

Advances in Photoreception: Proceedings of a Symposium on Frontiers of Visual Science

Committee on Vision, National Research Council

ISBN: 0-309-54339-8, 160 pages, 6 x 9, (1990)

**This PDF is available from the National Academies Press at:
<http://www.nap.edu/catalog/1570.html>**

Visit the [National Academies Press](#) online, the authoritative source for all books from the [National Academy of Sciences](#), the [National Academy of Engineering](#), the [Institute of Medicine](#), and the [National Research Council](#):

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Explore our innovative research tools – try the “[Research Dashboard](#)” now!
- [Sign up](#) to be notified when new books are published
- Purchase printed books and selected PDF files

Thank you for downloading this PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, [visit us online](#), or send an email to feedback@nap.edu.

This book plus thousands more are available at <http://www.nap.edu>.

Copyright © National Academy of Sciences. All rights reserved.

Unless otherwise indicated, all materials in this PDF File are copyrighted by the National Academy of Sciences. Distribution, posting, or copying is strictly prohibited without written permission of the National Academies Press. [Request reprint permission for this book](#).

Advances in Photoreception

**Proceedings of a Symposium on Frontiers of Visual
Science**

Committee on Vision
Commission on Behavioral and Social Sciences and Education
National Research Council

National Academy Press
Washington, D.C. 1990

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

L.C. card no. 90-60934

ISBN 0-309-04240-2

Available from: National Academy Press 2101 Constitution Avenue, N.W. Washington, D.C. 20418

S134

Printed in the United States of America

COMMITTEE ON VISION

SUZANNE MCKEE (Chair), Smith-Kettlewell Eye Research Foundation, San Francisco
ROBERT BOYNTON (NAS), Department of Psychology, University of California, San Diego
LYNN COOPER, Department of Psychology, Columbia University
RUSSELL LEE DEVALOIS (NAS), Department of Psychology, University of California, Berkeley
MERTON CLYDE FLOM, College of Optometry, University of Houston
DAVID L. GUYTON, Wilmer Ophthalmological Institute, Johns Hopkins University
DONALD HOOD, Department of Psychology, Columbia University
JAMES LACKNER, Ashton Graybiel Spatial Orientation Laboratory, Brandeis University
ROBERT SHAPLEY, Department of Psychology, New York University
LOUIS SILVERSTEIN, Honeywell, Inc., Phoenix, Ariz.
KENT A. STEVENS, Department of Computer and Information Science, University of Oregon
ANDREW B. WATSON, NASA Ames Research Center, Moffett Field, Calif.
PAMELA EBERT FLATTAU, Study Director
JOANNE ALBANES, Research Assistant
CAROL METCALF, Administrative Secretary
ROSE WHITE, Secretary

SYMPOSIUM PARTICIPANTS

ANTHONY J. ADAMS, School of Optometry, University of California, Berkeley

JAMES K. BOWMAKER, School of Biological Sciences, Queen Mary College, University of London

DONALD C. HOOD, Department of Psychology, Columbia University

WALTER MAKOUS, Center for Visual Science, University of Rochester

WILLIAM MILLER, School of Medicine, Yale University

V. HUGH PERRY, Department of Experimental Psychology, University of Oxford

EDWARD N. PUGH, Jr., Department of Psychology, University of Pennsylvania

ROBERT SHAPLEY, Department of Psychology, New York University

JULIE L. SCHNAPF, Department of Ophthalmology, University of California, San Francisco

DAVID R. WILLIAMS, Center for Visual Science, University of Rochester

JOHN I. YELLOTT, Cognitive Science Department, University of California, Irvine

PREFACE

The Committee on Vision is a standing committee of the National Research Council's Commission on Behavioral and Social Sciences and Education. The committee provides analysis and advice on scientific issues and applied problems involving vision. It also attempts to stimulate the development of visual science and to provide a forum in which basic and applied scientists, engineers, and clinicians can interact. Working groups of the committee study questions that may involve engineering and equipment, physiological and physical optics, neurophysiology, psychophysics, perception, environmental effects on vision, and visual disorders.

From time to time, the committee sponsors public meetings that feature papers on advances in vision research. The meetings are designed to aid the newcomer in reaching a preliminary understanding of the utility of the latest approaches to vision research and to challenge more experienced scientists, engineers, and clinicians alike to consider the appropriate role for these findings in the advancement of vision research and its application to practical problems.

In March 1988 the committee sponsored a Symposium on Frontiers of Visual Science at the National Academy of Sciences in Washington, D.C. The committee brought together 10 leading investigators in vision research to discuss advances in photoreception. Speakers included visual physiologists and psychophysicists, who discussed wavelength specificity of visual pigments, receptor transduction in health and disease, adaptation of photoreceptors to light, variations of the photoreceptor lattice across the retina, and photoreceptor sampling of visual images and aliasing. Comparisons of human and animal retinas also were discussed.

Funds for the symposium were provided from the general budget of the committee, which receives support from the departments of the Army, the

Navy, and the Air Force; the National Eye Institute; the National Institute on Aging; the National Aeronautics and Space Administration; the National Science Foundation, the Veterans Administration; the American Academy of Ophthalmology; and the American Academy of Optometry.

The committee gratefully acknowledges the efforts of the vision scientists who took time from their demanding schedules to participate in the symposium. The committee also thanks its staff officer, Pamela Ebert Flattau, for organizing the meeting and preparing the final report. Production of the report was effectively assisted by Carol Metcalf of the committee staff. Freelance editor Barbara A. Bodling helped improve the style and clarity of the report. To all these, we express our gratitude.

SUZANNE MCKEE, CHAIR
COMMITTEE ON VISION

CONTENTS

Introduction	1
<i>Robert Shapley</i>	
Optical and Molecular Design of Rods	2
<i>William H. Miller</i>	
Color and Wavelength Specificity of Receptors	17
Cone Visual Pigments in Monkeys and Humans	19
<i>James K. Bowmaker</i>	
Photosensitivity of Primate Photoreceptors	31
<i>Julie L. Schnapf</i>	
Transduction, Adaptation, and Disease Processes	39
The ERG and Sites and Mechanisms of Retinal Disease, Adapta- tion, and Development	41
<i>Donald C. Hood</i>	
Phototransduction in Vertebrate Rods: The Electrophysiological Approach to the cGMP Cascade Theory	59
<i>Edward N. Pugh, Jr., W.H. Cobbs, and J.D. Tanner</i>	
Partitioning Visual Processes	78
<i>Walter Makous</i>	

Spatial Effects of Photoreceptor Mosaic	103
The Distribution of Cones in the Primate Retina	105
<i>V. Hugh Perry</i>	
The Photoreceptor Mosaic as an Image Sampling Device	117
<i>John I. Yellott</i>	
The Invisible Cone Mosaic	135
<i>David R. Williams</i>	

INTRODUCTION

Robert Shapley

The visual interface between human visual perception and patterns of light in the world is the photoreceptor layer of the retina. All our experience of the appearance of objects and spatial or temporal patterns of illumination is filtered through the receptor array. All visual function thus depends on the characteristics of photoreceptors—for example, their responsiveness to different wavelengths of light, the gain and regulation of the biochemical cascade kicked off by the absorption of a photon of light, and the spatial arrangement of photoreceptors within the photoreceptor layer.

Our keynote speaker, William Miller, has contributed major insights to two of the areas of photoreceptor function: the biochemical cascade of phototransduction and the nature of photoreceptor assays. Dr. Miller was one of the first scientists to look at cyclic nucleotides as candidates for the internal messenger molecule that mediates transduction. The great discoveries of recent years about cyclic GMP, phosphodiesterase, transduction or G-protein, and the regulatory roles of calcium and guanylate cyclase are all results of experiments and models inspired by the initiatives of Dr. Miller and colleagues.

At the Frontiers of Vision symposium, Dr. Miller spoke briefly about optical effects in green rods and devoted the major part of his address to the topic of transduction. His talk is an introduction to a great problem, now apparently solved, by one of the great masters who contributed to its solution. Dr. Miller's insights into the major issues of transduction are penetrating. His summary of the process is a useful introduction to the rest of the symposium.

Optical and Molecular Design of Rods

William H. Miller

INTRODUCTION

Thank you, Bob Shapley, for such a kind introduction. I greatly appreciate it.

In trying to strike a keynote both melodious and strong enough to resonate with this audience, I think of Franz Boll, who proposed the photochemical theory of vision. Although he died of tuberculosis at the age of 30, before he had an opportunity to exploit his great discovery of visual purple, there is no doubt that he appreciated its potential as discussed in Hubbard's (1977) preface to her translation of his paper.

Not only did Boll propose the photochemical theory of phototransduction, but he also proposed a photophysical theory, which has found application in explaining the color of another of his discoveries, the green rod. Because I have been interested in transduction and the green rod, I want to share some of my thoughts on both of these problems with you.

GREEN ROD

Boll's green rod represents a design problem dealing with both visual pigments and image sampling. Neither the rod's color nor its role in sampling is completely understood. Allan Snyder and I have proposed that the green rod myoid is a wavelength filter (Miller and Snyder, 1972). Reuter and Virtanen (1972) showed that the green rod is the short-wavelength receptor for a color opponent process at photopic and mesopic light intensities. In contrast to the thin rods with pooled responses of many species, Jagger (1988) suggested that amphibian fat rods may individually sample

the eye's image. The green rod may function with both the rod and cone systems for image sampling and color vision.

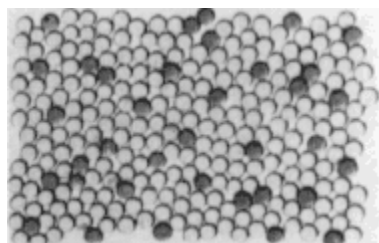


Figure 1
End-on appearance of pink and green rods. Pink rods are bleached; green rods retain their green color
SOURCE: Boll (1877).

Green Color Not Caused by Visual Pigment Alone

Boll was the first to describe the green rod. He proposed a photophysical theory of vision in addition to his photochemical theory because he was uncertain as to the origins of rod colors (Boll, 1877):

As against this photochemical theory of the nature of visual red and of the light sense, another formulation is possible which in contrast might be called the photophysical theory. This theory assumes that there is no special pigment that permeates the rod stroma, but attributes the red color of the rods to a purely physical phenomenon, the optical effect of stacking of intrinsically colorless platelets. Accordingly visual red would belong to the category of interference phenomena and more specifically to the narrower class of the so-called colors of thin films.

The photophysical theory of vision was not taken seriously after Kuhne's characterizations of rhodopsin. Nevertheless, this concept was plausible at the time. The dimensions of the rod platelets would not be known for 75 years, and in fact such platelets are found to produce interference colors in the optical trains of both invertebrate and vertebrate eyes. More to the point, a photophysical theory has merit when applied to the case of the amphibian green rod.

Figure 1 is a black and white reproduction of part of Boll's 1877 color plate showing the frog rod mosaic with its pink and green rods after illumination with short-wavelength light. The dark spots are green in the original color plate, and the light spots representing the bleached pink rods are colorless.

Figure 2 is a scanning electron micrograph of part of the photoreceptor layer of a preparation of an amphibian retina similar to that used by Boll. Although the green rod has a shorter outer segment than the pink rod, it nevertheless retains its green color after the pigment of the pink rod has

been bleached, as shown in Boll's figure. This was confirmed by Liebman and Entine (1968):

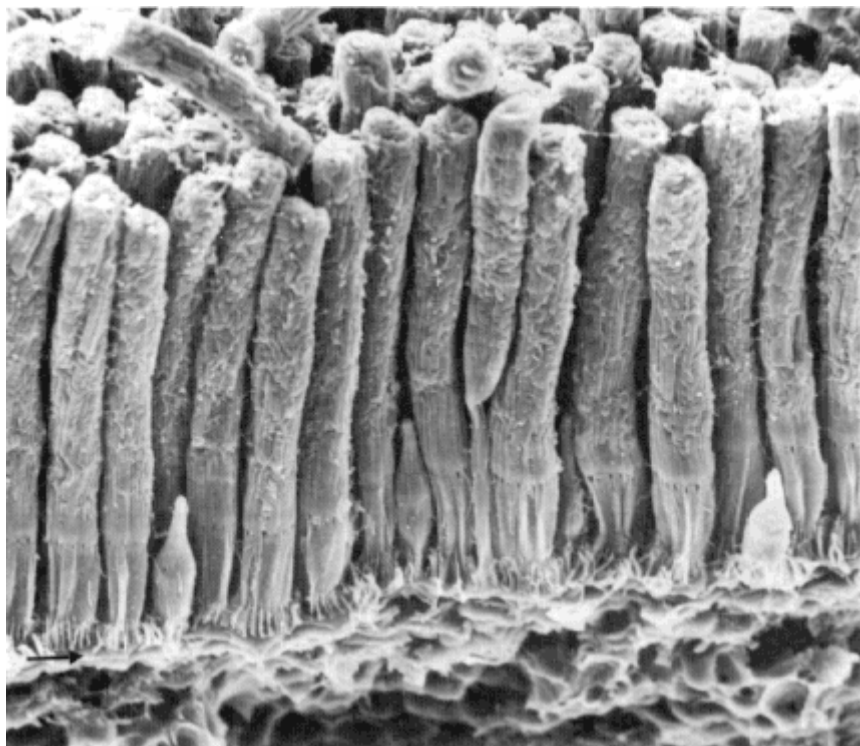


Figure 2
Scanning electron micrograph of amphibian photoreceptor layer showing a few single cones, many pink rods, and one green rod with its 1-micron diameter myoid.

NOTE: Arrow indicates the external limiting membrane.

It is of some interest that our green rods show no light absorption in the red-orange region that would explain results of Denton and Wyllie who found such absorption in end-on measurements using photographic densitometry (Denton and Wyllie, 1955). Nor, after bleaching, was an increased absorption found in the green region of the spectrum. . . . We can therefore state categorically that the locus of the green and red absorbing entities is not in pigment form in rod outer segments.

Similarly, a year earlier Dartnall (1967) also reached the conclusion that the yellow-red absorption may "even possibly be a structural color."

Myoids Are Waveguides

Miller and Snyder (1972) proposed that the myoid's waveguide properties in combination with the rod's 433-nm pigment could account for the green color. The assumption underlying this hypothesis is that the parts of the photoreceptor cells that are located sclerad to the external limiting membrane (horizontal arrow, Fig. 2) are light pipes. The green rod's aperture is its myoid near the external limiting membrane. Light that reaches the outer segment must first pass through the myoid. The physical properties of the myoid, its diameter and the difference between its refractive index and that of its surround, should determine the intensity and spectral distribution of the light that reaches the green-rod visual pigment.

The green rod's unbleached end-on color in white light results from absorption in the blue by its 433-nm pigment and "absorption" of yellow and red by an unknown structure. The myoid could be that unknown structure. Yellow-red light would be lost not by absorption but by not being propagated in the waveguide by total internal reflection. The light propagated in waveguides shows intensity patterns called modes, when the guide is viewed in cross section. For each mode the fraction of modal energy propagated inside the guide is a monotonic function of the ratio of guide diameter (d) to wavelength (λ) and of the difference in refractive index between the guide (n_1) and its cladding (n_2). The smaller d is relative to λ and the smaller $n_1 - n_2$, the less longer-wavelength light will reach the outer segment. The fraction of light energy inside the guide, η , is equal to

$$1 - \left(\frac{\lambda}{\pi d n_1} \right) \left(\frac{1}{\theta_c^2 - \theta^2} \right)^{1/2} \left(\frac{\theta}{\theta_c} \right)^2$$

where θ_c , the complement of the critical angle, is equal to $-(n_2/n_1)^2$ and θ is the angle of incidence, which is a function of the mode number (equation from Snyder, 1975:p.44).

The fraction of light (η) that would be inside a myoid of 2 μm (Figure 3 curve c), 1 μm (curve a), 0.8 μm (curve b) and 0.5 μm diameter (curve d) is plotted as a function of wavelength based on the two lowest order modes in Figure 3 from Miller and Snyder (1972).

Curve a of Figure 3 shows that the dark-adapted myoid with a diameter of 1 μm diameter would propagate yellow-red wavelengths poorly in comparison with shorter wavelengths. When viewed end on, the combination of decreased propagation in the yellow-red and absorption in the blue by the 432 nm pigment could cause the green appearance of the dark-adapted green rod. According to Arey (1916), the green rod myoid becomes slightly more tenuous when light-adapted. A 0.8 μm diameter would shift the curve to the left as shown by curve b, to mimic what was interpreted by Denton and Wyllie (1955) as an increased green absorption in the light-adapted

end-on preparation. An even smaller diameter myoid, as illustrated in curve d, would tend to cut off much of the light in the visible spectrum.

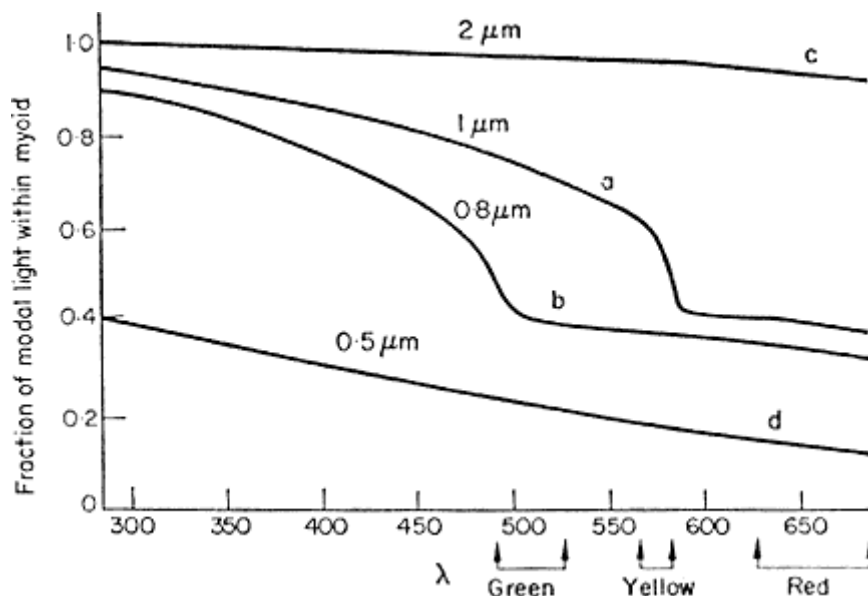


Figure 3
Theoretical effect of myoid diameter on fraction of light inside the myoid as a function of wavelength for the two lowest-order modes. SOURCE: Miller and Snyder (1972).

GREEN ROD FUNCTION

Green Rods May Be Color Receptors

If we accept that the green myoid in combination with the outer-segment 433-nm pigment can explain the rod's end-on color, that still leaves the problem of its function. Reuter and Virtanen (1972) have said:

Our experiments suggest the existence in the frog retina of separate channels for green rod responses. When during dark adaptation the red rods begin to contribute they enter the on-off and off channels of the class 1, 2 and 4 cells, *that is the cone but not the green rod channels*. This makes possible color vision in the mesopic state. The red rods simply contribute to the cone side of the opponent color system. This arrangement, together with the fact that the cone channels dominate and are sufficient for eliciting all the Maturanna et al. characteristics, guarantees the invariance of coding of brightness pattern information during the shift from photopic to scotopic vision. The main role of the green rod channels seems to be to "add color to the system."

TABLE 1 Amphibian Rod and Cone Parameters

	Green Rod	Pink Rod	Single Cone
Outer-segment diameter ^(a)	4 μ m	6 μ m	2.5 μ m
Dark-adapted entrance aperture (myoid at external limiting membrane)	$\approx 0.8\mu\text{m}^2$	$\approx 28\mu\text{m}^2$	$\approx 5\mu\text{m}^2$ (light adapted)
Outer-segment optical density, peak absorbance ^(a)	0.68	0.82	0.26
Relative anatomical sensitivity (absorbance \times capture area)	0.6	23	1
Sensitivity at peak wavelength (mV/quantum) ^(b,c)	≈ 10	≈ 10	1
Nyquist frequency based on posterior nodal distance = 5 mm (c/deg) ^(d)	2	7	?

^(a) Liebman and Entine (1968) (*Rana*)

^(b) Fain and Dowling (1973) based on intracellular recordings in intact retina, which may exaggerate sensitivity of rods because of coupling.

^(c) Reuter and Virtanen (1972) sensitivity at peak wavelength measured at ganglion cell.

^(d) Jagger (1988) (*Bufo*). Green rods comprise 11% of total in plate 1a, ref.2; Nyquist frequency calculated assuming pink rod number density = 18000/mm² and green rod number density = 1800/mm².

Are Individual Fat Rods Samplers?

The rods may provide spatial vision matched to the optics under mesopic conditions, according to Jagger (1988). Jagger argues that the rods could effectively sample the 6 cycles per degree image quality found over most of the eye's visual field. Some amphibian rod and cone parameters drawn from various sources are shown in [Table 1](#).

The green rod appears to be more sensitive than the single cone and may function as a color receptor. Comparison of the image sampling properties based on the information in the [Table 1](#) is difficult because of the wide range of eye sizes among amphibian species. Using a 5-mm posterior nodal distance (PND) eye as a sample eye size, the green rods by themselves would not be useful for spatial resolution because they could support only a Nyquist frequency of 2 cycles per degree. It seems likely, therefore, that the green and pink rods function as a unit for color and spatial vision based on a flat 6 cycles per degree optical quality over most of the retina. There would be some undersampling (Snyder et al., 1986) by the ganglion cells in the region of peak density where the potential maximum acuity would be 4.4 cycles per degree, and there would be less than 3 rods per ganglion cell away from the visual streak. Nevertheless, the theory that

the fat amphibian rods individually sample the retinal image is plausible, considering that there is much image processing prior to the ganglion cells in the amphibian retina and there is an early optokinetic acuity measure of 4.4 cycles per degree for *Rana* (reviewed in Jagger, 1988).

Rods and cones might be expected to have small diameters not only because separate optical channels should be as small as possible, consistent with the diffraction limited image, to maximize resolution but also to facilitate temporal resolution by limiting diffusion transit times. Rod outer segments of many animals ranging from fish to mammals are about 2 μm in diameter. Those of amphibians are 6 μm or larger. A design by which amphibian fat rods may provide the same effective resolution over the entire visual field in amphibians contrasts with that of many retinas with thin rods in which the rod responses are pooled to provide increased sensitivity at the expense of resolution. The green rods may have evolved to contribute to both color and form vision.

