvich L, eds. Handbook of Sensory Physiology, vol 7, pt 4, Visual Psychophysics. Berlin: Springer-Verlag, 1972:1–28.

AUTHOR'S ADDRESS:
Tom Cornsweet
School of Social Sciences
University of California
Irvine, California 92717