TRANSDUCTION

For the study of transduction that Boll initiated, it is now generally accepted that the rising phase of the vertebrate photoreceptor response is mediated by activation of phosphodiesterase (PDE) and that recovery is mediated at least in part by activation of guanylate cyclase (cyclase). But there is controversy regarding whether inhibition of PDE contributes to the initial phase of recovery. Hiroaki Kondo and I (Kondo and Miller, 1988) suggested that the initial phase of recovery is driven by increased rather than decreased PDE activity. Increased PDE activity may lead to decreased $[\text{Ca}]_i$ and activation of cyclase to both truncate the response and cause the smaller amplitude and faster recovering light-adapted response.

Photolyzed rhodopsin activates a cyclic nucleotide enzymatic cascade. The concentration of cyclic GMP is reduced causing decreased plasmamembrane conductance (Fesenko et al., 1985) and the decrease in dark current that is the response to illumination (see Pugh, in this volume; Stryer, 1986). Details of the recovery process and light-adapted response are still not well understood. That calcium plays a role is generally accepted: cytoplasmic calcium concentration ($[\text{Ca}]_i$) falls during the light response and that fall in $[\text{Ca}]_i$ is thought to either inhibit phosphodiesterase activity, activate guanylate cyclase activity, or both, to restore the concentration of cyclic GMP.

Is Response Recovery Mediated Initially by Cyclase Activation?

The development of the concept that recovery is mediated by activation of cyclase rests on the synthesis of two experimental facts: the finding of

McNaughton et al. (1986) that $[Ca]_i$ is decreased by illumination and the findings of Lolley and Racz (1982), Pepe et al (1986) and Koch and Stryer (1988) that very low $[Ca]_i$ activates cyclase. Hodgkin et al. (1985) combined the first hints of these findings with their observation that calcium does not close channels as rapidly as light:

Another possible explanation of the failure of calcium ions to close channels rapidly is that calcium might inhibit the production of a *substance which keeps channels open* rather than interacting directly with the channels.

What could that substance be? On one hand, Cyclic GMP is a candidate based on evidence that that substance increases conductance of an excised patch of ROS plasma membrane (Fesenko et al. 1985). On the other hand, though cyclic GMP increases channel conductance, whether that is because it binds the channel directly is not known, since the patch contains substances other than the ROS plasma membrane (Kolesnikov et al., 1987).

Nevertheless, measurements of $[Ca]_i$ have been continually refined. Dim illumination takes the dark resting level from 220 nM to ~140 nM (Ratto et al., 1988); strong illumination has been calculated to give a minimum value of 60 nM (Miller and Korenbrot, 1987). At the same time the evidence that cyclase is activated by nM levels of calcium has become convincing. Lolley and Racz (1982) were the first to use calcium at nM concentrations. Pepe et al. (1986) were able to duplicate Lolley and Racz's result on rat rods in dim red light, but only when their homogenized toad rods were illuminated. On the other hand, Koch and Stryer (1988) found that the activity of unilluminated bovine ROS cyclase was regulated by a highly cooperative calcium-dependent molecular switch (not calmodulin). Guanylate cyclase was activated up to 20-fold when $[Ca]_i$ was reduced from 200 to 50 nM from a basal rate of 7 μ M cyclic GMP/sec/ROS to 100 μ M cyclic GMP/sec/ROS. Since $[Ca]_i$ may decrease from about 200 to 60 μ M as a result of illumination, Koch and Stryer's evidence leads to the suggestion that the free concentration of cyclic GMP decreases only transiently in response to a flash, because low Ca rapidly triggers restoration of cyclic GMP levels. Pugh (in this volume) recounts other in press evidence to the same effect.

Disagreement Regarding Role of PDE Inactivation

While there is agreement that low $[Ca]_i$ aids recovery by activating cyclase, there is still no agreement as to whether the low $[Ca]_i$ also inhibits PDE. There is conflicting biochemical evidence regarding the rate of PDE inactivation. Recovery of PDE to its basal level is controlled by the rate of hydrolysis of GTP bound to alpha-transducin, T_a . On the one hand, the rate constant for transducin's GTP hydrolysis was measured to be 0.02 sec^{-1} (Bennett and Dupont, 1985), while other measurements show a

half life for $T_a \cdot \text{GTP}$ of 3 sec at 23° for cattle rods (Vuong and Chabre, 1988). Decreased PDE activity would of course aid recovery of cyclic GMP levels near the plasma-membrane channel. In this vein, Torre et al. (1986) conclude:

The results are consistent with the idea that the rapid drop in Ca_i , which has recently been shown to accompany the light response, is involved in terminating the light response, and that Ca_i is thereby involved in setting the operating point and sensitivity of phototransduction. From the comparison with other work we infer that Ca_i appears to act, at least in part, by means of control of cGMP phosphodiesterase activity. . . . Hence we conclude that *elevated* Ca_i leads to *enhanced* activity of PDEase.

An alternative hypothesis holds that lowered $[\text{Ca}]_i$ is involved in initially terminating the light response not by inactivating PDE as suggested by Torre et al. but rather by activating cyclase.

Light Response Recovery Because of PDE Activation?

Uniform pC cyclic GMP injections into ROS at the resting potential cause cyclic GMP depolarizing responses of uniform amplitude and duration, as if the cyclic GMP responses provide an index of PDE activity. Corresponding with PDE activation by light, cyclic GMP responses are antagonized in a graded manner with increasing intensities of illumination (Miller, 1982).

Uniform 5 pC cyclic GMP pulses were delivered before during and after a 0.1 sec bright light flash in the experiment shown in [Figure 4A](#). The cyclic GMP pulses are numbered 1-10. Each downward spike was an injection artifact. The numbered upward depolarizations were the cyclic GMP responses. The increased PDE activity that mediates the receptor potential was sufficient to reduce the amplitude and speed the recovery of cyclic GMP response number 2 in comparison with that of number 1. Although the receptor potential recovered rapidly, PDE activity (as reflected in the amplitude and recovery rate of the cyclic GMP response) continued to increase and did not recover to the preflash levels until more than a minute after the receptor potential recovered. The receptor potential recovered rapidly despite apparently high PDE activity. The light flash intensity was reduced by half and the flash duration was lengthened in [Figure 4B](#). The recovery rate of the cyclic GMP responses was more nearly matched to that of the receptor potential. Finally in [Figure 4C](#), in which the light flash was a log unit dimmer than that of [Figure 4B](#), the recovery rate of the receptor potential matches the recovery rate of the cyclic GMP response (and presumably PDE activity).

The initial very rapid recovery from a bright flash occurs despite an apparent high rate of PDE activity, which implies that the concentration of cyclic GMP must be restored by cyclase activation rather than PDE

inactivation. A similar conclusion follows from the experiments of Ames et al. (1986) showing that cyclic GMP metabolic turnover increases with stimulus intensity (until saturation). Ames et al. (1986) found that cyclic GMP metabolic flux increased about 4.5-fold for a 3-log unit increase in light intensity with little change in measurable cyclic GMP levels. The dimmest intensity caused an estimated 30 photoisomerizations/ROS/sec to turn over about 10^4 cyclic GMP molecules/Rh*; the brightest continuous illumination turned over 573 cyclic GMP's/Rh*/sec for a total of 4×10^6 cyclic GMP's/ROS. Koch and Stryer's maximal cyclic GMP synthesis of 100 μM /ROS/sec (on bovine in vitro rather than rabbit in vivo) corresponds to 5×10^7 cyclic GMP's/ROS/sec based on a ROS cytoplasmic volume of 10^{-12} liters. The experiments in Figure 4 showing recovery despite high PDE activity, those of Ames et al. (1986) showing increased metabolic flux of cyclic GMP with increasing light stimulus intensity and of Koch and Stryer (1988) showing a steep Ca dependence of cyclase are most easily

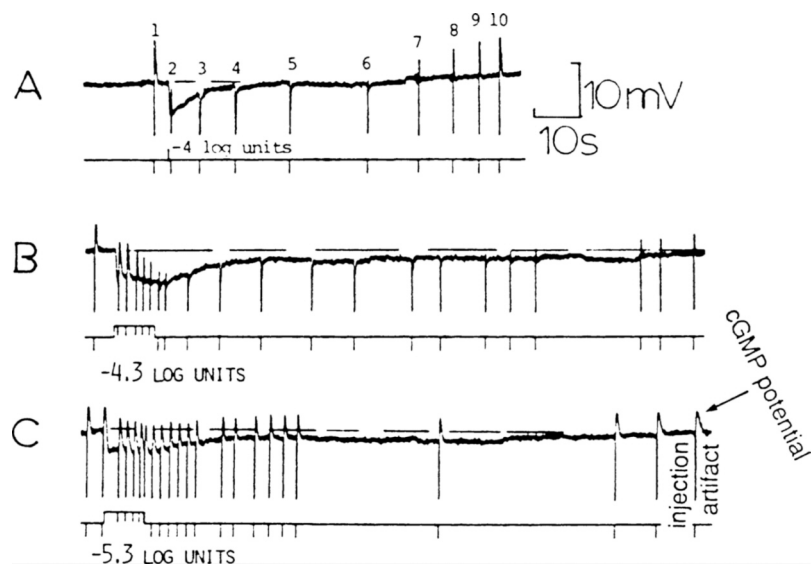


Figure 4

Voltage recordings from a single rod of the isolated frog retina for various intensities and durations of illumination. The 5-pC injections of cyclic GMP are superimposed on the traces as a measure of PDE activity. "Down" on the signal trace indicates injection; "up" signals light. Receptor potential is a hyperpolarization indicated by downward movement of the recording trace. Downward spikes are injection artifacts; cyclic GMP responses are up.

explained by a model in which the initial stage of recovery is mediated by an increase rather than a decrease in PDE activity.

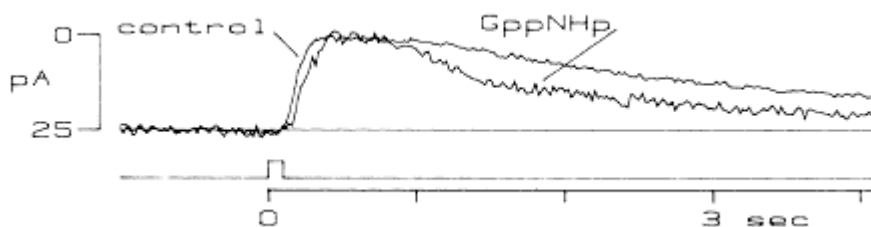


Figure 5

Suction current responses to light flash before and after application of p[NH]ppG to rod outer segment by whole-cell patch. Response after p[NH]ppG is scaled up by factor of 2.17 for normalization. Response after p[NH]ppG is truncated presumably by cyclase activation as indicated in Figure 6.

We obtained additional support for this hypothesis by testing whether the response to a light flash can recover when PDE activated by the flash is inhibited from deactivation by the introduction of the nonhydrolyzable analog of GTP, guanyl-5'-yl imidodiphosphate, p[NH]ppG, into the outer segment using the whole-cell patch clamp technique (Kondo and Miller, 1988). The p[NH]ppG both reduces the amplitude of the response and hastens the initial segment of recovery (Figure 5).

A previous experiment in which a hydrolysis resistant analog of GTP was patched into the outer segment gave a different result. There was no response recovery when 500 μ M GTP- γ -S was applied to the isolated outer segment (Sather and Detwiler, 1987). The difference between these two results is difficult to explain in the absence of information about the K_m 's of the various analogs and of GTP for cyclase, but a preliminary report of 1-2 mM p[NH]ppG patched into the inner segment of the isolated rod (Lamb and Mathews, 1986) appears to show results similar to those of the whole cell rather than the isolated outer segment.

The effects of p[NH]ppG may be explained by a model in which activated transducin leads to a new steady state of lowered dark current and hence sensitivity. An added test flash rapidly leads to still lower cytoplasmic calcium concentration, and because of the steepness of the calcium dependence of cyclase, the concentration of cytoplasmic cyclic GMP is driven rapidly higher to truncate the response. The later and slower component of recovery would be mediated by T_a -GTPase. This analogy between the effects of p[NH]ppG and the effects of light adaptation leads to the suggestion (Kondo and Miller, 1988) that transducin activated by photolyzed rhodopsin is the signal that controls the initial rod response recovery following a bright flash as well as the rapid initial recovery and

reduced sensitivity of the light-adapted response by means of acceleration of the PDE-cyclase cycle as indicated in Figure 6.

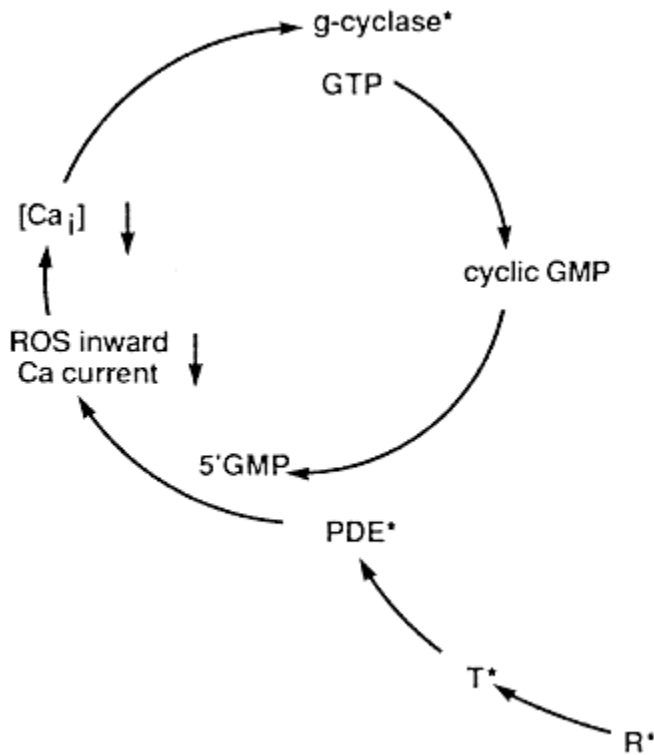


Figure 6

Schematic of hypothesized molecular mechanism of initial response recovery and of light adaptation of rod photoreceptor. Light revs up POE-cyclase cycle, causing it to race. Light adaptation reduces the ROS inward calcium current and $[Ca]_i$, which activates g-cyclase to rapidly increase cyclic GMP levels, which truncates response.

CONCLUSION

Remarquez bien que les nez ont été faits pour porter des lunettes, aussi nous-avons des lunettes.

Candide, Ch. 1:38-39

The speculators continue to give rise to suggestions for optical and physiological photoreceptor functions following the good example of Franz Boll, who was the first to practice this art in visual science. The suggestions presented at this symposium are grist for the mill. Because observation

continually improves, these speculations on the path to knowledge can never be accepted as final. Getting there is all the fun.

ACKNOWLEDGMENTS

This work was supported mainly by National Institutes of Health grant EY03196 and in part by NIH Vision Center grant EY00785, Research to Prevent Blindness, Inc., and the Connecticut Lions Eye Research Foundation.

Note added in proof: There have been significant developments since this manuscript was written. 1. Additional evidence that calcium may mediate the light-adapted response was published; reviewed by: E. Pugh and J. Altman (1988) A role for calcium in adaptation. *Nature* 334:16–17. 2. Studies showing cyclic GMP gated channel activity lent additional support for the theory that cyclic GMP may mediate transduction by directly binding the light-sensitive channel: J.W. Clack and P.J. Stein (1988) Opsin exhibits cGMP-activated single-channel activity. *Proc. Natl. Acad. Sci. USA* 85:9806–9810 and U.B. Kaupp, T. Niidome, T. Tanabe, S. Terada, W. Bonigk, W. Stuhmer, N.J. Cook, K. Kangawa, H. Matsuo, T. Hirose, T. Miyata, and S. Numa (1989) Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel. *Nature* 343:762–766.

References

- Ames, A. III, T.F. Walseth, R.A. Heyman, M. Barad, R.M. Graeff and N.D. Goldberg 1986 Light-induced increases in cGMP metabolic flux correspond with electrical responses of photoreceptors. *The Journal of Biological Chemistry* 261:13034-13042.
- Bennett, N., and Y. Dupont 1985 The G-protein of retinal rod outer segments (Transducin). *The Journal of Biological Chemistry* 260:4156-4168.
- Boll, F. 1877 On the anatomy and physiology of the retina. (Translated by Ruth Hubbard with the help of Helene Hoffmann). *Vision Research* 17:1253-1266.
- Dartnall, H.J.A. 1967 The visual pigment of the green rods. *Vision Research* 7:1-16.
- Denton, E.J., and J.H. Wyllie 1955 Study of the photosensitive pigments of the pink and green rods of the frog. *Journal of Physiology (London)* 127:81-89.
- Fain, G.L., and J.E. Dowling 1973 Intracellular recordings from single rods and cones in the mudpuppy retina. *Science* 180:1178-1180.
- Fesenko, E.E., S.S., Kolesnikov, and A.L. Lyubarsky 1985 Induction by cyclic GMP of cationic conductance in plasma membrane of retinal rod outer segment. *Nature* 313:310-313.

- Hubbard, R. 1997 Preface to the English translations of Boll's *On the anatomy and physiology of the retina* and of Kuhne's *Chemical processes in the retina*. *Vision Research* 17:1247-1248.
- Hodgkin, A.L., P.A. McNaughton, and B.J. Nunn 1985 The ionic selectivity and calcium dependence of the light-sensitive pathway in toad rods. *Journal of Physiology (London)* 358:447-468
- Jagger, W.S. 1988 Optical quality of the eye of the cane toad, *Bufo marinus*. *Vision Research* 28:105-114.
- Koch, K-W. and L. Stryer 1988 Highly cooperative feedback control of retinal guanylate cyclase by calcium. *Nature* 334:64-66.
- Kolesnikov, S.S., A.B. Jainazarov, and E.E. Fesenko 1987 Time-dependent cGMP-activated conductance of detached patches of ROS plasma membrane. *FEBS Letters* 222:37-41.
- Kondo, H., and W.H. Miller 1988 Rod light adaptation may be mediated by acceleration of the phosphodiesterase-guanylate cyclase cycle. *Proceedings of the National Academy of Sciences, USA* 85:1322-1326.
- Lamb, T.D., and H.R. Mathews 1986 Incorporation of hydrolysis-resistant analogues of GTP into the rod photoreceptors isolated from the tiger salamander. *Journal of Physiology (London)* 381:58p.
- Liebman, P.A., and G. Entine 1968 Visual Pigments of frog and tadpole (*Rana pipiens*). *Vision Research* 8:761-775.
- Lolley, R.N., and E. Racz 1982 Calcium modulation of cyclic GMP synthesis in rat visual cells. *Vision Research* 22:1481-1486.
- McNaughton, P.A., L. Cervetto, and B.J. Nunn 1986 Measurement of the intracellular free calcium concentration in salamander rods. *Nature* 322:261-263.
- Miller, D.L., and J.I. Korenbrot 1987 Kinetics of light-dependent Ca fluxes across the plasma membrane of rod outer segments. *Journal of General Physiology* 90:397-425.
- Miller, W.H. 1982 Physiological evidence that light-mediated decrease in cyclic GMP is an intermediary process in retinal rod transduction. *Journal of General Physiology* 80:103-123.
- Miller, W.H., and A.W. Snyder 1972 Optical function of myoids. *Vision Research* 12:1841-1848.
- Pepe, I.M., I. Panfoli, and C. Cugnoli 1986 Guanylate cyclase in rod outer segments of the toad retina. *FEBS Letters* 203:73-76.
- Ratto, G.M., R. Payne, W.G. Owen, and R.Y. Tsien 1988 Cytosolic free $[Ca^{2+}]$ of rods in the bullfrog retina measured with FURA2. *Biophysical Journal* 53:473a.
- Reuter, T., and K. Virtanen 1972 Border and colour coding in the retina of the frog. *Nature* 239:260-263.

- Sather, W.A., and P.B. Detwiler 1987 Intracellular biochemical manipulation of phototransduction in detached rod outer segments. *Proceedings of the National Academy of Sciences, USA* 84:9290-9294.
- Snyder, A.W. 1975 Photoreceptor optics-theoretical principles. In "Photoreceptor Optics," A.W. Snyder and R. Menzel, eds. Berlin: Springer-Verlag.
- Snyder, A.W., T.R.J. Bossomaier, and A. Hughes 1986 Optical image quality and the cone mosaic. *Science* 231:499-501.
- Stryer, L. 1986 Cyclic GMP cascade of vision. *Annual Reviews of Neuroscience* 9:87-119.
- Torre, V., H.R. Matthews, and T.D. Lamb 1986 Role of calcium in regulating the cyclic GMP cascade of phototransduction in retinal rods. *Proceedings of the National Academy of Sciences, USA* 83:7109-7113.
- Vuong, T.M., and M. Chabre 1988 As measured by fast microcalorimetry, bound GTP in transducin is hydrolysed within 3 seconds in cattle rods at 23°C. *Biophysical Journal* 53:472a.

COLOR AND WAVELENGTH SPECIFICITY OF RECEPTORS

Cone Visual Pigments in Monkeys and Humans

James K. Bowmaker

Color vision in humans is based on three spectrally different classes of photoreceptors, that is, three classes of cones containing different visual pigments. The absorbance spectra of these different pigments were first established by direct microspectrophotometric measurements from individual cones by Marks et al. in 1964. Their data, along with the data of Brown and Wald (1964), showed the expected clear evidence for three types of cones with maximum absorbance, λ_{\max} , at about 560, 530, and 440 to 460 nm. Since then various attempts have been made to establish with more precision the λ_{\max} and the shape of the absorbance spectra of the three pigments in humans (Dartnall et al., 1983) and to compare them with those found in Old World monkeys (Bowmaker et al., 1983; MacNichol et al., 1983). The visual pigments of New World monkeys have also been investigated and correlated with the marked variations in color vision found in these species (e.g., Mollon et al., 1984b).

This paper reviews the present position, as determined from microspectrophotometry, and addresses the following questions: How many visual pigments are available to primates? Are there spectrally preferred positions for the pigments? If so, can they be related to anomalous pigments in humans? In addition, some controversy has arisen as to the spectral location of short-wave receptors in primates (Mansfield et al., 1984), and I shall present further evidence from both Old and New World monkeys suggesting differences between species.

Microspectrophotometric recording from primate photoreceptors is difficult: the cone outer segments are small, with diameters of only 1 to 2 μm , and they break down structurally within a relatively short time. Thus, the transverse absorbance spectra recorded from them have maximum densities of only about 0.015 to 0.020 with low signal-to-noise ratios. (For

a review see MacNichol et al., 1983). In analyzing the data we set rigid criteria by which to select records from individual cones and then estimate the λ_{\max} by fitting a template curve to the absorbance spectrum (see Mollon et al., 1984, for details). Ideally, the template curve should be expressed in such a way that it remains constant in shape when displaced throughout the spectrum, and my co-workers and I have previously found that the Dartnall nomogram (based on frog rhodopsin with λ_{\max} at 502 nm) expressed on a scale of $\lambda^{1/4}$ is a reasonable approximation (Barlow, 1982; Mollon et al., 1984b). More recently, Mansfield (1985) and MacNichol (1986) have demonstrated that a transformation to F/Fmax (frequency/frequency maximum) produces a better fit to experimental data, and we have adopted this template in the analysis of all the data presented here. The difference in the λ_{\max} determined from the two different templates is of course zero near 500 nm (both based on frog rhodopsin), but it increases to about 2 nm at 565 nm, with the F/Fmax value being the shorter, since at longer-wavelengths the transformation, when expressed on a wavelength basis, yields a slightly broader curve than the $\lambda^{1/4}$ transformation.

LONG- TO MIDDLE-WAVE PIGMENTS

Old World Monkeys and Humans

The cone pigments in macaques are well established. It is clear that the two pigments in the red-green spectral region have λ_{\max} at about 565 and 535 nm (Bowmaker et al., 1983; MacNichol et al., 1983). We have now had the opportunity to look at three other groups of Old World primates: the baboon, *Papio papio* (Bowmaker et al., 1983), and two types of guenon, the talapoin, *Cercopithecus (Miopithecus) talapoin*, and the moustached guenon, *Cercopithecus cephus*. These species extend the range of platyrrhines studied. In addition, *C. cephus* is of particular interest because of its facial coloration. The moustached guenon, or blue-faced monkey, is characterized by a strikingly blue face surrounded by yellow hair, and we hoped, perhaps naively, that its retina would have a somewhat higher concentration of short-wave cones. We were fortunate in obtaining data from six such cones, about 10 percent of the total, more than double the numbers we have recorded from other species (Table 1).

It is clear from Figure 1 that the distributions of cones in the four species of Old World primates are similar and that the mean λ_{\max} for the long-wave cones and that for the middle-wave cones are almost identical, with overall means of 565.2 and 535.1 nm (Table 1). The one rhesus monkey listed in the present data yielded relatively few long-wave cones with a mean λ_{\max} at about 562 nm, but our earlier data (Bowmaker et al., 1983) suggest a λ_{\max} closer to 565 nm.

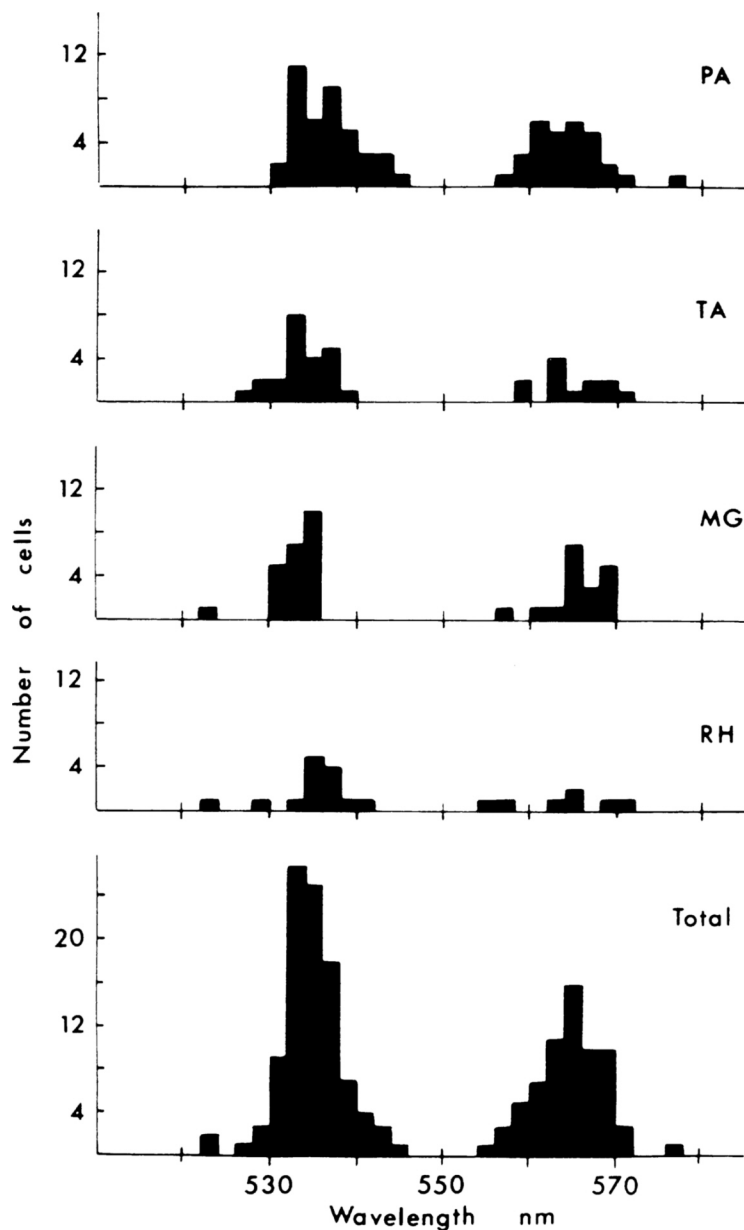


Figure 1
Distribution of values of peak sensitivity of individual cones (maximally sensitive in the red-green spectral range) from four species of Old World primates: PA, *Papio papio* (baboon) (four animals); TA, *Cercopithecus talapoin* (talapoin); MG, *Cercopithecus cephus* (moustached guenon); and RH, *Macaca mulatta* (rhesus). The bottom panel is a summary histogram of the four species.

TABLE 1 Old World Primate Visual Pigments (λ_{max} in nm)

	Short-wave	Middle-wave	Long-wave
<i>Macaca mulatta</i> (Rhesus) (1)	432.8 ± 4.7 (1)	534.6 ± 3.1 (14)	561.5 ± 2.1 (7)
<i>Cercopithecus cephus</i> (1)	430.5 ± 2.2 (6)	533.0 ± 1.1 (23)	566.5 ± 1.4 (18)
<i>Cercopithecus talapoin</i> (1)	430.1 ± 2.5 (2)	534.3 ± 1.3 (23)	566.1 ± 2.3 (12)
<i>Papio papio</i> (baboon) (4)	428.5 ± 3.3 (4)	536.4 ± 1.2 (40)	564.5 ± 1.2 (30)
Total Old World (7)	430.9 ± 0.9 (13)	535.1 ± 0.5 (100)	565.2 ± 0.9 (67)
Human (5)	419.0 ± 2.1 (4)	531.4 ± 0.7 (71)	563.7 ± 0.6 (102)

NOTE: Numbers in parentheses are either numbers of animals or numbers of cells analyzed. Max values are ± S.D.

Are the cone pigments in humans similar to those in Old World primates? We have recently made measurements from another five human eyes enucleated because of melanoma; the distributions of the long- and middle-wave cones are shown in [Figure 2](#). These distributions are remarkably similar to those of the Old World primates, with mean λ_{max} of 563.7 and 531.4 nm ([Table 1](#)): a similarity reflected in the spectral sensitivity measurements obtained from suction electrode techniques both in macaques (Baylor et al., 1987) and in humans (Schnapf et al., 1987). The bimodality in the populations of both long- and middle-wave cones suggested in our earlier data from human eyes (Dartnall et al., 1983) was not apparent in the present study.

A difference that is apparent between Old World monkeys and humans, however, is the ratio of long- to middle-wave cones. In each of the four species of Old World primates, there are fewer long-wave cones with an overall ratio of 0.77 (101:131), whereas in the human data the situation is reversed, with a ratio of long to middle of 1.67 (164:98). These ratios are based on the total number of cells identified, not the numbers of records that passed our criteria, as listed in [Table 1](#). Similar ratios were found in our earlier studies: 0.73 for macaques (Bowmaker et al., 1983) and 1.3 for humans (Dartnall et al., 1983). [The ratios given by Boynton (1988, [Table 1](#)) are incorrect: he reversed our numbers of long- and middle-wave cones.] In their suction electrode measurements of macaque foveal cones, Baylor et al. (1987) found a ratio closer to 1.0 with 192 long-wave cones and 211 middle-wave cones. It is not clear whether the difference we found

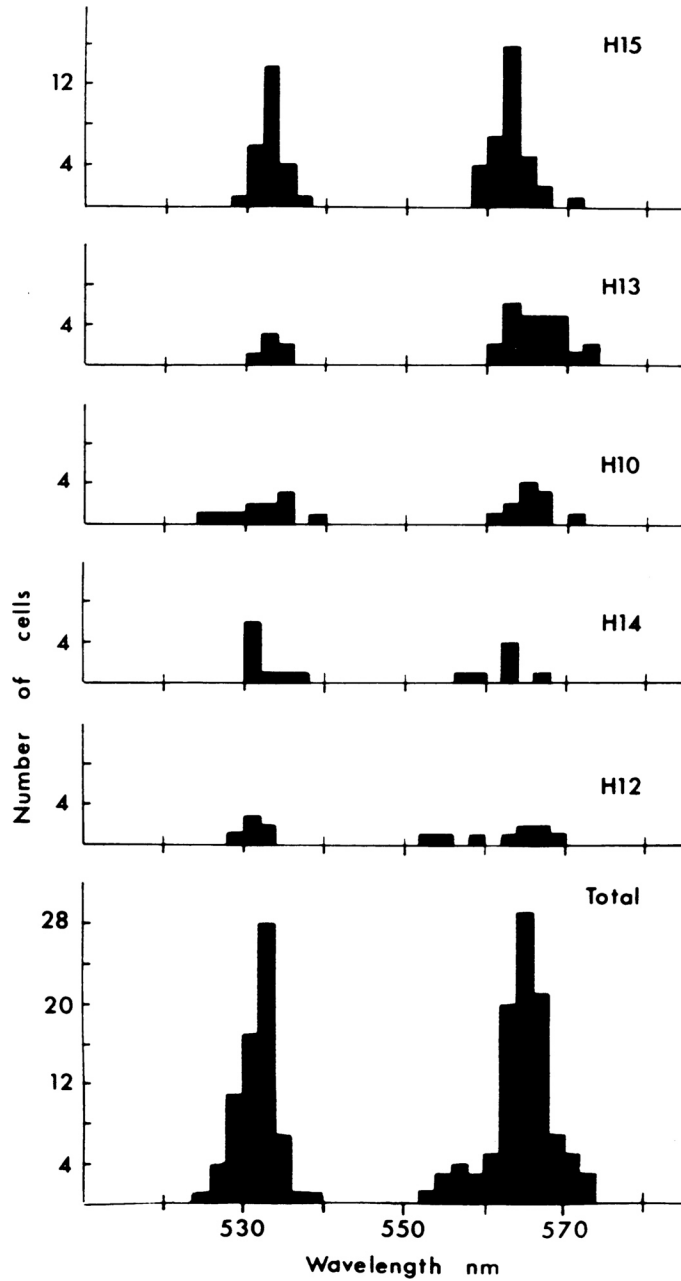


Figure 2
Distribution of values of peak sensitivity of individual cones (maximally sensitive in the red-green spectral range) from the eyes of five persons. The bottom panel is a summary histogram of the five individuals.

between Old World monkeys and humans is real or simply the result of sampling biases.

New World Monkeys

The cone pigments present in New World monkeys are markedly different from those of Old World monkeys, for New World monkeys show a striking polymorphism of visual pigments maximally sensitive in the red-green spectral region. This first became apparent in studies of color vision in the squirrel monkey, *Saimiri sciureus* (Jacobs et al., 1981). It is now clear that this species has three pigments available in the middle to long wavelengths with λ_{\max} (determined from microspectrophotometry) at about 536, 550, and 563 nm (Mollon et al., 1984b). The pigments with λ_{\max} at about 536 and 563 nm are very similar in spectral location to those of Old World monkeys, but the P550 appears to be unique to New World species.

The cone pigment complement in squirrel monkeys varies between individuals such that all males possess only one pigment in the red-green spectral region. This, combined with a short-wave pigment (see below), allows the animal only dichromatic color vision (Mollon et al., 1984b). However, the single pigment in the middle- to long-wave region may be any one of the three available to the species. In the females the position is more complex in that some animals are dichromats, resembling the males, whereas others enjoy trichromacy and possess, in addition to the short-wave pigment, a combination of any two of the three available longer-wave pigments (Bowmaker et al., 1985, 1987). The polymorphism of pigments and the resulting variation in color vision between individuals can be explained by a genetic model (Mollon et al., 1984b; Jacobs and Neitz, 1985) that is distinctly different from that proposed for the inheritance of color vision in humans and, by inference, in all Old World primates. The model postulates that

- there is only one genetic locus for a pigment in the red-green spectral range;
- there are three alleles that can occur at the locus, the three alleles corresponding to the three slightly different opsins of the photopigments;
- the locus is on the X-chromosome; and
- in those females that are heterozygous at the locus, only one of the two alleles is expressed, owing to Lyonization or X-chromosome inactivation, so that only one pigment is manufactured in any one cell.

The polymorphism expressed in squirrel monkeys is not unique to the cebid monkeys. It occurs in a similar form in the other major group of New World monkeys—the callitrichids (Travis et al., 1985; Jacobs et al., 1987). In the common marmoset, *Callithrix jacchus jacchus*, three pigments are available in the red-green spectral region, but these are different from those in *Saimiri*, with λ_{\max} at about 543, 557, and 564 nm (Travis et al., 1988). The P564 is again similar to the long-wave pigment in Old World monkeys, but the two pigments at shorter wavelengths are not common to Old World species, although there are reports of cones from macaques containing a pigment with λ_{\max} at about 543 nm (Bowmaker et al., 1978; MacNichol et al., 1983). As with *Saimiri*, it appears that all males are dichromats, pairing one of the three pigments with a short-wave pigment common to all individuals, whereas females may be either dichromatic or trichromatic. At present we have good evidence for only one form of trichromacy, the P564 combined with the P543 (Travis et al., 1988) but with indications (unpublished) of the other two predicted variants. In its simplest form the genetic model suggests that two-thirds of females are trichromatic. However, within our limited sample of marmosets, there appear to be fewer trichromats than the model predicts.

Spectral Location of Pigments

Within the primates both from the Old World and the New World, there would appear to be distinct spectral locations for cone photopigments in the red-green spectral range. Thus, Old World monkeys (including humans) have pigments at 535 and 565 nm, while New World monkeys may have one of these but may also have intermediate pigments with λ_{\max} at about 543, 550, and 557 nm. This is illustrated in Figure 3, which shows the pigments found in each species, and there are indications of specific spectral locations each separated by about 6 to 7 nm. The same spacing can be applied within mammals generally (see Jacobs, 1987) and is reminiscent of the idea of "clustering" of visual pigments first proposed by Dartnall and Lythgoe (1965) for extracts of rod visual pigments (see also Knowles and Dartnall, 1977, pp. 195–199). Presumably there are constraints on the structure of opsins such that variations in their amino acid sequences do not result in visual pigments with a continuum of spectral sensitivities but in pigments that lie at discrete spectral locations.

Can such a hypothesis reveal anything about the spectral location of anomalous pigments in colour-deficient human observers? The work of Nathans et al. (1986) has shown that the opsin sequences of the long- and middle-wave cone pigments in humans are more than 90 percent homologous and that in anomalous trichromats the hybrid genes probably

encode for the pigments that lie between the normal long- and middle-wave pigments. The clustering of primate pigments implies that there are potentially three spectral locations for these pigments—543, 550, and 557 nm. It is interesting that in analyses of psychophysical measurements on anomalous observers, Pokorny et al. (1973) estimated that the normal and anomalous pigments are separated by only 6 to 7 nm. Thus, protanomalous observers may have a P535-P542 combination, whereas deuteranomalous observers may have a P564 combined with a P557. Unfortunately, no direct measurements of anomalous pigments by either microspectrophotometry or suction electrode techniques have been made. The only such measurements from the eye of a color-deficient observer are those from a deuteranope, where only short- and long-wave cones were identified (Mollon et al., 1984a).

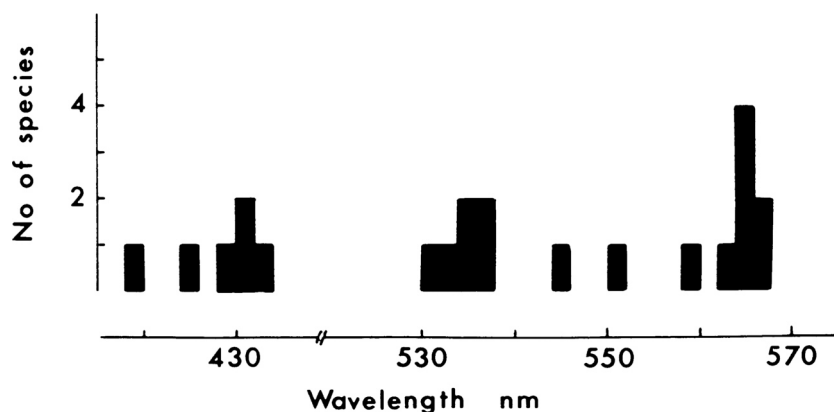


Figure 3

Distribution of values of the mean peak sensitivities of the cone types identified in humans, four species of Old World primates, and two species of New World primates. Note the dominant peak wavelengths at about 565, 535, and 430 nm and the suggestion of clustering with a wavelength interval of about 5 to 8 nm.

SHORT-WAVE CONES

Old World Monkeys and Humans

Attempts to establish by microspectrophotometry the spectral location and shape of the absorbance spectra of short-wave cone pigments have been made more difficult by the relative rarity of short-wave cones, normally less than 10 percent of the cone population, and by the problems of short-wave scatter and the probable contamination of spectra by photoproducts (for a discussion, see Mansfield et al., 1984). The best estimate for the λ_{max} of short-wave cones in macaques would appear to be about 430 nm

(MacNichol et al., 1983; Mansfield et al., 1984); our more recent data from four species of platyrrhines support this value (Table 1).

In the moustached guenon we obtained data from six short-wave cones; the mean absorbance spectrum is shown in Figure 4a. These prebleach spectra may be contaminated by short-wave light scatter and by photostable pigments. Each cone was exposed to white light to bleach the photopigment, and the spectrum was measured again, from which a difference spectrum representing the bleached photopigment could be derived (Figure 4a). The difference spectrum should eliminate problems caused by light scatter and photostable pigments, but it may be distorted by photoproducts. The λ_{\max} of the prebleach and that of the difference spectra were then determined by fitting the template curve to the right-hand limb. As can be seen, both sets of data are best fitted by a template with λ_{\max} at about 431 nm. Similar fits were obtained for the rhesus, talapoin, and baboons, and the overall λ_{\max} was again 431 nm. This value is identical to that reported by Baylor et al. (1987) from suction electrode measurements of spectral sensitivity in short-wave cones from macaques.

In humans the situation appears to be different. In Figure 4b the mean spectra from four human short-wave cones are shown, with a template curve with λ_{\max} 420 nm superimposed. Although the difference spectrum is narrower than the template, it is clear that the spectra lie at wavelengths distinctly shorter than those of Old World monkeys. The λ_{\max} of 420 nm agrees with our earlier published data for human short-wave cones (Dartnall et al., 1983).

New World Monkeys

In the cebid, *Saimiri sciureus*, we have data from 16 short-wave cones. The spectra, both prebleach and difference, are best fitted by a template curve with λ_{\max} at about 430 nm (for spectra, see Mollon et al., 1984b; Bowmaker et al., 1985). The photopigment is clearly spectrally very similar to those in Old World monkeys. However, the short-wave cones in the callitrichid, *Callithrix jacchus*, appear to be different. The mean spectra from 14 cones are shown in Figure 4c and are best fitted by a template with λ_{\max} at about 424 nm (Travis et al., 1988).

Spectral Location of Pigments

As with the photopigments maximally sensitive in the red-green spectral region, the short-wave pigments of primates also appear to cluster at distinct spectral locations—those of Old World monkeys and squirrel monkeys cluster at about 431 nm, that of the marmoset at about 424 nm, and that of humans at about 420 nm (Figure 3). The mean spectra for the three pigments are shown together in Figure 4d, where the spectral displacements

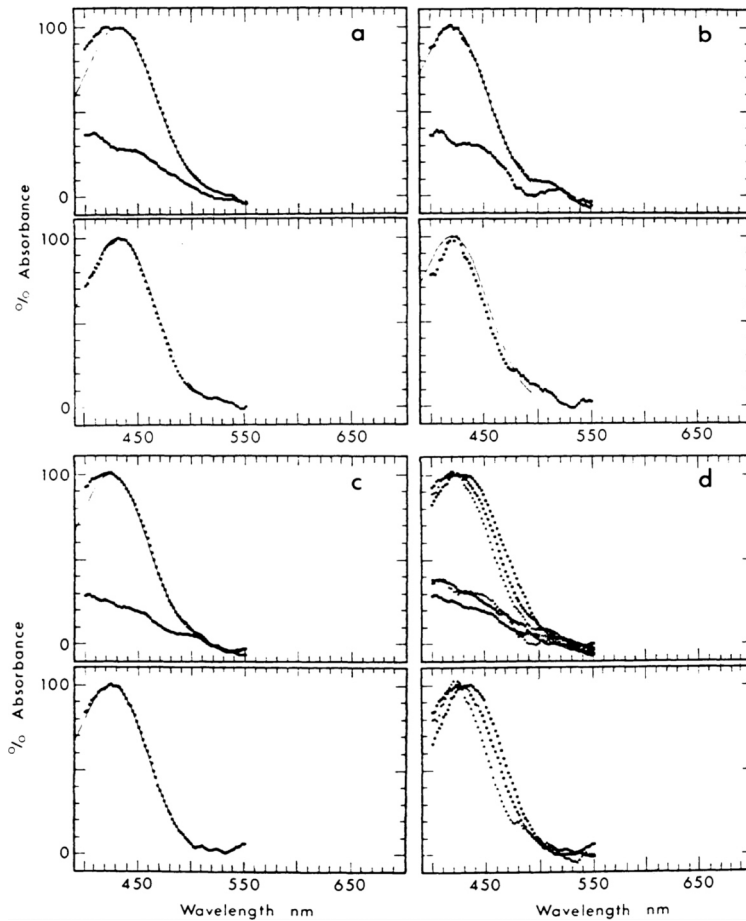


Figure 4

Absorbance and difference spectra from short-wave cones of primates. The top panel of each pair plots the mean absorbance spectra before and after bleaching, and the bottom panel plots the difference spectrum derived from the bleaching. The prebleach spectra and difference spectra have been normalized to 100 percent. In the upper left are the mean spectra from six cones from *Cercopithecus cephus* (moustached guenon), with a 431-nm template superimposed. In the upper right are the mean spectra from four cones from human eyes, with a 420-nm template. In the lower left are the mean spectra from 14 cones from *Callithrix jacchus* (common marmoset), with a 424-nm template. In the lower right is a comparison of the mean spectra from Old World primates (longest-wave), *Callithrix* (middle-wave), and humans (shortest-wave).

are clear. Again, this is reminiscent of the 6-nm spacing of the longer-wave pigments (Figure 3) and may imply that there are constraints on the structure of visual pigments that determine their spectral sensitivities. Clearly, to elucidate such problems, more information is needed on the amino acid sequences of a number of primate opsins as well as a further understanding of the tertiary arrangement of opsin and retinal within the disc membranes of the cone outer segments.

ACKNOWLEDGMENT

The microspectrophotometry summarized here was conducted with support by the Medical Research Council of the United Kingdom grant G8512759N.

References

- Barlow, H.B. 1982 What causes trichromacy? A theoretical analysis using combfiltered spectra. *Vision Research* 22:635-643.
- Baylor, D.A., B.J. Nunn, and J.L. Schnapf 1987 Spectral sensitivity of cones of the monkey *Macaca fascicularis*. *Journal of Physiology* 390:145-160.
- Bowmaker, J.K., H.J.A. Dartnall, J.N. Lythgoe, and J.D. Mollon 1978 The visual pigments of rods and cones in the rhesus monkey, *Macaca mulatta*. *Journal of Physiology* 274:329-348.
- Bowmaker, J.K., J.D. Mollon, and G.H. Jacobs 1983 Microspectrophotometric results for Old and New World primates. In *Colour Vision: Physiology and Psychophysics*, J.D. Mollon and L.T. Sharpe, eds. London: Academic Press.
- Bowmaker, J.K., G.H. Jacobs, D.J. Spiegelhalter, and J.D. Mollon 1985 Two types of trichromatic squirrel monkey share a pigment in the red-green spectral range. *Vision Research* 25:1937-1946.
- Bowmaker, J.K., G.H. Jacobs, and J.D. Mollon 1987 Polymorphism of photopigments in the squirrel monkey: a sixth phenotype. *Proceedings of the Royal Society of London B* 231:383-390.
- Boynton, R.M. 1988 Colour vision. *Annual Review of Psychology* 39:69-100.
- Brown, P.K., and G. Wald 1964 Visual pigments of single rods and cones of the human retina. *Science* 144:45-51.
- Dartnall, H.J.A., and J.N. Lythgoe 1965 The spectral clustering of visual pigments. *Vision Research* 5:81-100.
- Dartnall, H.J.A., J.K. Bowmaker, and J.D. Mollon 1983 Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal Society of London B* 220:115-130.
- Jacobs, G.H. 1987 Cone pigments and color polymorphism: a comparative perspective. In *Frontiers of Visual Science: Proceedings of the 1985 Symposium*. Committee

- on Vision, National Research Council. Washington, D.C.: National Academy Press.
- Jacobs, G.H., and J. Neitz 1985 Color vision in squirrel monkeys: sex-related differences suggest the mode of inheritance. *Vision Research* 25:141-143.
- Jacobs, G.H., J.K. Bowmaker, and J.D. Mollon 1981 Behavioural and microspectrophotometric measurements of colour vision in monkeys. *Nature* 292:541-543.
- Jacobs, G.H., J. Neitz, and M. Crognale 1987 Color vision polymorphism and its photopigment basis in a callitrichid monkey (*Sanguinus fuscicollis*). *Vision Research* 27:2089-2100.
- Knowles, A., and H.J.A. Dartnall 1977 The photobiology of vision. In *The Eye*, vol. 2B, H. Davson, ed. New York: Academic Press.
- MacNichol, E.F. 1986 A unifying presentation of photopigment spectra. *Vision Research* 26:1543-1556.
- MacNichol, E.F., J.S. Levine, R.J.W. Mansfield, L.E. Lipetz, and B.A. Collins 1983 Microspectrophotometry of visual pigments in primate photoreceptors. In *Colour Vision: Physiology and Psychophysics*, J.D. Mollon and L.T. Sharpe, eds. London: Academic Press.
- Mansfield, R.J.W. 1985 Primate photopigments and cone mechanisms. In *The Visual System*, A. Fein and J.S. Levine, eds. New York: Alan R. Liss.
- Mansfield, R.J.W., J.S. Levine, L.E. Lipetz, B.A. Collins, G. Raymond, and E.F. MacNichol 1984 Blue-sensitive cones in the primate retina: microspectrophotometry of the visual pigment. *Experimental Brain Research* 56:389-394.
- Marks, W.B., W.H. Dobbelle, and E.F. MacNichol 1964 Visual pigments of single primate cones. *Science* 143:1181-1183.
- Mollon, J.D., J.K. Bowmaker, H.J.A. Dartnall, and A.C. Bird 1984a Microspectrophotometric and psychophysical results for the same deuteranopic observer. In *Colour vision deficiencies VII*. G. Verriest, ed. The Hague: W. Junk.
- Mollon, J.D., J.K. Bowmaker, and G.H. Jacobs 1984b Variations in colour vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proceedings of the Royal Society of London B* 322:373-399.
- Nathans, J., D. Thomas, and D.S. Hogness 1986 Molecular genetics of human color vision: the genes encoding blue, green and red pigments. *Science* 232:193-202.
- Pokorny, J., V.C. Smith, and I. Katz 1973 Derivation of the photopigment absorption spectra in anomalous trichromats. *Journal of the Optical Society of America* 63:232-237.
- Schnapf, J.L., T.W. Kraft, and D.A. Baylor 1987 Spectral sensitivity of human cone photoreceptors. *Nature* 325:439-441.
- Travis, D.S., J.K. Bowmaker, and J.D. Mollon 1985 Polymorphism of retinal photopigments in the common marmoset (*Callithrix jacchus*). *Perception* 14:A16.
- 1988 Polymorphism of visual pigments in a callitrichid monkey. *Vision Research* 28:481-490.

Photosensitivity of Primate Photoreceptors

Julie L. Schnapf

Our ability to discriminate different wavelengths of light depends on the different spectral sensitivities of the three kinds of cones in our retinas. These spectral functions have been determined in rods and cones of macaque and human retinas by recording their photocurrents. The spectra obtained in these experiments provide the receptor basis for human color-matching functions and luminosity functions. Results have been described in detail by Baylor et al. (1984, 1987) and Schnapf et al. (1987, 1988).

METHODS

Using a suction pipette, photocurrent was recorded from single rods and cones of the monkey *Macaca fascicularis*. Psychophysical experiments (DeValois et al., 1974) suggest that this monkey's photoreceptors are similar to those of humans. A limited number of electrical recordings were also made in cones of a single human eye (Schnapf et al., 1987), and the results support that suggestion.

Photoreceptors were illuminated by light incident perpendicular to the long axes of their outer segments so that pigment self-screening was negligible. Assuming then that the quantal efficiency of bleaching is constant for the wavelengths used, the spectra obtained here should be proportional to the probability that a single photopigment will absorb light as a function of wavelength.

Spectral sensitivity was obtained by the method of criterion response (Naka and Rushton, 1966; Baylor et al., 1984). Brief flashes of light were applied at a test wavelength and a reference wavelength (500 nm) to determine the relative amount of light at the two wavelengths required to elicit responses of criterion size. The reciprocal of the flash photon densities

of the two lights gave the relative spectral sensitivity. The responses of the photoreceptors appeared to obey the "principle of spectral univariance," depending only on the number of photons absorbed and not the wavelength (Naka and Rushton, 1966). Results were obtained from 51 rods and cones in the macaque retina and 6 cones in a human retina.

RESULTS

Spectral Sensitivity

Spectral sensitivities of the macaque photoreceptors are shown in [Figure 1](#). The symbols plot the normalized average values obtained from the blue (\square), green (Δ), and red cones (\circ) and the rods (\bullet) of 12 macaque monkeys. The curves near the cone spectra are drawn by eye, while the rod curve is the Dartnall nomogram with a peak sensitivity at 491 nm. The peaks of the cone spectra lie at about 430, 530, and 560 nm.

Spectra obtained from human cones are shown in [Figure 2](#). The points plot the average normalized spectra from five red cones (A) and one green cone (B). The smooth curves are the same as those drawn through the corresponding macaque spectra of [Figure 1](#). Spectra obtained from the two species are indistinguishable. The amplitude and waveform of the light responses are also very similar for the two species.

The spectra of individual cones varied little for cells within a class. For a given class of cone, the standard deviation of the positions of the spectra along the abscissa was less than 1.5 nm. (Spectral position was estimated by fitting a straight line to the descent of sensitivity at low wavenumber.) This amount of variability is similar to that estimated from human color-matching experiments (MacLeod and Webster, 1983; Neitz and Jacobs, 1986).

When plotted on a normalized wavenumber scale, where wavenumber is divided by the wavenumber of maximum absorption, the absorption spectra of the four macaque photoreceptors have a common shape. The action spectra shown here confirm this observation, although the normalized rod spectrum is slightly broader than that of the cones. It is not known what, if any, physical mechanism might be implied by this scaling principle. It appears, however, that the shape of the normalized spectrum is species specific, as the normalized action spectrum of photoreceptors of the ground squirrel differs significantly from that of the macaque (Kraft, 1988).

Color Matching

How can the action spectra of the cones be compared to human color vision? Since the electrical response in a cone outer segment depends

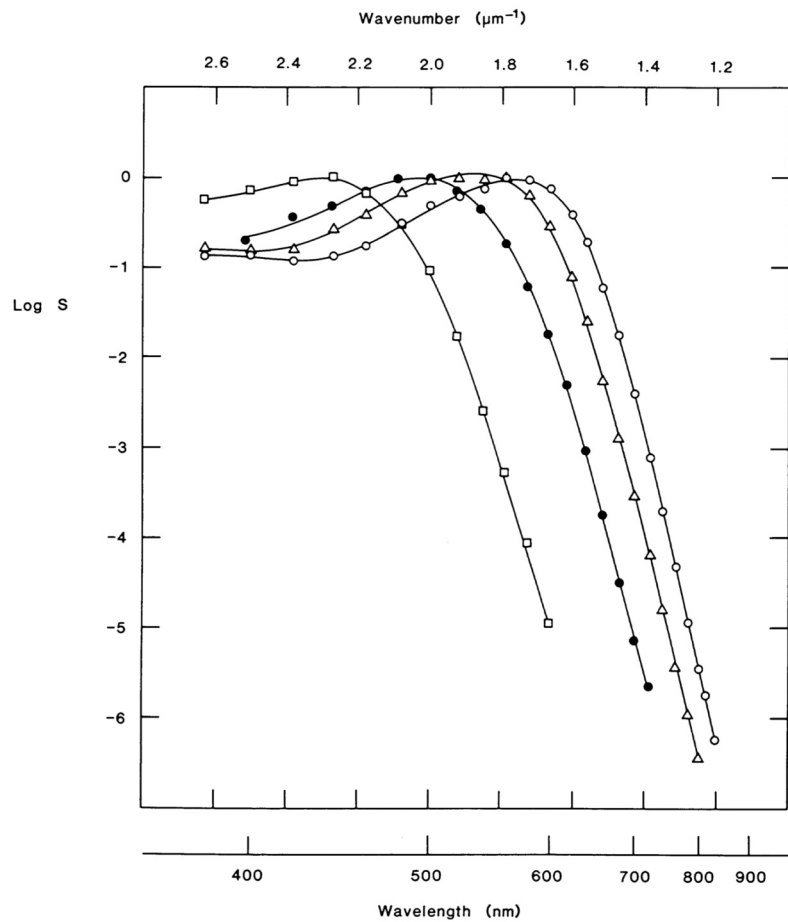


Figure 1
Spectral sensitivities of macaque photoreceptors. Points plot the average normalized sensitivities of 5 blue cones (\square), 20 green cones (Δ), 16 red cones (\circ), and 10 rods (\bullet), as a function of wavenumber. Wavelength scale above.
SOURCE: Schnapf et al. (1988).

only on the number of photons absorbed and not wavelength, it follows that stimuli comprised of different combinations of wavelengths will be indistinguishable to us if the number of quantal absorptions evoked in our cones is identical. The intensities of the component wavelengths required to equate quantal absorptions elicited by physically different stimuli may be calculated from the spectral sensitivity functions of the three cones.

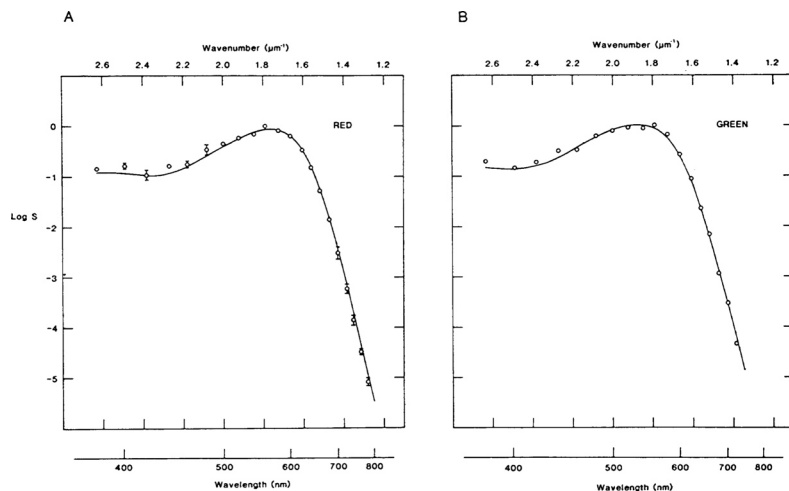


Figure 2

Spectral sensitivity of human cones. Points plot the average normalized sensitivities of five red cones in A and one green cone B as a function of wavenumber; wavelength above. The bars in A are the standard deviations. The curves are the same as those for the corresponding spectra in Figure 1. SOURCE: Schnapf et al. (1987).

The color-matching experiments of Stiles and Burch (1955) were used for the calculations. In these psychophysical experiments, observers adjusted the intensities of long-, middle-, and short-wavelength "primaries" so that when the three lights were superimposed the stimulus looked identical to a single "test" wavelength of unit intensity. The smooth curves in Figure 3 show the intensities of the primaries (incident on the cornea) plotted as a function of the wavelength of the test light. The symbols indicate the expected color-matching functions calculated from the spectral sensitivity measurements of the macaque cones of Figure 1. The macaque spectra were first adjusted to take into account differences in the way the light was applied in the two kinds of experiments. Specifically, the effects of light absorption by the human lens and macular pigment were taken into account as well as photopigment self-screening within the cone outer segments in the psychophysical experiment. The forms of the preretinal absorption spectra were taken from Wyszecki and Stiles (1982); self-screening was assumed to follow Beer's Law. The absolute magnitude of the lens,

macular, and photopigment densities (given in Figure 3's caption) were chosen to give the best fit between the two sets of data. A good correspondence was obtained between the psychophysical and adjusted physiological measurements, supporting the idea that color-matching functions may be derived simply from the absorption characteristics of the cone pigments and that the macaque and human cone pigments are very similar.

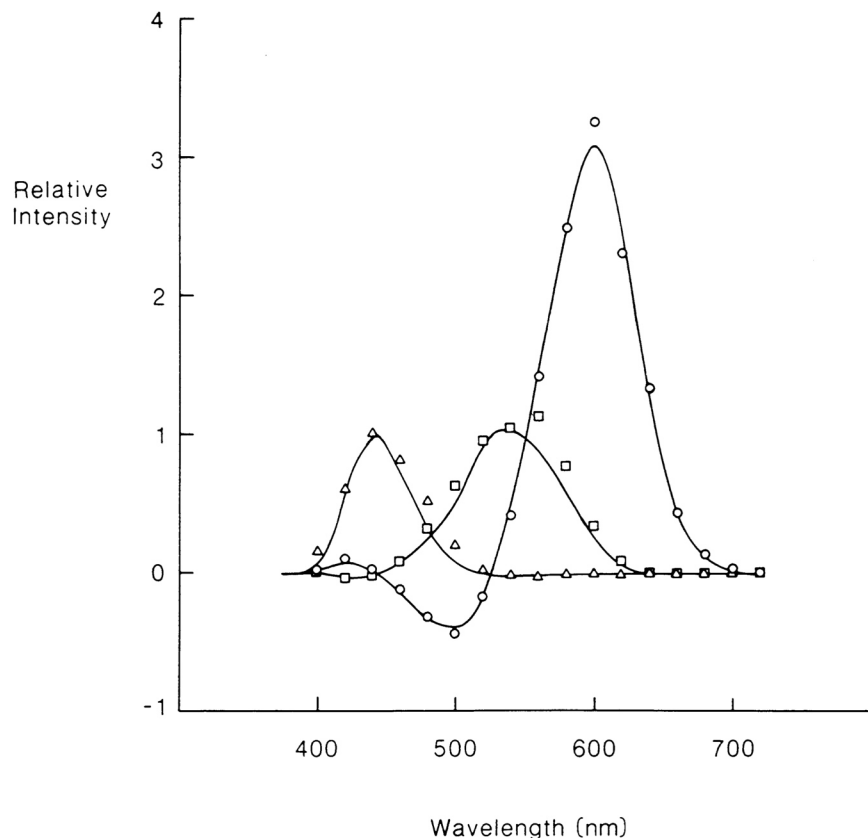


Figure 3

Color-matching functions. The smooth curves show the relative amount of light required of the three primaries to match a test light of unit intensity as a function of the wavelength of the test light (Stiles and Burch, 1955). The points plot the predicted results based on the spectral sensitivity functions of the macaque cones in Figure 1 after correction for absorption by the lens and macular pigment and photopigment self-screening. The correction factors used were: macular pigment density at 460 nm, 0.29; lens density at 400 nm, 1.22; and peak axial photopigment density, 0.27. the wavelengths of the primaries are 444 (Δ), 526 (□) and 645 (○) nm. SOURCE: Baylor et al. (1987).

Luminosity

Another reflection of the spectral sensitivity functions as seen in psychophysical experiments is luminosity. Luminosity describes the efficiency of light (incident on the cornea) in stimulating the visual system as a function of wavelength. In light-adapted conditions the photopic luminosity function is thought to reflect a linear sum of the excitation of the red and green cones. Under dark-adapted conditions the scotopic function is thought to depend on the excitation of the rods. These two functions are given by the smooth curves in Figure 4. The continuous line is the photopic function of Vos (1978). The open circles plot the weighted sum of the "corrected" spectral sensitivities of the red and green cones of the human retina (Figure 2); the corrections were the same as those applied in Figure 3. The corrected spectra were weighted so that the green spectrum contributed only about half as much (0.52) to the sum as did the red spectrum. This weighting was chosen to maximize the fit between the calculated and measured functions.

The scotopic luminosity function of Crawford (1949) is shown by the dashed curve, and the filled symbols plot the "corrected" spectral sensitivity function of the macaque rods (Figure 1). The correction factors here (given in Figure 4's caption) were chosen to give the best fit between the points and the curve. A good match was obtained between the two luminosity functions and the corresponding physiological results, consistent with the idea that the scotopic function is set by the filtered rod spectrum and that the photopic function is determined by the summed activity of the red and green cones. The scaling factor (0.52) weighting the contributions of the two cone spectra may reflect the relative numbers of the two cone types in the human eye or the relative strengths of their signals to higher-order neurons.

Bleaching Cross Section

The bleaching cross-sectional area, A , of a photopigment is a measure of the effective physical area over which a single pigment captures light. For steady light of intensity I , the product IA gives the expected rate of isomerization per pigment. This value has been previously estimated for rhodopsin by measuring, in the presence of steady bright light, the optical density of a dilute solution of rhodopsin as a function of time. The density falls exponential in time as the fraction of unbleached rhodopsin declines; the time constant of decay is $(IA)^{-1}$. The estimated values of A for rhodopsin (for light of optimal wavelength) were in the range of $7-10 \times 10^{-17} \text{ cm}^2$ (Dartnall, 1972).

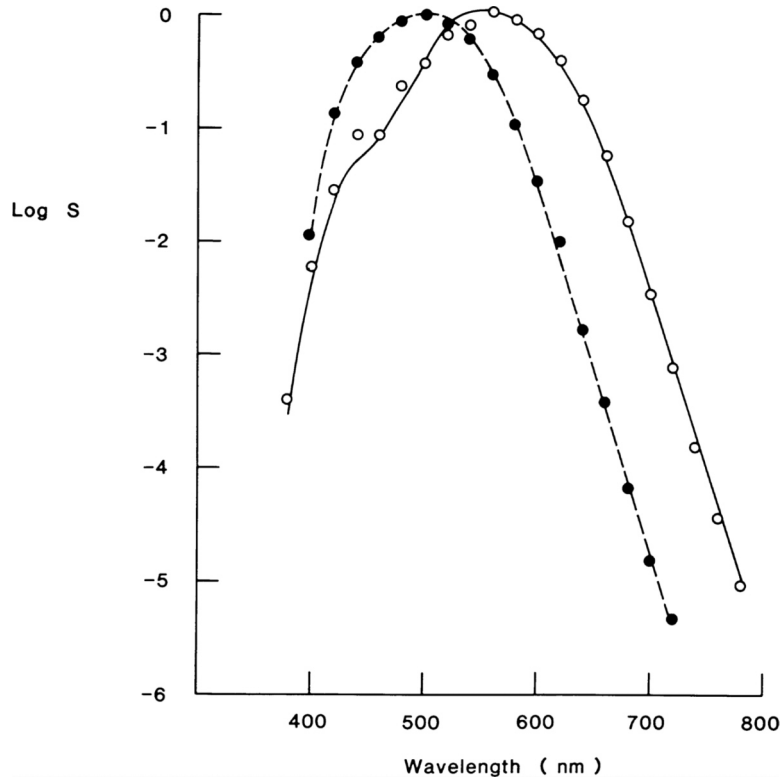


Figure 4
Luminosity functions and comparison to rod and cone spectra. The continuous curve is the photopic luminosity function of Vos (1978); the dashed curve is the scotopic luminosity function of Crawford (1949). The symbols plot the correlated spectral sensitivities of the macaque rods (●) and the weighted sum of the human red and green cones (○). The cone spectra were corrected as in Figure 3. The corrections for the rods were: lens density at 400 nm, 1.54; and peak axial photopigment density, 0.35.
SOURCE: Schnapf et al. (1988).

Similar density measurements for cone pigments are not available because of difficulties in pigment extraction. However, measurements of the photocurrent of single cones appear to provide an alternative method of estimating A . In the presence of steady bleaching light, the photocurrent was found to decline exponentially in time, with a time constant that was inversely related to I . These observations are consistent with the idea that the photocurrent is tracking the fraction of unbleached pigment in the cone outer segment. If so, A may be calculated from the intensity of the light and

the time constant of decay. The mean value of A calculated from several cones was found to lie within the range of that estimated for rhodopsin.

ACKNOWLEDGMENTS

This work was supported by grant EY05750 from the National Eye Institute and by an award from Research to Prevent Blindness, Inc.

References

- Baylor, D.A., B.J. Nunn, and J.L. Schnapf 1984 The photocurrent, noise and spectral sensitivity of rods of the monkey *Macaca fascicularis*. *Journal of Physiology* 357:575-607.
- 1987 Spectral sensitivity of cones of the monkey *Macaca fascicularis*. *Journal of Physiology* 390:145-160.
- Crawford, B.H. 1949 The scotopic visibility function. *Proceedings of the Physical Society* B62:321-334.
- Dartnall, H.J.A. 1972 Photosensitivity. Pp. 122-145 in *Photochemistry of Vision*, H.J.A. Dartnall, ed. New York: Springer-Verlag.
- DeValois, R., H.C. Morgan, M.C. Polson, W.R. Mead, and E.M. Hull 1974 Psychophysical studies on monkey vision--I. Macaque luminosity and color vision tests. *Vision Research* 14:53-67.
- Kraft, T.W. 1988 Transduction by cones in the golden-mantled ground squirrel, *Citellus lateralis*. *Biophysical Journal* 53:474a.
- MacLeod, D.I.A., and M.A. Webster 1983 Factors influencing the colour matches of normal observers. Pp. 81-92 in *Colour Vision*, J.D. Mollon and L.T. Sharpe, eds. London: Academic Press.
- Naka, K.I., and W.A.H. Rushton 1966 S-potentials from colour units in the retina of fish (*Cyprinidae*). *Journal of Physiology* 185:536-555.
- Neitz, J., and G.H. Jacobs 1986 Polymorphism of the long-wavelength cone in normal human colour vision. *Nature* 323:623-625.
- Schnapf, J.L., T.W. Kraft, and D.A. Baylor 1987 Spectral sensitivity of human cone photoreceptors. *Nature* 325:439-441.
- Schnapf, J.L., T.W. Kraft, B.J. Nunn, and D.A. Baylor 1988 Spectral sensitivity of primate photoreceptors. *Visual Neuroscience* 1:255-261.
- Stiles, W.S., and J.M. Burch 1955 Interim report to the Commission Internationale de l'Eclairage, Zurich, 1955, on the National Physical Laboratory's investigation of colour-matching (1955). *Optica Acta* 2:168-181.
- Vos, J.J. 1978 Colorimetric and photometric properties of a 2° fundamental observer. *Color Research and Applications* 3:125-128.
- Wyszecki, G., and W.S. Stiles 1982 *Color Science: Concepts and Methods, Quantitative Data and Formulae*. New York: Wiley.

TRANSDUCTION, ADAPTATION, AND DISEASE PROCESSES

The ERG and Sites and Mechanisms of Retinal Disease, Adaptation, and Development

Donald C. Hood

INTRODUCTION

What do ambient lights, retinal disease, and the development of the neonatal retina have in common? One answer is that each can alter the retina's sensitivity to light. A second is that the same noninvasive techniques have been used with human subjects to assess the sites and mechanisms of disease, adaptation, and developmental processes. One technique is the recording of the electroretinogram (ERG). This paper presents a general approach for using ERG data to assess the sites and mechanisms of a change in retinal processing. The approach is illustrated using data from an adaptation paradigm. Application to clinical and developmental questions is considered. Since the receptor is the focus of this symposium, the emphasis here is on testing the hypothesis that sensitivity changes that accompany a disease, an adaptation process, or a developing retina have their locus at the receptors.

The ERG is a gross potential recorded from the eye. [Figure 1A](#) shows ERGs recorded from a normal adult. Each record is the response of the eye to a flash of a different intensity. For higher flash intensities, the characteristic a- and b-waves of the ERG can be seen.

The ERG is a complex potential. The potential measured is actually the algebraic sum of a number of individual components. [Figure 2](#) shows an analysis of the ERG by Brown (1968). The analysis is similar to Granit's classic analysis (Granit, 1947). One of the two main potentials contributing to the ERG is a corneally negative potential generated by the receptors. This component was called p-III by Granit and is labeled the rod late receptor potential in [Figure 2](#). The two corneally positive potentials, labeled D.C. component and b-wave in [Figure 2](#), were shown

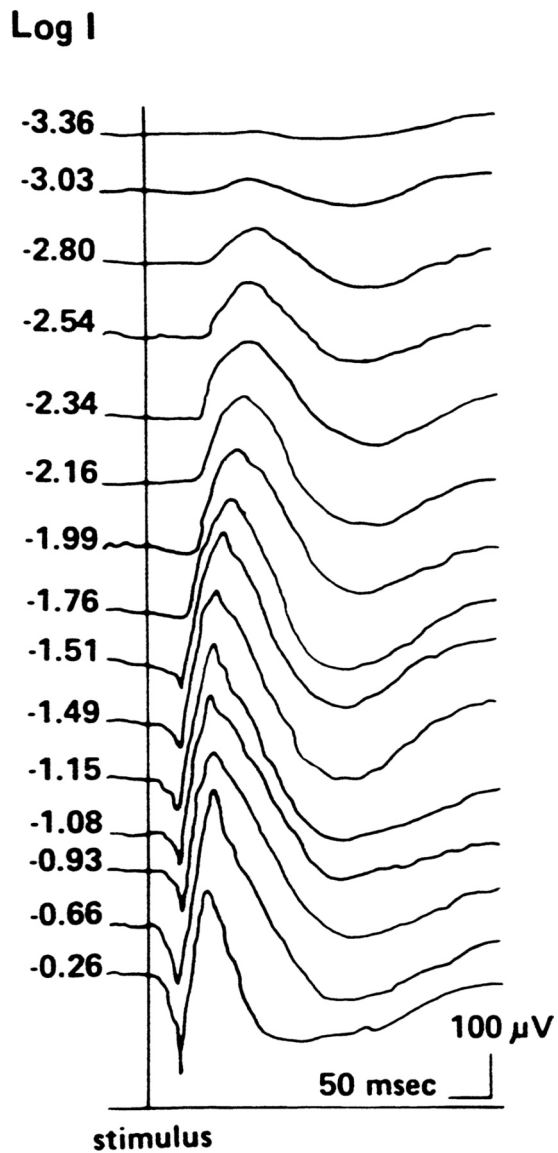


Figure 1A
ERG responses from a normal adult. Each record shows the response to a single flash of light. The log of the flash intensity ($\log \text{cd/m}^2$) is shown next to each record. SOURCE: Modified from Massof et al. (1984).

by Brown to be generated by cells in the inner nuclear layer. These potentials comprise Granit's p-II. The peak-to-trough amplitude of the b-wave provides a reasonably good measure of the magnitude of the potentials generated in the inner nuclear layer. Some investigators measure the a-wave to obtain information about the receptors. Use of the a-wave as a measure of the magnitude of the receptor response is more risky or rather requires additional assumptions.

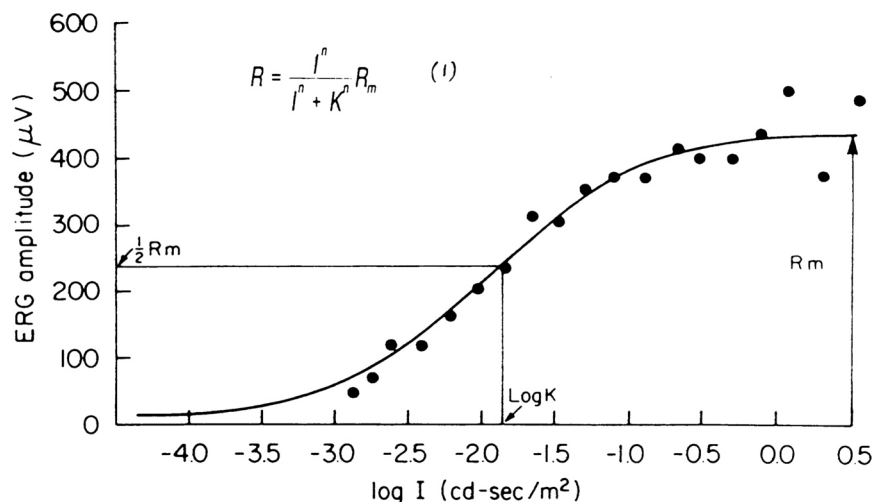


Figure 1B
 Intensity-response functions for the b-wave. The filled symbols are the peak-to-trough b-wave amplitudes plotted against the log intensity of the flash.
 SOURCE: Courtesy of M. Johnson.

The ERG was the focus of a great deal of research in the 1940s, 1950s, and 1960s. Many of the basic functions of the retina were inferred from this work. As a laboratory technique, it has been supplanted for most purposes by intracellular recording. In the clinic, however, the ERG still provides a powerful tool for diagnosing the type and progression of retinal disease. The development of the focal ERG and the use of quantitative analyses have enhanced clinical interest in the ERG. Figure 1B illustrates one quantitative approach. The filled symbols depict intensity-response data from recordings such as those in Figure 1A. The smooth curve through the data is given by Equation (1) in Figure 1, where R is the peak-to-trough b-wave amplitude, I is the flash intensity, K is the semisaturation constant, and R_m is the maximum response. The exponent n is usually close to 1. To simplify the presentation, it is assumed that n is equal to 1.0.

Retinal diseases produce both an increase in K and a decrease in R_m . Mary Johnson, Bob Massof, and I have been developing an approach to

inferring the retinal site and mechanism of disease action from changes in K and R_m (Johnson and Hood, 1988; Johnson and Massof, 1988). To evaluate this approach, I have analyzed adaptation data from Fulton and Rushton (1978). This analysis is summarized below.

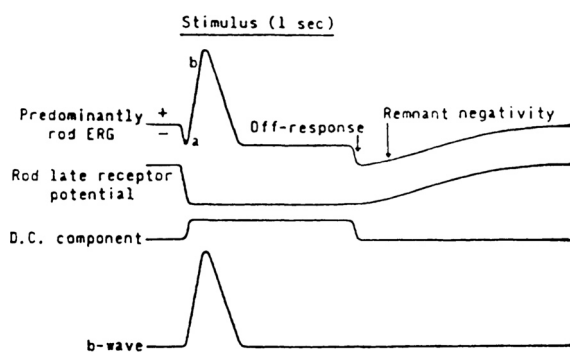


Figure 2

An analysis of the components of the dark-adapted rod ERG. Reprinted with permission from Vision Research, 8, Kenneth T. Brown, "The Electroretinogram: Its Components and Their Origin," © 1968, Pergamon Press.

INFERRING THE SITE OF ADAPTATION FROM ERG DATA

In the adaptation paradigm, intensity response data (as in Figure 1) are collected for flashes presented on steady ambient light. Fulton and Rushton (1978) obtained intensity-response data for b-waves recorded on steady adapting fields ranging in intensity from no field (dark adapted) to 3.2 log scotopic trolands. They fitted Equation (1), with n set equal to 1.0, to each set of intensity-response data.

To illustrate our approach, assume that Fulton and Rushton's data are well fitted by Equation (1) and that the rod system has been successfully isolated. As the adapting intensity is increased, their intensity-response curves move down and to the right when plotted as in Figure 1B. In terms of Equation (1), the maximum response, R_m , decreases and the value of the semisaturation constant, K , increases with increases in the intensity of the steady adapting field. In Figure 3 the change in $\log R_m$ is plotted against the change in $\log K$. Each data point is for a different adapting field intensity. The values of R_m and K are expressed relative to their dark-adapted values. The value 0 log relative K or R_m represents no change from the dark-adapted value (dashed lines). Data for the higher-adapting intensities were omitted to help assure that the rod system was isolated and to keep the range of sensitivity changes close to those seen below for retinal diseases.

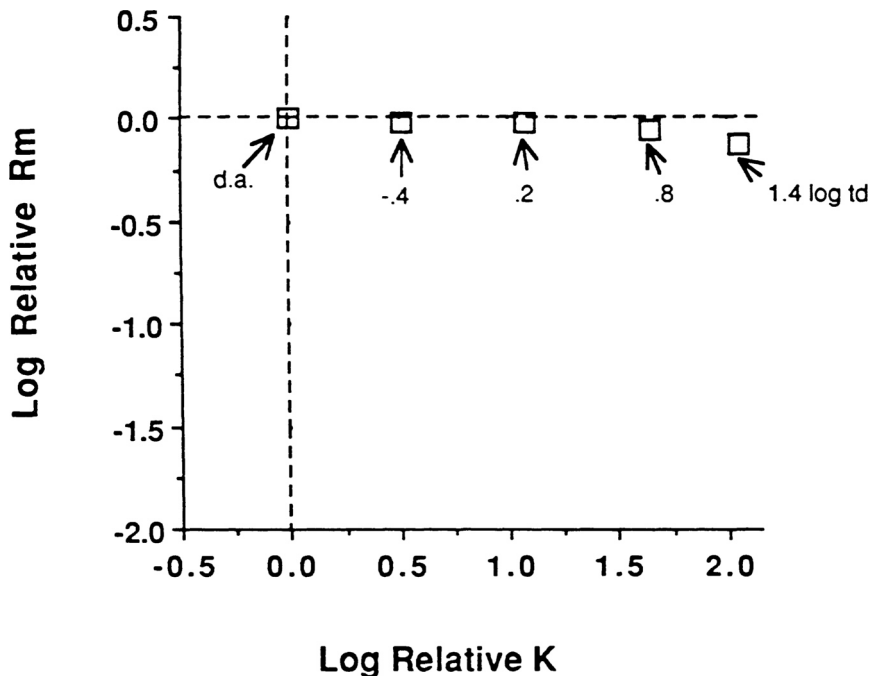


Figure 3

Log relative R_m versus log relative K is shown for a range of adapting intensities from the Fulton and Rushton (1978) study. The values of R_m and K , estimated from their fit of Equation (1), are expressed relative to the dark-adapted values.

A Model of the ERG

Ambient lights produce a small decrease in R_m and a large increase in K . How many of these changes can be attributed to the rod receptor and how many to the cells of the inner nuclear layer? A simplified version (Johnson and Hood, 1988) of the model we are developing is summarized in Table 1. We assume that the pooled rod receptor response, A , is given by Equation (2), where K_a is the semisaturation constant of the receptors and A_m is the maximum pooled response. Assume further that the response, B , of the b-wave generators in the inner nuclear layer is also described by a function of the same form, where the pooled receptor response A is the input to the b-wave generator [Equation (3)]. The parameters K_b and B_m are the semisaturation constant and maximum amplitude of the b-wave generators. Finally, we assume that the peak-to-trough b-wave amplitude measured is proportional to the response of the b-wave generators [Equation (4)]. After appropriate substitution and algebraic manipulation, Equation (5) shows

that R , the b-wave response one measures, has the form of Equation (1), where R_m and K are given by Equations (6) and (7), respectively.

TABLE 1 Static Model of ERG

Pooled response of rod receptors:

$$A = \frac{1^n}{1^n + K_a^n} A_m \quad (2)$$

Response of b-wave generators:

$$B = \frac{A}{A + K_b} B_m \quad (3)$$

ERG b-wave response:

$$R = \beta B \quad (4)$$

where β is a constant

by substitution

$$R = \frac{1^n}{1^n + K^n} R_m \quad (5)$$

where

$$R_m = \frac{\beta B_m}{1 + (K_b/A_m)} \quad (6)$$

$$K = \frac{K_a^n}{1 + (A_m/K_b)} \quad (7)$$

Notice that the semisaturation K and the maximum response R_m of the b-wave depends on three of the four parameters of the model. In other words, changes in the receptor's semisaturation constant, K_a , only affect K and changes in B_m only affect R_m . Both K and R_m are affected by the ratio of A_m to K_b . The dashed vertical and horizontal lines in Figure 4 show the predicted changes in K and R_m for increases in K_a (a decrease in receptor sensitivity) or a decrease in B_m (a decrease in the response amplitude of the cells generating the b-wave).

The semisaturation constant, K_b , of the b-wave generator and the maximum amplitude of the receptor potential, A_m , both contribute to the measured values of K and R_m [see Equations (6) and (7)]. To obtain a measure of the influence of A_m and K_b on K and R_m , we can set the ratio of A_m/K_b in Equations (6) and (7). A reasonable estimate of this

ratio is about $10^{2.5}$; it was obtained in the following way: The value of K typically measured for the b-wave is on the order of $10^{2.5}$ times smaller than the semisaturation of the receptors, K_a , measured for the a-wave (Fulton and Rushton, 1978; M.A. Johnson, personal communication) or inferred from human psychophysics (e.g., Sakitt, 1976; Hood and Greenstein, 1988a, 1988b). Substituting $(10^{-2.5}) \times K_a$ for K_2 in Equation (7) provides a value of the ratio K_b/A_m . It too will be approximately $10^{-2.5}$. By specifying the value of K_b/A_m , Equations (6) and (7) provide estimates for changes in R_m and K , with increases in K_a and decreases in A_m .

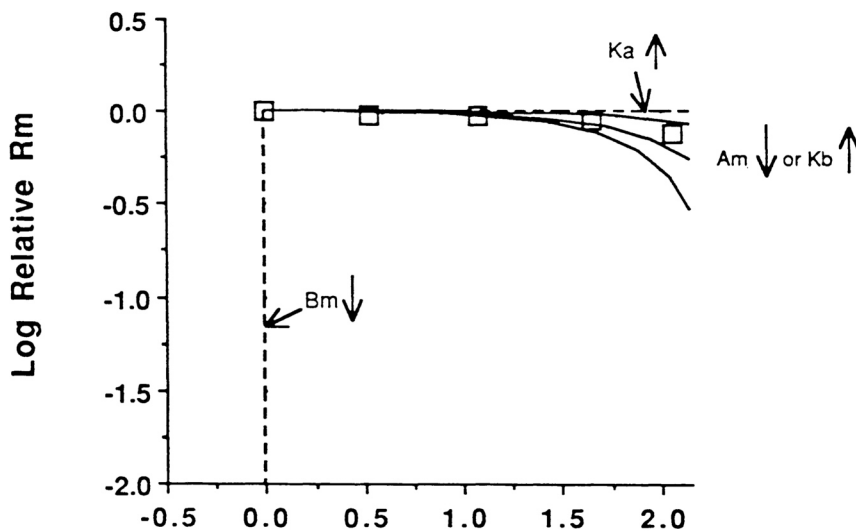


Figure 4
 The data from Figure 3 are shown with theoretical curves based on changes in the parameters of the model in Table 1.

The three solid curves in Figure 4 show the predicted changes in $\log K$ and $\log R_m$ for increases in K_b or decreases in A_m for three values of the ratio K_b/A_m . The middle curve was derived assuming a K/K_a ratio of $10^{-2.5}$. The two other solid curves are for K/K_a ratios of 10^{-3} (upper solid curve) and 10^{-2} (lower solid curve). Under these assumptions, an increase in K_b or a decrease in A_m largely affects K and has little effect on R_m . [For a quantitatively similar analysis of the ERG, see Arden et al. (1983). For a quantitatively similar analysis of psychophysical data, see Hood and Greenstein (1988b). For the general notion that the rods are linear over a wide range of conditions, see Rushton (1965).]

How do we answer the question: Are the changes in K and R_m produced by the steady adapting field occurring at the receptors? There are two approaches one can take to answer this question: (1) an explicit model of the receptors can be assumed or (2) the a-wave can be used to obtain estimates of K_a and A_m . Let us consider the first approach.

A Model of Adaptation of the Rod Receptor

The prevailing model of adaptation of the human rod receptor is response compression.¹ Response compression as a model of receptor adaptation was first described by Boynton and Whitten (1970). More recently, Baylor et al. (1984) have shown that this model fits the data from single primate rods. Figure 5 is a summary of a response compression model. The solid curve in the top panel is the dark-adapted intensity-response function. Adapting fields, shown by dashed lines, are assumed to produce responses as if they were flashes of light. The adapting-field-induced desensitization of the receptor occurs because of the restricted or compressed response range. When the total response is plotted versus the total intensity of the adapting field plus the flash (see Figure 5A), there is no change in K or R_m . In the Fulton and Rushton experiment, an incremental response, the ERG, was measured as a function of an incremental light, the flash. The bottom panel in Figure 5B shows the predictions of a response compression model in terms of incremental responses and incremental light intensities. For this incremental response function, response compression yields an increase in K and a decrease in R_m (Normann and Perlman, 1979).

If a particular value of the dark-adapted K_a is assumed, the response compression model provides predicted values of K_a and A_m as a function of adapting intensity. By substituting these values of K_a and A_m in Equations (6) and (7), the changes in R_m and K , relative to their dark-adapted values, can be derived. The solid curves in Figure 6 show the predicted changes in R_m (panel B) and K (panel A) assuming a response compression model and a dark-adapted value for K_a of 102.3 scotopic trolands (td). The open symbols, from Fulton and Rushton and Figure 3, are the changes in $\log K$ (panel A) and $\log R_m$ (panel B) plotted as a function of adapting field intensity. We can reject the receptor as the site of most of the sensitivity

¹ In the clinical literature, "response compression" is often used to signify a decrease in response amplitude without a change in K . Here we use the term to mean a decrease in responsiveness secondary to a decrease in the polarization of the receptor. The response compression model described in Figure 5 results in an equal change in ($\log K_a$) and ($-\log A_m$) from their dark-adapted values.

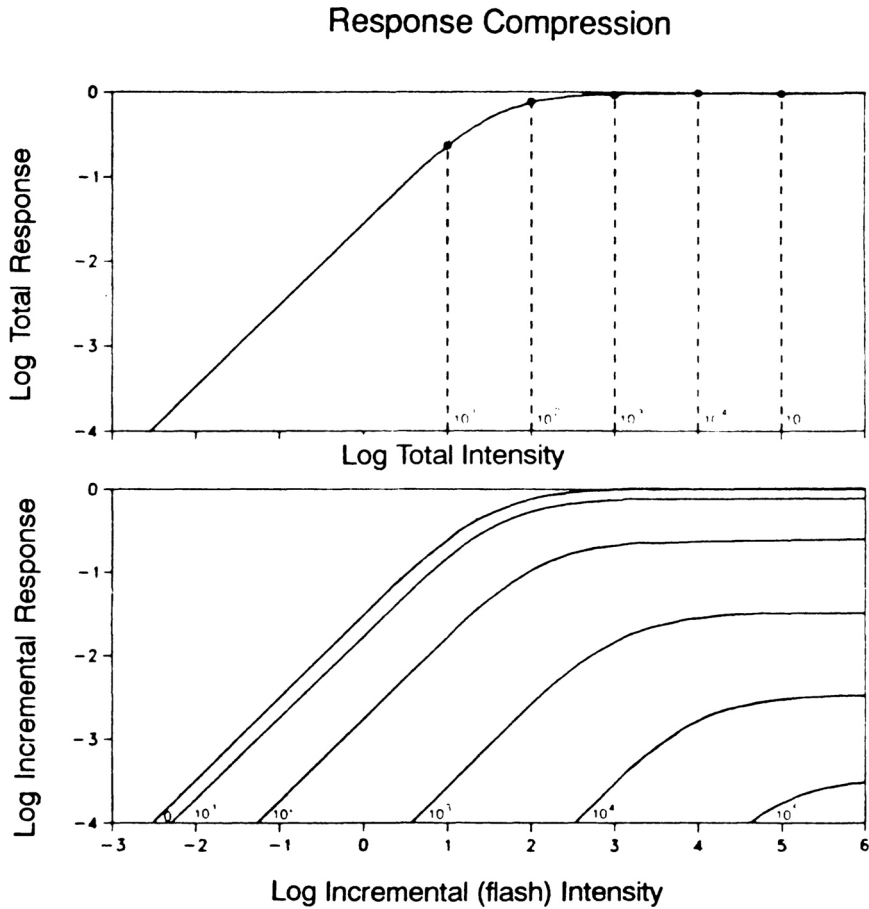


Figure 5

Response compression. The solid curve in the upper panel and the left-most curve in the lower panel are the dark-adapted intensity-response functions for a hypothetical receptor. The upper panel plots the total response amplitude versus the total light, flash plus adapting field, for adapting fields of different intensities (dotted lines). The lower panel shows the intensity-response functions in terms of the incremental response above the response of the adapting field versus the log of the incremental intensity above the intensity of the adapting field. SOURCE: Adapted from Hood and Finkelstein (1986).

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

changes occurring with light adaptation over the range of adapting intensities shown.² From this analysis the changes observed in the semisaturation constant, K , of the ERG can be largely attributed to an increase in K_b . The small change in R_m can be attributed to changes in B_m .³

Using A-Wave Data to Estimate K_a and A_m

As an alternative to assuming a model of receptor adaptation, we can assume that the a-wave provides a measure of the receptor's response. Fulton and Rushton (1978, their Figure 3a) measured intensity-response functions for the a-wave. The a-wave amplitude is not easily measured in all subjects or at all intensities. As a measure of the a-wave amplitude, Fulton and Rushton used the slope of the a-wave. If we assume that this measure is a good substitute for the rod receptor's response amplitude, then their intensity-response functions provide a measure of the change in K_a and A_m for each adapting intensity. Using Equations (6) and (7) and the Fulton and Rushton values of K_a and A_m for each adapting intensity, the predicted changes in K and R_m attributable to receptor changes were calculated. The small crosses in Figure 6 show the expected changes in $\log K$ and R_m based on the observed changes in K_a and A_m of the a-wave. The conclusions from this analysis are the same; most of the change in K is caused by an increase in K_b and the small change in R_m by a decrease in B_m . (The agreement between the Fulton and Rushton a-wave data and the response compression model of the receptors is better than it appears in Figure 6. If the model were fitted to the a-wave data, the solid curve in Figure 6A would be shifted vertically. Most of the variation from the model is caused by the dark-adapted data point.)

Assessing the Approach

With both procedures for estimating K_a and A_m , we reject the receptor as the primary site of sensitivity change for adapting fields up to 1.4 log scotopic trolands. In fact, we can conclude that the small changes in R_m are attributable to changes in B_m of the inner nuclear layer b-wave generators and that the change in $\log K$ is largely due to an increase in K_b (cellular adaptation in the inner nuclear layer). This is the conclusion we should have expected from previous physiological and psychophysical work (see Shapley and Enroth-Cugell, 1985, and Walraven et al., 1989, for reviews).

² This conclusion is restricted to the range of adapting intensities in Figure 6. At higher adapting field values, the changes in K_a and A_m make major contributions to changes in both K and R_m .

³ According to the model, the difference between the solid curve and the data provide measures of $\log K_b$ (panel A) and $\log B_m$ (panel B).

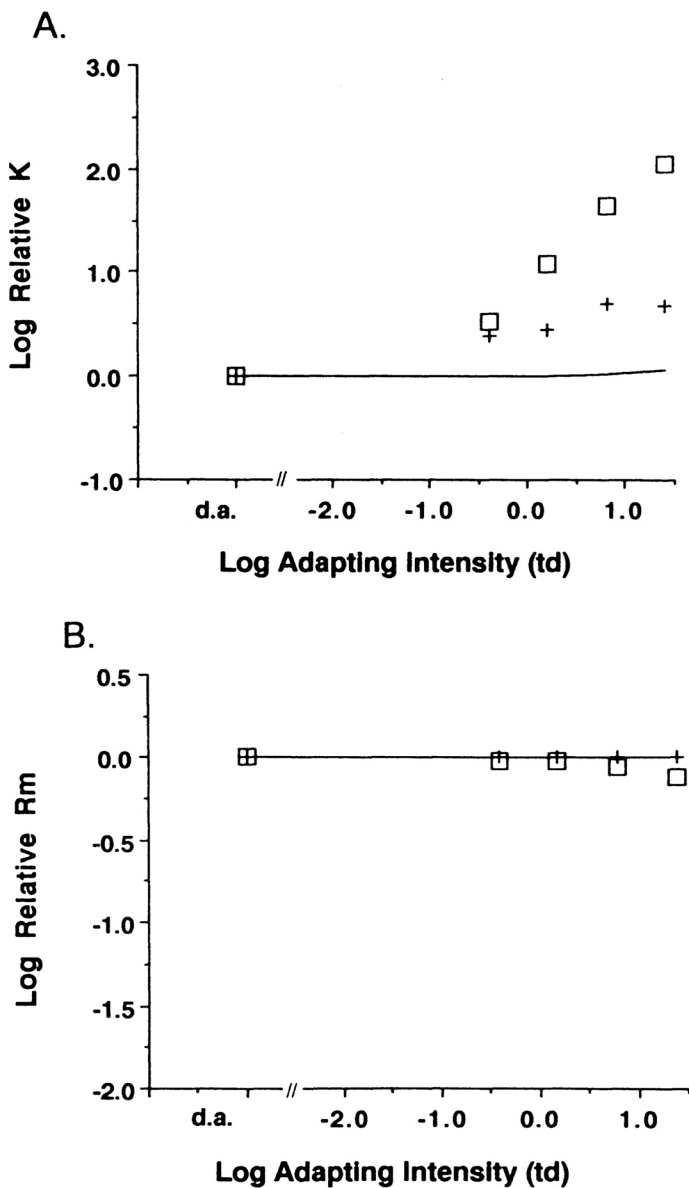


Figure 6

(A) Log relative R_m and (B) log relative K versus log adapting field intensities. The values 0 log K or R_m represents a dark-adapted value. The open symbols are from the Fulton and Rushton (1978) study. The solid curves are the predicted changes in log K and log R_m based on the response compression model of the rod receptors. The crosses show the predicted change in log K and log R_m derived from the Fulton and Rushton a-wave data and Equations (6) and (7).

The main purpose here, however, was not to test hypotheses about rod system adaptation, but rather to evaluate an approach to inferring sites and mechanisms of disease-related sensitivity changes from ERG data. If the results of this analysis turned out to be inconsistent with previous work, the ERG would not have been the topic of this paper. The results are, however, encouraging. Let us consider the application of the approach to the problem of retinal disease.

RETINAL DISEASE AND ERG CHANGES

Various retinal diseases have been shown to increase $\log K$ and decrease $\log R_m$ of the b-wave. Figure 7 shows data from three clinical studies. Each data point represents the $\log R_m$ and $\log K$ for a single eye from a single patient. Open symbols are from two studies (Birch and Fish, 1987; Massof et al., 1984) measuring ERGs of patients with retinitis pigmentosa (RP), and the filled symbols are from a study examining the ERGs of patients with central vein occlusion (Johnson and Hood, 1988). Both diseases decrease the maximum b-wave response and increase the value of K . However, the patterns of the results differ substantially. The patients with central vein occlusion show relatively larger changes in K than the patients with RP.

To determine whether the changes in K and R_m are due to receptor changes, the approach described above can be applied. We can assume a model of receptor change and ask if it describes the changes in K and R_m , or we can use a-wave recordings to obtain estimates of K_a and A_m . RP will be used to illustrate the first approach and central vein occlusion the second.

Consider RP (filled symbols) first. RP is an inherited disease that affects the pigment epithelium and thus the receptors. A loss in visual pigment has been documented with both retinal densitometry and psychophysical matching techniques. Histopathological studies show outer segments that have been shortened and receptors that are twisted and distorted. Ultimately the disease results in the loss of receptor cells. Table 2 lists seven hypothesized actions of RP on the receptors (labeled 2 through 8). The other columns show the effects on the model's parameters (column 2) and the predicted changes in K and R_m (column 3). Notice that except for hypothesis 8 the receptor hypotheses translate into changes in K_a and A_m . And these changes in K_a and A_m result largely in changes in K . Based on these hypotheses, there should be little or no change in R_m .

Are the changes in K and R_m produced by RP occurring at the receptors? Suppose we assume an explicit model of the receptor. If we assume a value of K_a between 1.8 and 2.8 log scotopic trolands, the curves labeled increased K_a and decreased A_m in Figure 4 show the range of

predicted changes in R_m and K for hypotheses 2 through 7 in Table 2. The envelope of these curves is reproduced in Figure 7. We cannot reject the receptor as the major cause of the change in log K .

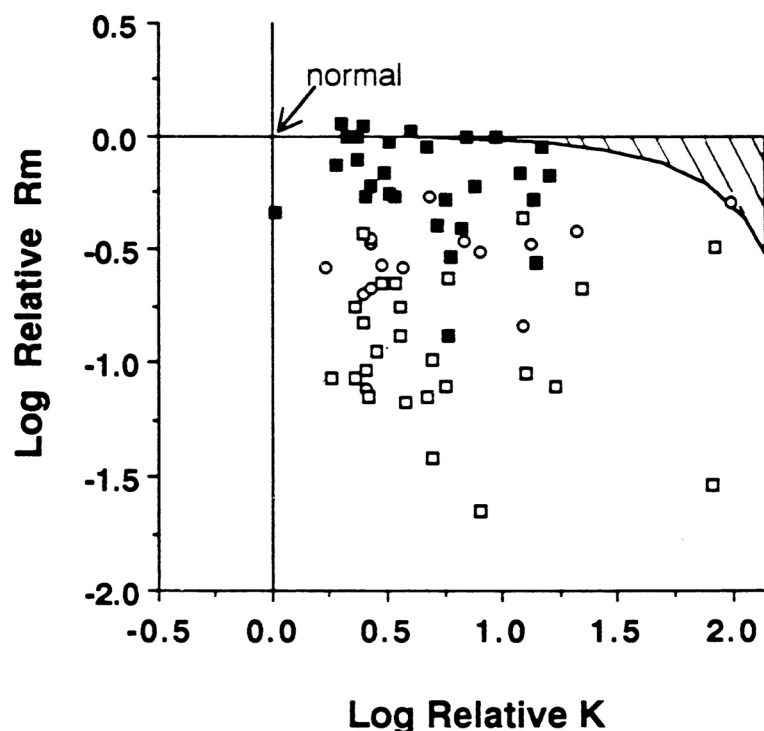


Figure 7

Log relative R_m versus log relative K for ERG intensity-response functions recorded from patients with retinal disease. Each data point is for a single eye of a single patient. The filled symbols represent data from patients with central vein occlusion (Johnson and Massof, 1988). The open symbols represent data from RP patients from two studies (Birch and Fish, 1987—squares; Massof et al., 1984—triangles). The R_m and K values are divided by mean normal data for the RP patients and by the parameters for the unaffected eye for the central vein occlusion patients. The 0.0 value represents normal mean value or the value for the normal eye.

Can we, however, reject the receptor as the only cause of the changes in the ERG with RP? Hypotheses 1 through 7 can be rejected, as each predicts little or no change in log R_m . A loss of visual pigment, shortened or distorted receptors, or a decrease in the responsiveness of the receptors cannot account for the ERG results. Only one receptor hypothesis is capable of explaining the large decrease in log R_m . Hypothesis 8 in Table 2, if coupled with one or more of the other receptor hypotheses, could in theory account for the changes in both K and R_m . Large patches of receptor

loss could lead to a loss of b-wave generators and a decrease in both B_m and R_m . However, the rod visual fields in these patients are probably too large to produce a sufficient decrease in B_m based simply on a loss of patches of receptors⁴ (Birch et al., 1987). Although additional analyses of field data must be done and electrical shunting must be considered, it is likely that the ERG data will be consistent with recent psychophysical data (Greenstein and Hood, 1986; Hood and Greenstein, 1988b) in supporting an inner nuclear component to the sensitivity loss observed in RP.

TABLE 2 Hypothesized Disease Action and Predicted ERG Changes

Disease Action	Change in Model	Predicted b-Wave K	R_m
Prereceptor			
1. Preretinal screening (e.g., cataract, increase in macular pigment, blood)	$\uparrow K_a$	\uparrow	normal (n)
Receptor			
2. Decreased quantal catch (loss of pigment)	$\uparrow K_a$	\uparrow	n
3. Decreased quantal catch (misaligned or misshapened receptors)	$\uparrow K_a$	\uparrow	n
4. Shortening of outer segments	$\uparrow K_a \downarrow A_m?$	\uparrow	$\cong n$
5. Random loss of receptors	$\downarrow A_m$	\uparrow	$\cong n$
6. Increased membrane potential (response compression)	$\downarrow A_m = \uparrow K_a$	\uparrow	$\cong n$
7. Decrease in response amplitude	$\downarrow A_m$	\uparrow	$\cong n$
8. Large regional loss of receptors	$\downarrow B_m$	n	\downarrow

⁴ To generate predictions for a regional or patchy loss of receptors, a spatial dimension must be added to the model in Table 1. Assume that there are a finite number of b-wave generators and that the response of each is given by

$$B_i = \frac{A_i}{A_i + K_b} B_m$$

for all i , where A_i is the pooled response of the receptors feeding into the i th b-wave generator. Note that we are assuming that all b-wave generators have the same K_b and B_m . Assuming an equal contribution of all b-wave generators to the b-wave recorded at the electrode, Equation (4) becomes

$$R = \beta \sum_i \beta_i$$

A patchy loss of receptors removes a subset of b-wave generators by removing their receptor pools. A loss of all the receptors contributing to A_i eliminates the response of B_i . Under these assumptions a regional loss of one-half of the visual field, by itself, would reduce $\log R_m$ by 0.3 log unit.

The open symbols in [Figure 7](#) represent data from patients with central vein occlusion. In this disease the blood flow to the inner nuclear layer of one eye is disrupted. The changes in K and R_m of this eye, relative to the normal eye, are shown by the open square. Ophthalmologists must determine which of these patients will go on to develop neovascularization of the iris. Neovascularization not only leads to further loss of vision but can also result in the loss of the eye. Johnson et al. (1988) have shown that changes in the ERG, especially $\log K$, are four times more sensitive than fluorescein angiography in identifying patients at risk.

Are the large increases in $\log K$ due to receptor changes (K_a or A_m) or to changes in the inner nuclear layer (K_b)? Johnson and Massof (1988) have analyzed a-wave data and found that the increase in $\log K$ is largely due to an increase in K_a ; that is, the increase in the semisaturation constant of the b-wave in these patients is largely due to an increase in the semisaturation constant of the receptors. It appears that changes in receptor sensitivity may be a good predictor of which patients will develop neovascularization. This is particularly interesting since central vein occlusion disrupts blood flow to the inner nuclear layer while leaving the choroidal blood supply of the receptors intact. Disrupting the blood flow to the inner nuclear layer appears to compromise receptor responsiveness.

DEVELOPMENT AND ERG CHANGES

Over the first 6 months or more of life the most fundamental of visual capabilities, the detection of light, continues to improve. Fulton and Hansen (1982) have measured ERG intensity-response functions of neonates. The changes they measured in $\log R_m$ and $\log K$ for infants ranging in age from 1.5 to 12 months are shown in [Figure 8](#) (filled symbols). The values of R_m and K are expressed relative to the mean for a group of adults. The adaptation data (open squares) from [Figure 3](#) are shown for comparison. The developmental data show larger changes in R_m for equivalent changes in sensitivity than do the adaptation data. The theoretical curves are the same as those in [Figure 4](#).

From these data and the analysis above, we can reject the receptor as the only cause for the developmental changes. Since the changes in R_m are largely attributable to an increase in the response amplitude of the inner nuclear layer cells, these ERG data suggest some postnatal development of the inner nuclear layer. The nature of the postreceptor development can be further constrained by testing explicit hypotheses about the development of the inner nuclear layer and by analyzing behavioral data (Hood, 1988).

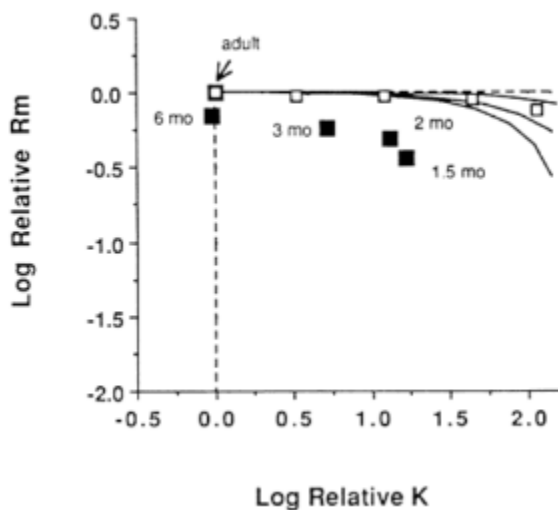


Figure 8

Log relative R_m versus log relative K for EGR intensity-response functions recorded from infants. The filled symbols are for infants 1.5 to 12 months old and for adults (from Fulton and Hansen, 1982). The open symbols represent the adult adaptation data from Figure 3. The smooth theoretical curves are the same as in Figure 4.

SUMMARY

The ERG is a complex potential; we should be cautious in interpreting changes it brought about by disease or development. However, carefully measured intensity-response functions combined with explicit models of the ERG and with data from psychophysical paradigms (e.g., visual fields, matching, increment threshold) provide the possibility of identifying sites and mechanisms of disease action for specific diseases and individual patients.

ACKNOWLEDGMENTS

This research was supported by National Eye Institute grant EY-02115 to D.C. Hood and V. Greenstein. I appreciate the helpful comments

provided by M. Johnson, V. Greenstein, and D. Birch on an earlier version of the manuscript and am particularly grateful to M. Johnson for her tutorials and for reacquainting me with the ERG.

References

- Arden, G.B., R.M. Carter, C.R. Hogg, D.F. Powell, W.J.K. Ernst, G.M. Clover, A.L. Lyness, and M.P. Quinlan 1983 A modified ERG technique and the results obtained in x-linked retinitis pigmentosa. *British Journal of Ophthalmology* 67:419-430.
- Baylor, D.A., B.J. Nunn, and J.L. Schnapf 1984 The photocurrent, noise, and spectral sensitivity of rods of the monkey *Macaca fascicularis*. *Journal of Physiology* 357:575-607.
- Birch, D.G., and G.E. Fish 1987 Rod ERGs in retinitis pigmentosa and cone-rod degeneration. *Investigative Ophthalmology and Visual Science* 28:140-150.
- Birch, D.G., W.K. Herman, J.M. deFaller, D.T. Disbrow, and E.E. Birch 1987 The relationship between rod perimetric thresholds and full-field ERGs in retinitis pigmentosa. *Investigative Ophthalmology and Visual Science* 28:954-965.
- Boynton, R.M., and D.N. Whitten 1970 Visual adaptation on monkey cones: recordings of late receptor potentials. *Science* 170:1423-1426.
- Brown, K.T. 1968 The electroretinogram: its components and their origin. *Vision Research* 8:633-678.
- Fulton, A.B., and R.M. Hansen 1982 Background adaptation in human infants. *Documenta Ophthalmologica Proceeding Series* 31:191-197.
- Fulton, A.B., and W.A.H. Rushton 1978 The human rod ERG. Correlation with psychophysical responses in light and dark adaptation. *Vision Research* 18:793-800.
- Granit, R. 1947 *Sensory Mechanisms of the Retina*. London: Oxford University Press.
- Greenstein, V.C., and D.C. Hood 1986 Test of the decreased responsiveness hypothesis in retinitis pigmentosa. *American Journal of Optometry and Physiological Optics* 63:22-27.
- Hood, D.C. 1988 Testing hypotheses about development with ERG and incremental threshold data. *Journal of the Optical Society of America* 5:2159-2165.
- Hood, D.C., and M.A. Finkelstein 1986 Visual sensitivity. In *Handbook of Perception and Human Performance*, vol. 1, K. Boff, L. Kaufman, and J. Thomas, eds. New York: Wiley.
- Hood, D.C., and V.C. Greenstein 1988a Blues (s) cone vulnerability: a test of a fragile receptor hypothesis. *Applied Optics* 27:1025-1029.
- 1988b Increment threshold (tvi) data and the site of disease action. *OSA Technical Digest* 3:2-5.

- Johnson, M.A., and D.C. Hood 1988 A theoretical interpretation of ERG abnormalities in central retinal vein occlusion. *OSA Technical Digest* 3:84-87.
- Johnson, M.A., and R.W. Massof 1988 Photoreceptor sensitivity loss in patients with central retinal vein occlusion and iris neovascularization. *Investigative Ophthalmology and Visual Science* (Suppl.):179.
- Johnson, M.A., S. Marcus, M.J. Elman, and T.J. McPhee 1988 Electroretinographic abnormalities associated with neovascularization in central occlusion. *Archives of Ophthalmology* 75:513-517.
- Massof, R.W., L. Wu, D. Finkelstein, C. Perry, S.O. Stair, and M.A. Johnson 1984 Properties of electroretinographic intensity-response functions in retinitis pigmentosa. *Documenta Ophthalmologica* 57:279-296.
- Normann, R.A., and I. Perlman 1979 Evaluating sensitivity changing mechanisms in light-adapted photoreceptors. *Vision Research* 19:391-394.
- Rushton, W.A.H. 1965 The Ferrier lecture: visual adaptation. *Proceedings of the Royal Society of London B* 162:20-46.
- Sakitt, B. 1976 Psychophysical correlates of photoreceptor activity. *Vision Research* 16:129-140.
- Shapley, R., and C. Enroth-Cugell 1985 Visual adaptation and retinal gain controls. *Progress in Retinal Research* 3:263-346.
- Walraven, J., C. Enroth-Cugell, D.C. Hood, D.I.A. MacLeod, and J. Schnapf 1989 The control of visual sensitivity: receptor and postreceptor processes. Ch 5 in *Visual Perception: The Neurophysiological Foundations*, L. Spillman and J. Werner, eds. New York: Springer-Verlag.

Phototransduction in Vertebrate Rods: The Electrophysiological Approach to the cGMP Cascade Theory

Edward N. Pugh, Jr., W.H. Cobbs, and J.W. Tanner

This paper reviews recent developments in the field of phototransduction, with a special focus on electrophysiological work that constrains and informs the transduction theory now called the cyclic GMP cascade theory (for other recent reviews, see Lamb, 1986; Pugh and Cobbs, 1986; Stryer, 1986; Yau et al., 1986; Pugh and Miller, 1987).

INTRODUCTION

The Nature of Phototransduction

Phototransduction may be defined as the sequence of events that transpire in a photoreceptor from the absorption of a photon to the production of a physiologically significant electrical signal. The latter ideally would be specified as a light-induced membrane polarization at the synaptic region sufficient to alter transmitter release to a degree detectable by a subjacent bipolar cell, but practically can be equally well specified as a light-induced change in cell membrane potential or even in outer-segment membrane current.

Research in phototransduction took an exciting turn in 1970 with the publication of Hagins et al.'s (1970) landmark paper on the dark currents and photocurrents of rat rods. [Figure 1](#) summarizes their finding, overlaying the theoretical interpretation of their data on a scanning electron micrograph of a frog retina. In the dark there is a continually flowing, net outer-segment-inward membrane current, whose spatially integrated magnitude is about 50 to 70 pA in the 25- μ m outer-segment rat rod at 37°C (rat rods are essentially identical to human rods) and is about 40 to 50 pA in the 50- μ m frog or toad rod. A little appreciation of the magnitude

of the dark current can be had from considering some numbers: since 1 pA of current corresponds to 6.25×10^6 monovalent charges/second, a 50-pA dark current flowing into a human rod outer segment with a water volume of c. 100 fl demands a complete turnover of the ions carrying the current every 5 seconds (50 pC of monovalent charge into 100 fl corresponds to 5 mM of the charge carrier; we can assume that the internal $c[\text{Na}]$ is no higher than 50 mM). For the rod membrane potential to be at steady state in the face of the inward current, there must be a balancing outward current in the inner segment. General principles of cellular physiology, ion substitution experiments, and other work lead to the conclusions that

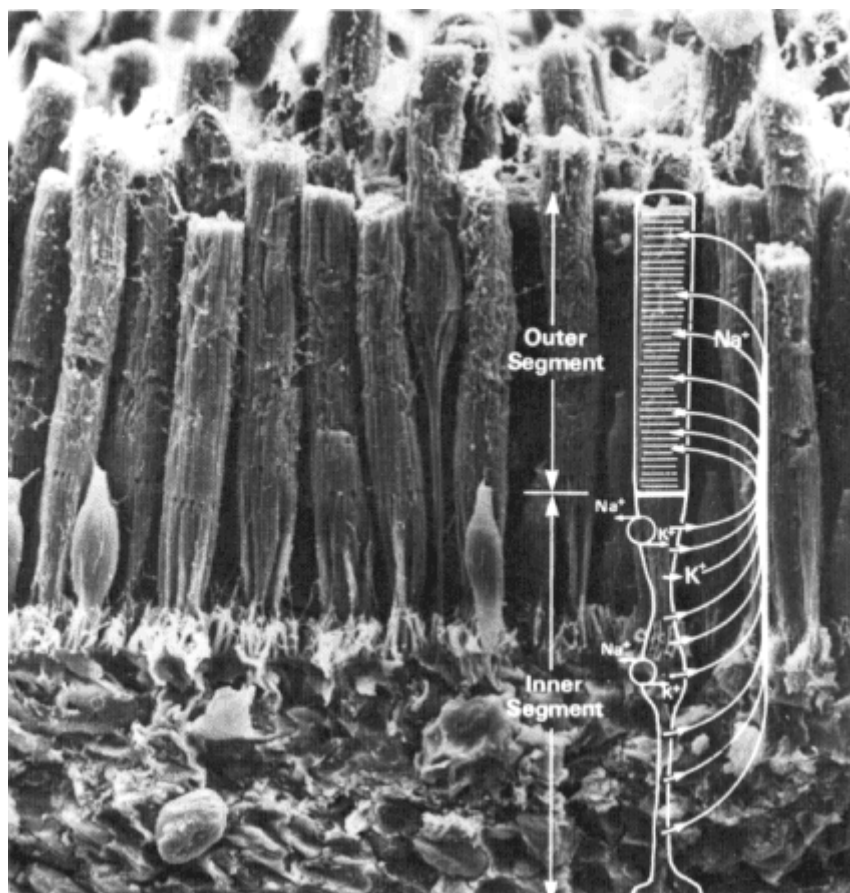


Figure 1
A scanning electron micrograph of the photoreceptor layer of the frog retina on which has been overlaid a schematic of the "dark" or "circulating" current of a rod and the ionic species that carry the current.
SOURCE: Courtesy of W.H. Miller.

(1) the outer-segment inward current is carried primarily by Na^+ , (2) the balancing inner-segment outward current is carried primarily by K^+ , and (3) there exists a potent ATP-requiring Na/K exchange that maintains the dark ionic balance. The latter exchange mechanism has been localized by Stirling and Lee (1980) to be primarily in the ellipsoid region and on the calices of inner-segment membrane that climb up the inner part of the outer segment.

Hagins et al. (1970) showed that transverse flashes of light delivered to the outer-segment layer produce local membrane currents—photocurrents—and hypothesized that the photocurrent was merely the suppression of the dark current. Penn and Hagins (1972) showed that the maximal photocurrent was in fact the complete suppression of the dark current. That the rod should hyperpolarize to light follows immediately from the fact that light suppresses the outer-segment dark current: since that latter current is a depolarizing current, its suppression (in the absence of compensatory reaction) must necessarily drive the membrane potential toward the reversal potential of any K^+ conductance. Thus, by 1970 the nature of the initial electrophysiological event in phototransduction was made crystal clear: it is *the local suppression in the outer segment of inward membrane current*.

Two Fundamental Constraints

We would now like to emphasize two additional hard inferences drawn in the early 1970s, inferences that drove research on the theory of phototransduction through much of the 1970s and early 1980s. The first was that there must exist a specific conductance in the outer-segment membrane that carries the dark current and that is closed by light—the *light-sensitive conductance*, or g_v . Hagins et al.'s (1970) work and subsequent work with the suction electrode (Baylor et al., 1979; Lamb et al., 1981) showed g_{hv} to be distributed more or less uniformly throughout the outer-segment membrane and also showed that the effect of a single photon on the conductance was highly localized, perhaps to 2 to 3 μm . A second hard inference that propelled transduction research was that there must exist an *internal messenger* in rods.

Light-Sensitive Conductance

The inference that there must be a specific light-sensitive conductance motivated many "fingerprint" analyses: experiments for characterizing this entity operationally. Chief among these are current/voltage relations, noise power spectra, and ionic permeation experiments. Biophysical characterization might be said to have begun with Bader et al.'s (1979) effort to

determine the current/voltage (I/V) relation of the conductance; the relation was definitively determined in the elegant triple electrode study of Baylor and Nunn (1986). One striking feature of this I/V relation is its near-zero slope over the entire physiological range of voltages, -25 mV to -75 mV. As Baylor has pointed out, this unusual I/V relation is ideal for producing virtually no cable losses in the outer segment, thus allowing the full effect of a local dark-current suppression to be converted into a voltage change in the inner-segment membrane. Another basic characterization of g_{hv} was carried out by Bodoia and Detwiler (1984), who measured noise power spectra that could be attributed to noise produced by the conductance per se. Baylor et al. (1980) had previously characterized two other forms of outer-segment noise that manifest themselves through the light-sensitive conductance: (1) "shot noise" events that can, with virtual certainty, be attributed to the thermal isomerization of rhodopsin and (2) a broader band noise that corresponds to some as yet unspecified transducer process.

Internal Messenger Hypothesis

The second inference that drove much transduction work was the need for an internal messenger, required because of several different kinds of evidence that the rod disks (which one could presume are the site of most photon absorptions, based on the action spectrum of night vision) are physically separated from the outer-segment plasma membrane by about the distance of a synaptic cleft. In 1970 Yoshikami and Hagins, in analogy with other systems such as muscle, proposed the divalent cation Ca^{2+} to be the transmitter. About the same time Bitensky et al. (1971), in analogy with Sutherland's fundamental insight, proposed that a cyclic nucleotide was the messenger. It is not our goal to review the incredible burgeoning field of cyclic nucleotide research that followed the Miller and Bitensky hypothesis, nor even a fraction of the elegant work that went into developing tests of the calcium hypothesis (for recent reviews, see *Physiological Reviews*, 1987). Rather, we would like simply to underscore the essence of the messenger hypothesis: light-driven changes in its concentration were hypothesized to cause g_{hv} to close rapidly.

Calcium as Internal Messenger

One line of evidence that strongly argued against Ca^{2+} playing the role of internal messenger was developed by Lamb and colleagues (Matthews et al., 1985; Lamb et al., 1986). Using a combined suction electrode and whole-cell gigaseal electrode (such as illustrated in [Figure 2](#)), they infused the calcium buffer/chelator BAPTA into rods and demonstrated that the

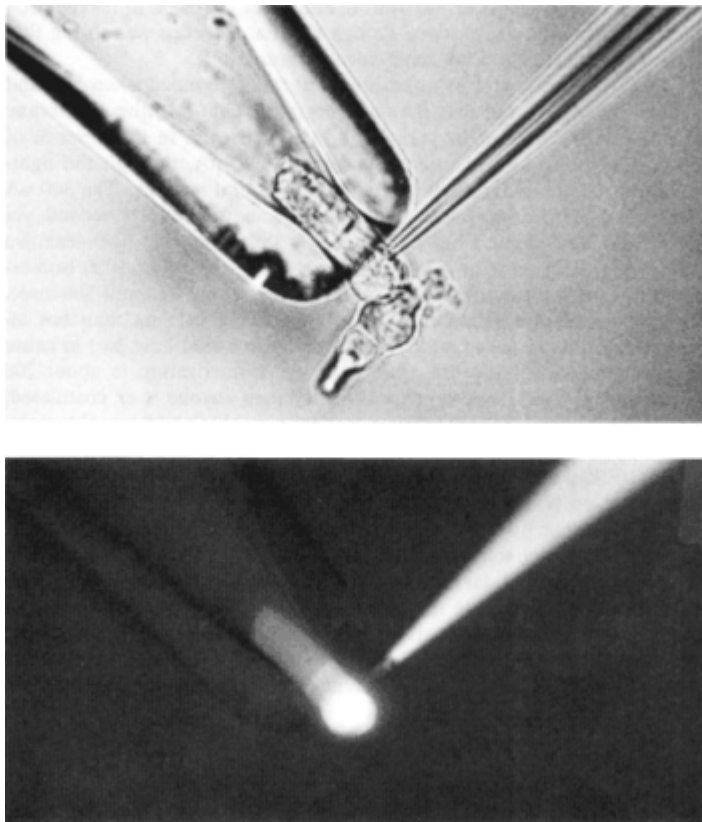


Figure 2

A pair of photographs of a salamander rod whose outer segment is held in a suction pipette and whose inner segment has been impaled with a gigaseal electrode; a cone was fortuitously attached to the rod in the inner-segment region. The top picture was taken with transmitted light, the bottom picture with epifluorescence. The cell was initially drawn into the suction pipette under deep infrared illumination. It was then recorded from for approximately 1 min after intracellular access was gained with the gigaseal electrode (as in Figure 3), which contained, in addition to an intracellular medium, 5 mM 5(6)-carboxyfluorescein. The epifluorescence light was then turned on, and the picture on the right was taken; finally, the transmission photograph was made. This experiment demonstrates that substance infused into the inner segment penetrates to the outer segment. The relative differences in intensity of fluorescence of the outer and inner segments probably reflect the differences in the water space—in the outer segment more than half of the volume is occupied by the disks.

initial phase of the photocurrent was not altered by the chelator (as required by the calcium hypothesis), even though internal evidence was strong that much of the chelator did not have calcium bound to it.

The Ca^{2+} messenger hypothesis was mortally wounded when Yau and Nakatani (1984) showed that the outer-segment, light-sensitive membrane current could be carried by *pure* Ca^{2+} , and that even in the absence of voltage control an inward calcium current of 300 pA, through the light-sensitive conductance, could be sustained for several seconds. The 300 pA of calcium current corresponds to an influx of about 10^9 Ca^{2+} /second, yet even so required about 5 sec to completely shut down the light-sensitive conductance. In contrast, a 200-msec flash delivering about 1000 isomerizations during the calcium influx shut down the current in about 100 msec. A simple calculation shows that, to produce more calcium than has already come in through the dark current, the flash would have had to cause the release of more than 10^6 Ca^{2+} ions per isomerization in about 200 msec, a much greater release stoichiometry than anyone ever postulated; indeed, the maximal reliable releases by light ever obtained from disk membrane suspensions were on the order of a few calcium ions per isomerized rhodopsin! Yau and Nakatani's (1984) experiments also demonstrated the existence of an Na/Ca exchange in the outer-segment plasma membrane, a mechanism that uses the energy in the Na gradient to evict Ca^{2+} . In a subsequent experiment, Yau and Nakatani (1985) showed that about 5 percent of the normal dark current is carried by calcium moving together with sodium; thus, Ca^{2+} in the outer segment actually *declines* during the light response. This interpretation was directly confirmed by McNaughton et al. (1986), who administered the coup to the hypothesis that an elevation in internal calcium in the outer segment carries the message of excitation between disk and plasma membrane.

Cyclic GMP as Internal Messenger

Bill Miller and Grant Nicol (Nicol and Miller, 1978; Miller and Nicol, 1979) were responsible for the beginnings of the electrophysiological testing of the cGMP-transmitter hypothesis: they discovered that injection of cGMP into outer segments depolarizes rods in a magnitude and with a duration that is dose dependent and that light antagonizes the depolarizing effect of cGMP. Miller (1982) extended the work with microelectrode potentiometric recording and presented a series of detailed phenomena consistent with the hypothesis that cGMP acts to open an outer-segment conductance that is shut by light. Consistent with the Miller-Bitsensky hypothesis, MacLeish et al. (1984) showed that cGMP infused into rods induced an increased inward current that was reversed by light. At about the same time, we (Cobbs and Pugh, 1985; Matthews et al., 1985) developed

a combined suction electrode and intracellular electrode salamander rod preparation like that used by Baylor and Nunn (1986), incorporating a gigaseal electrode rather than a microelectrode, as illustrated in Figure 2. Some of our results are shown in Figure 3.

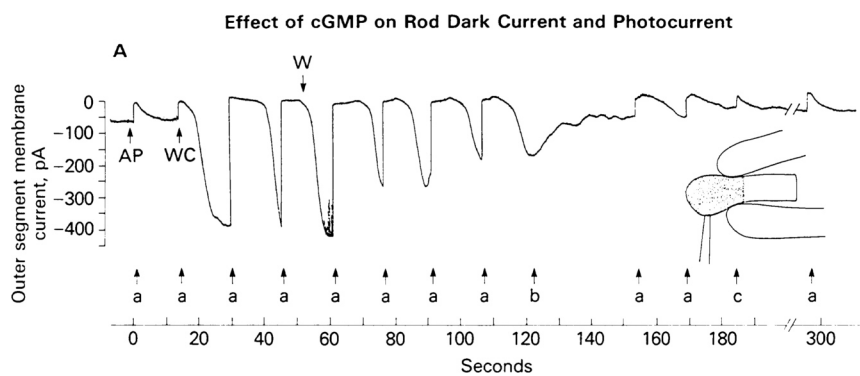


Figure 3

An experiment like that illustrated in Figure 2 in which the gigaseal pipette contained 5 mM cyclic GMP. The suction electrode records only outer-segment membrane current. At AP an attached patch (having a resistance greater than a gigohm) was formed; at WC the patch was reaptured by slight negative pressure and whole-cell recording mode commenced. A large inward current developed rapidly, which was entirely light suppressible. At W the whole-cell electrode was withdrawn and the membrane patch resealed; over the next 5 min the cell recovered completely to baseline. Positions marked with arrows indicate 20-msec flashes: a, 1000 isomerizations; b, 100; c, 10. The inset b is a tracing from the infrared video record of the experiment. SOURCE: Cobbs and Pugh (1985).

Since the suction electrode in Figure 3 records only from the outer-segment membrane, this experiment and similar ones by Matthews et al. (1985) demonstrated that the current induced by cGMP is located in the outer segment. In another set of experiments with the same technique (Cobbs et al., 1985), it was demonstrated that cGMP has the same effect on cones—that is, it increases an outer-segment inward current that light suppresses.

The cyclic GMP messenger hypothesis came into full flower with the exciting discovery by Fesenko et al. (1985) of a cGMP-gated cation conductance in excised patches of frog outer-segment membrane, as illustrated in Figure 4.

Fesenko et al. presented several arguments for identifying the cGMP-gated conductance with the light-sensitive conductance, including I/V relation, noise spectrum, and ion specificity, although all these fingerprints were somewhat discrepant from what was expected. Soon thereafter, however, identification of the cGMP-activated conductance, g_{cGMP} , with the light-sensitive conductance, g_{hv} , was made firm by Yau and Nakatani (1985) and

cGMP-Sensitive Conductance of Rod Outer Segment Membrane

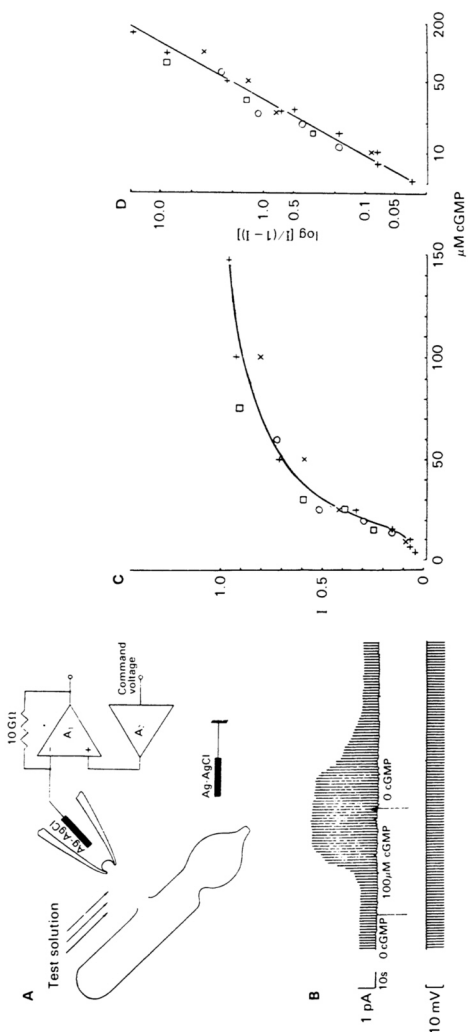


FIGURE 4 (A) Scheme used by Fesenko et al. (1985) in their discovery of the cGMP-gated conductance of the rod outer-segment membrane. A small patch of outer-segment membrane was excised from the outer segment with the patch-clamp technique of Neher and Sakmann. (B) Voltage pulses were then applied across the membrane patch in the presence or absence of various concentrations of cGMP. (C) The concentration dependence of the cGMP-induced current of several membrane patches, with the theoretical curve given by the Hill equation for cooperative activation, $I/I_{max} = [cGMP]^N / ([cGMP]^N + K_D^N)$, with $N = 1.8$, and $K_D = 30 \mu M$. (D) The same results plotted in the form $\log [I/(1 - I)]$ versus $\log [cGMP]$, which linearizes the curve, so that the slope gives N . SOURCE: Replotted from Fesenko et al. (1985, Figures 1 and 2).

by Matthews (1986, 1987). Matthews, for example, demonstrated conclusively that the light-suppressible channel noise recorded from cell-attached, outer-segment membrane patches had a spectrum identical to that recorded when the identical membrane patch was pulled and perfused with cGMP.

The cGMP Cascade: A Biochemical Sketch

Figure 5 shows a sketch of the so-called cGMP cascade of the outer segment. This diagram summarizes not only the electrophysiological work cursorily reviewed above but also a virtual flood of biochemistry that followed in the wake of Bitensky et al.'s (1971) hypothesis (space does not even permit even a cursory review of the relevant biochemical literature; for extensive reviews, see Miller, 1981; Stryer, 1986; Pugh and Cobbs, 1986; Liebman et al., 1987).

According to the canonical version of the cascade theory, the sequence of events is as follows: light does only one thing—isomerize the rhodopsin chromophore, causing it to undergo a conformation change that converts it into an enzyme. As an enzyme, R^* catalyzes the binding of GTP to a protein, GTP-binding protein, which is present in an amount of about one copy for every 10 rhodopsins. Several lines of evidence say that one R^* —in the absence of the inactivation process—can activate several hundred G-proteins/second (Liebman and Pugh, 1979, 1982; Vuong et al., 1984). Under normal intracellular conditions, however, R^* is inactivated fairly rapidly, apparently by phosphorylation and the binding of "48K protein." G^* is not an enzyme but rather a general messenger with its own limited lifetime. While active, G^* in turn activates, by direct binding, the third protein in the cascade, phosphodiesterase (PDE). PDE is a powerful enzyme capable of catalyzing the breakdown or hydrolysis of about 2000 cGMPs per second. One should keep in mind, however, when describing the "power" of PDE that it is likely that the *free* cGMP in the rod outer segment is far lower than the k_m of the enzyme, as guessed in 1985 (Cobbs and Pugh, 1985), and that the hydrolysis rate is always scaled by the k_m . According to the theory, PDE* next reduces the local free cGMP; as the cGMP unbinds from the conductance, it is not replaced, causing g_{hv} to close.

This canonical cGMP cascade theory requires fleshing out, even to be a complete qualitative ("what does what") theory and, indeed, is well known to contain several unresolved problems. Most of these problems have to do with the kinetics of the restorative reactions, which are shown in the open arrows in Figure 5. A key problem that remained unresolved after calcium was rejected as the excitational messenger was the nature of its role, if any, in transduction. This problem now appears near resolution. Having been rejected as the excitational message, intracellular calcium concentration has now made a major comeback as a restorative feedback signal and likely

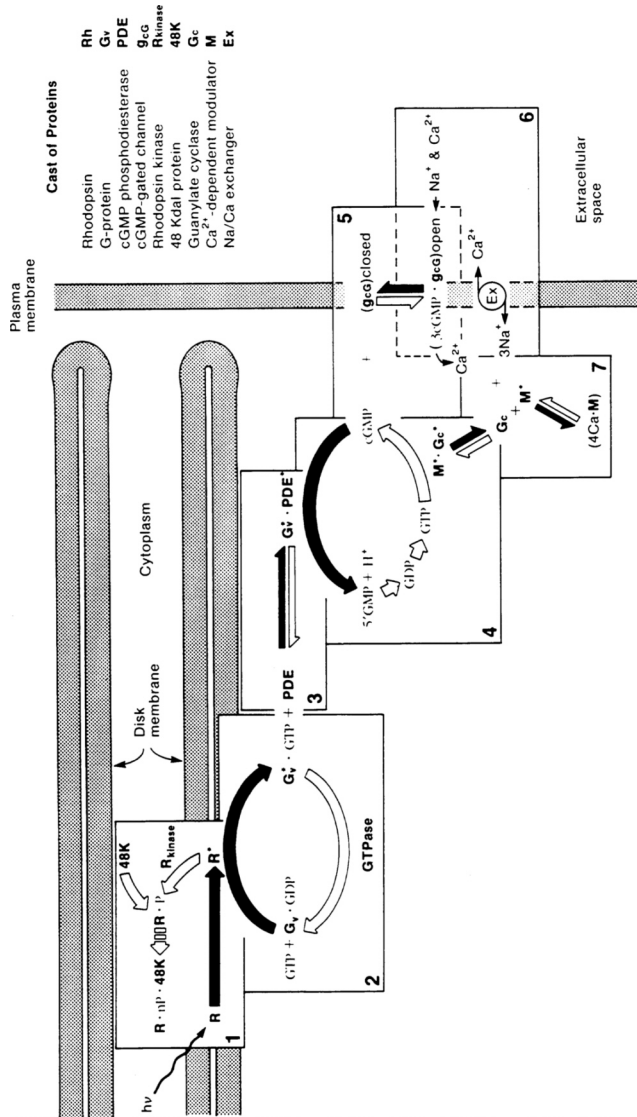


Figure 5
 Schematic of the cGMP-cascade theory derived from the work of many investigators since 1971. Each box in the diagram represents an identified step in the cascade: the solid arrows represent reactions that lead to the closure of the cGMP-gated conductance; the open arrows represent reactions that restore the conductance to the open state. Arched arrows represent enzymatically catalyzed steps (which produce amplification); straight arrows represent binding interactions. The proteins that play key roles in the cascade are identified. In this schematic the role of calcium is elaborated in boxes 5 through 7.
 SOURCE: Pugh and Altman (1988).

as the internal adaptational message (Matthews et al., 1988; Nakatani and Yau, 1988). The primary way by which dynamic changes in internal Ca^{2+} affect the cascade appears to be via the guanylate cyclase reaction (Koch and Stryer, 1988; Hodgkin and Nunn, 1988). The figure reflects thinking and experimentation up to about early 1989.

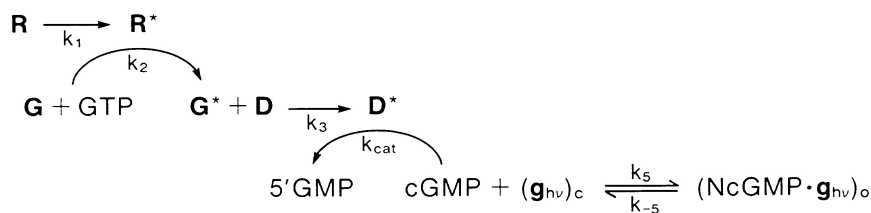


Figure 6

A formal chemical schematic of the "activation" steps in the cascade. R^* , G^* , and D^* represent the activated forms of the proteins rhodopsin, G-protein, and phosphodiesterase, respectively; \mathbf{g}_{hv} represents the cGMP-activated conductance, which is opened by the cooperative binding of $N = 2$ to 3 cGMP molecules. A differential equation representation of this reaction sequence is presented in Cobbs and Pugh (1987b).

FORMAL THEORY

The remainder of this paper sketches our recent efforts to develop a formal representation of the reactions that lead to the closure of \mathbf{g}_{hv} (reactions represented by the filled arrows in Figure 5) and to test this representation against responses of voltage-clamped rods to intense, brief flashes isomerizing up to 20 percent of the rhodopsin. The strategy employed is diametrically opposite those pioneered by Penn and Hagins (1972) and Baylor et al. (1974), both of whom developed formal analyses of the dim flash or linear response and then attempted to incorporate appropriate saturating nonlinearities to extend the analysis. Rather, we build the theory to handle the responses to the most intense flashes first and then change one critical parameter to see if it can explain the responses to less intense flashes. A priori testing of a theory against the most nonlinear range of the response of the rod seems a poor strategy. Nonetheless, we hope to show that this strategy not only makes a great deal of sense but also actually yields rich insight into the system.

Figure 6 shows a chemical representation of the relevant cascade reactions. A differential equation representation of this chemical model can be found in Cobbs and Pugh (1987).

One of the interesting formal features of the cGMP theory is that it relies on a "negative" messenger; that is, the signal or photocurrent is produced by the messenger's *removal*. This feature is no doubt shared by messengers in other systems but is nonetheless remarkable. Because

of this feature, the maximum speed of the photocurrent can in principle be rate limited by any of several steps up to and including the rate of unbinding of the cGMP from the conductance and its attendant closure. In contrast, as pointed out by Penn and Hagins (1972), a transduction theory that employs a positive messenger whose release rate is a monotonic function of the number of R*s produced by a brief flash predicts that the photocurrent should not reach velocity saturation over the whole range of bleach intensities. This absence of rate limitation is predicted because the more intense the flash the faster the messenger should be released, and thus the second-order binding reaction of the positive messenger to the conductance (which closes it) should proceed at a speed ever increasing with flash intensity.

Photocurrent Velocity and Delay Saturation

Figure 7 shows the voltage-clamp photocurrents of two rods stimulated with a series of intense 20 μ sec flashes. Note that the *velocity* of the voltage-clamp photocurrent appears to saturate at a relatively modest fractional isomerization, about 0.001. By virtue of the voltage-clamp technique it can be confidently asserted that this velocity limitation is not due to capacitative loading—the membrane potential could be shown to have changed less than a millivolt during the response. Above 0.001 fractional isomerization, the velocity-saturated photocurrent seems to translate laterally to the left with increasing bleach fraction, and this translatory behavior itself saturates, reaching its left-most position at about 0.1 fractional isomerization. This latter saturation phenomenon we call delay saturation. The average apparent latency of the velocity- and delay-saturated voltage-clamp photocurrent is about 7 msec (Cobbs and Pugh, 1987).

Theoretical Account of Velocity Saturation

Can a formal representation of the cGMP cascade theory account for these phenomena, and is the account quantitatively consistent with the biochemical data on the reactions? Theory suggests two ways in which the approximately exponential velocity-saturated photocurrent could arise: one is simply by virtue of the rate of cGMP unbinding from the channel and its closure; the second is by virtue of the limitation posed by the maximal PDE activity on the rate at which cGMP may be removed from the cytoplasm. Employing numerical values taken from the biochemical literature (Table 3 and 4 in Cobbs and Pugh, 1987) for most of the theory parameters, we found that a velocity limitation imposed either by the intrinsic rate of cGMP unbinding and channel closure (derived from noise power spectra) of the conductance or by the maximal PDE activity (or a combination

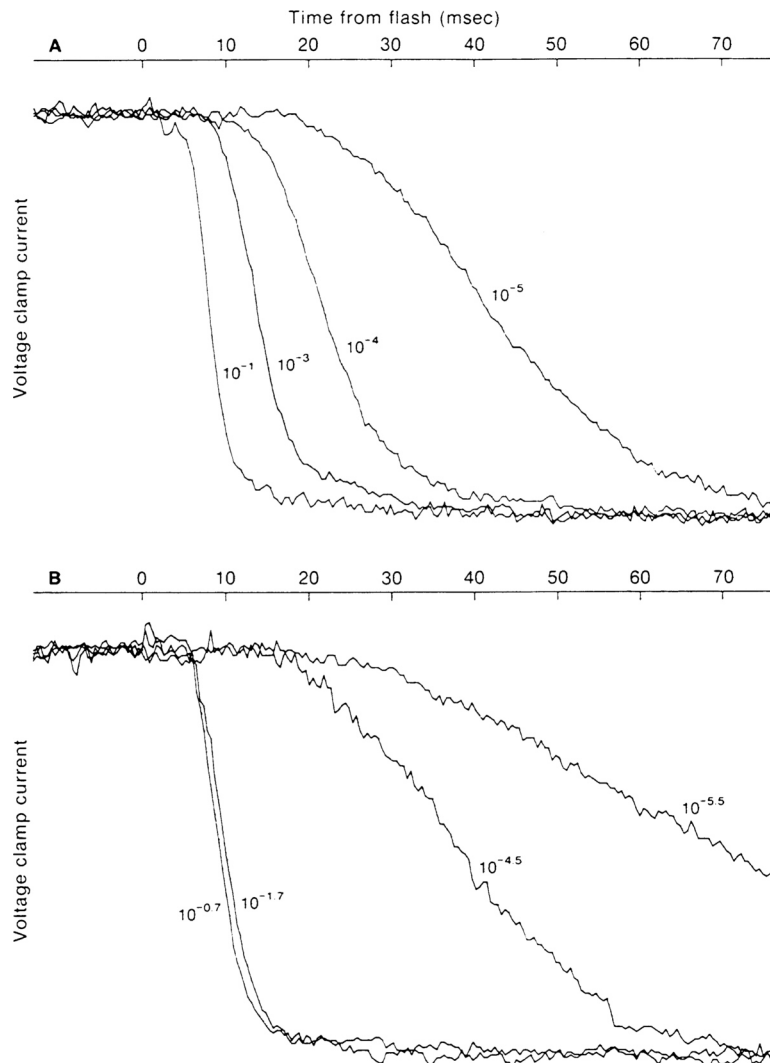


Figure 7
 Voltage-clamp photocurrents of two rods stimulated with a series of intense 20- μ sec flashes, which isomerize the fraction of rhodopsin given by the parameter near each curve. In part A note that there is almost no change in photocurrent velocity for the three most intense flashes, even though the left-most current was produced with a stimulus 1000 times more intense (10^{-1} vs. 10^{-4}): this is the phenomenon of velocity saturation. In part A note that the 10-fold increase from 10^{-4} to 10^{-3} produces about the same lateral translation as the 100-fold increase from 10^{-3} to 10^{-1} and that in part B the 10-fold change from $10^{-1.7}$ to $10^{-0.7}$ produces virtually no change in the lateral position of the photocurrent: this is the phenomenon of delay saturation. Reprinted from Cobbs and Pugh (1987) with permission.

About this PDF file: This new digital representation of the original work has been reproduced from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

of both) produced solutions of the equations consistent with our data. These two hypotheses about velocity saturation appear to be resolvable by testing the theory against the photocurrents of rods infused with cGMP or its weakly hydrolyzable analog, 8-Bromo-cGMP (Zimmerman and Baylor, 1986; Barkdoll et al., 1988). The extension of the theory must deal with two serious complications: (1) the diffusion of the nucleotide is relatively slow; (2) the rod outer segment has such a high longitudinal resistance (1 to 3 megohm/ μm) that its cable properties must be taken into consideration when modeling the cGMP-induced currents (Cameron and Pugh, 1988). Our work on extensions of the theory to incorporate diffusion and cable properties is nearly complete, and we should know shortly which feature of the cascade imposes the photocurrent velocity limitations.

Theoretical Account of Delay Saturation and Translation

Formal theory also provides interesting insight into the translatory behavior that is seen at fractional isomerizations of about 0.001, and the c. 7 msec saturation limit of the translation that occurs at about 0.1 fractional isomerization. It is interesting that the delay saturation occurs at about 0.1 fractional isomerization, at which intensity there is one photolyzed rhodopsin for every G-protein (the ratio of the two proteins being about 10/1). One would thus expect delay saturation to set in at this flash intensity, given that the time that it takes an isomerized rhodopsin to encounter a G-protein is short. And indeed the time must be short; otherwise one isomerized rhodopsin could not activate 500 to 1000 G-proteins/second (Fung and Stryer, 1980; Liebman and Pugh, 1982). Further insight can be had by detailed comparison of theory and data.

Figure 8a compares average velocity- and delay-saturated photocurrents of a population of 11 rods (thicker, noisy trace) with a family of theoretical curves (thinner traces). The left-most theoretical curve is a particular parameterization of the theory, optimized to fit the data trace. All the other theoretical curves are generated by changing a single parameter, the only rate constant in the theory expected to depend on light intensity. This rate constant can be expressed as the pseudo-first-order rate at which each rhodopsin activates the pool of G-protein to which it has access before the entire G-protein pool is activated (at these flash levels all the G-protein is predicted to become activated). To produce the series of theoretical curves this one rate constant has been decremented in a geometric series, each decrement approximately twofold. Clearly, the theory provides a natural account of the translatory behavior. The underlying cause of translation lies in the very nature of cascaded reactions followed by a rate-limiting step, not in any details of the parameters.

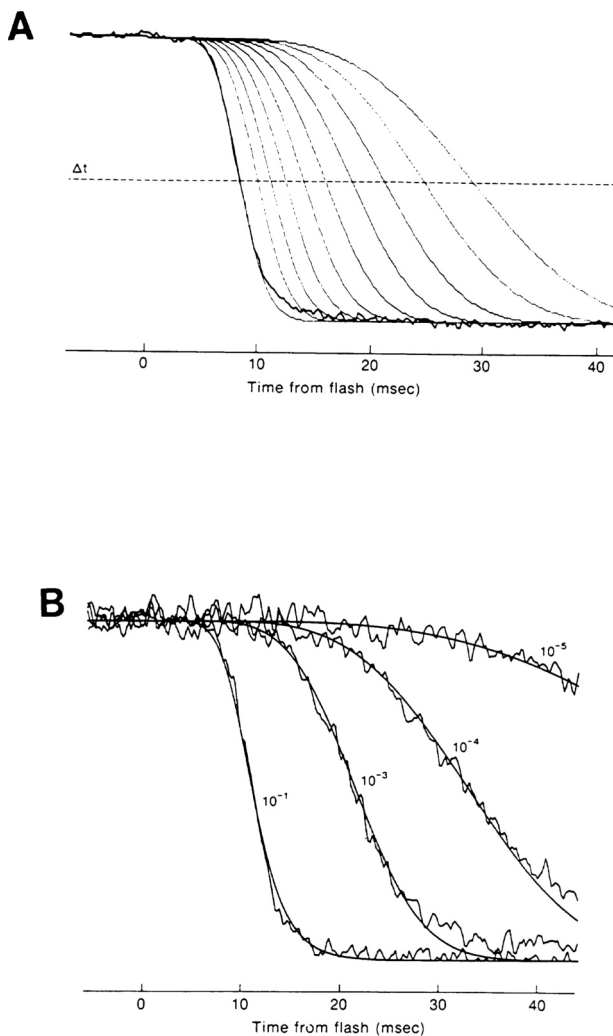


Figure 8

(A) The thickened noisy trace is the mean velocity- and delay-saturated photocurrent of 11 rods stimulated with 20- μ sec flashes that isomerized at least 2 percent of the rhodopsin. The thinner traces are solutions of the differential equation representation of the cascade equations in Figure 6, as given in Cobbs and Pugh (1987b), while the parameters are those used in Figure 15A-C of the same paper. Only a single rate parameter, k_{20} , the effective rate with which each rhodopsin activates its respective pool of G-protein, was varied to produce the theoretical traces. From far left to far right $1/k_{20} = 1.5, 6.0, 12.5, 25, 50, 100, 200, 400, 800,$ and 1600 msec. (B) Voltage-clamp photocurrents (noisy traces) of a single rod stimulated with 20- μ sec flashes isomerizing the fraction rhodopsin indicated. The smooth traces are generated by a representation of the cGMP-cascade theory: only one parameter varies between the curves: k_{20} , the pseudo-first-order rate with which rhodopsin activates G-protein. From left to right, $1/k_{20} = 2.07, 167, 1225,$ and 13,038 msec.

An approximate functional dependence of the pseudo-first-order rate of G-protein activation on light intensity can be derived from first-passage theory (Szabo et al., 1980), if one assumes that the rate-limiting step in activation of G-protein by rhodopsin is intramembrane lateral diffusion of rhodopsin (Liebman and Pugh, 1981, 1982); this in turn can be used to predict the delay (Δt in Figure 8a). The agreement of predicted delay with observation is quite good (see Cobbs and Pugh, 1987, Figure 12C). To test the functional dependence in more detail, however, we need to be able to estimate from photocurrents produced by different intensities on individual cells the rate constant itself, not just the delay. Figure 8B shows how this can be done. A family of responses from a single rod is simultaneously fit with an instantiation of the theory by a parameter search routine; the only parameter that is allowed to vary between curves is the rate constant in question. Since we have independent knowledge of the fraction isomerized, we can thus infer the dependence of the rate constant on isomerization. We are in the process of confronting the theory in this fashion with the individual records of many cells, in effect using the theory as a measuring tool for estimating dependence of the rate on fractional isomerization and thus testing the lateral diffusion hypothesis.

Efforts to build a formal representation of the cascade that could account for the voltage-clamp responses could not account for the magnitude (7 msec) of the saturated delay unless one were to postulate that about 3 or 4 first-order delay steps of order 2 msec intervened between the binding of rhodopsin to G-protein and the velocity-saturation mechanism (see Cobbs and Pugh, 1987b, Appendix, for details). A natural mechanism for such additional minor delay components would be in the binding of GTP, the conformation change of the G-protein, the binding of G-protein to PDE, and the latter conformational change as it becomes an active enzyme. Here again, theory, applied to the photocurrents, allows us to "see" a biochemical process that might otherwise escape attention.

SUMMARY

The hypothesis that there is an internal messenger in vertebrate phototransduction has offered a rich challenge to scientists in many fields and has brought together a wide array of tools and insights. With most of the fundamental qualitative features of phototransduction now understood (i.e., the identity of the molecular entities involved and the roles they play), the field of phototransduction is entering a phase in which formal theories linking photoreceptor biochemistry and electrophysiology can play an important role, permitting deeper understanding of this wonderful process.

References

- Bader, C.R., P.R. MacLeish, and E.A. Schwartz 1979 A voltage-clamp study of the light response in solitary rods of the tiger salamander. *Journal of Physiology* 296:1-26.
- Barkdoll, A.E., III, A. Sitarayya, and E.N. Pugh, Jr. 1988 Kinetics of the hydrolysis of 8-Bromo-cyclic GMP by the light-activated phosphodiesterase of toad rods. *Journal of Neurochemistry* 50:839-846.
- Baylor, D.A., and B.J. Nunn 1986 Electrical properties of the light-sensitive conductance of salamander rods. *Journal of Physiology* 371:115-145.
- Baylor, D.A., A.L. Hodgkin, and T.D. Lamb 1974 The electrical response of turtle cones to flashes and steps of light. *Journal of Physiology* 242:685-727.
- Baylor, D.A., T.D. Lamb, and K.W. Yau 1979 The membrane current of single rod outer segments. *Journal of Physiology* 288:589-611.
- Baylor, D.A., G. Matthews, and K.W. Yau 1980 Two components of electrical dark noise in toad retinal rod outer segments. *Journal of Physiology* 309:591-621.
- Bitensky, M.W., R.E. Gorman, and W.H. Miller 1971 Adenyl cyclase as a link between photon capture and changes in membrane permeability of frog photoreceptors. *Proceedings of the National Academy of Sciences, USA*, 68:561-562.
- Bodoia, R.D., and P.B. Detwiler 1984 Patch-clamp recordings of the light-sensitive dark noise in retinal rods from the lizard and frog. *Journal of Physiology* 367:183-216.
- Cameron, D.A., and E.N. Pugh, Jr. 1988 Outer segment cable properties set upper limit to cGMP-induced clamp current of salamander rods. *Society for Neuroscience Abstracts*, Pt. 1, p. 34, #19.2.
- Cobbs, W.H., and E.N. Pugh, Jr. 1985 Cyclic GMP can increase rod outer segment light-sensitive current 10-fold without delay of excitation. *Nature* 313:585-587.
- 1987 Kinetics and components of the flash photocurrent of isolated retinal rods of the larval salamander, *Ambystoma tigrinum*. *Journal of Physiology* 394:529-572.
- Cobbs, W.H., A.E. Barkdoll III, and E.N. Pugh, Jr. 1985 Cyclic GMP increases photocurrent and light-sensitivity of retinal cones. *Nature* 317:64-66.
- Fesenko, E.E., S.S. Kolesnikov, and A.L. Lyubarsky 1985 Induction by cyclic GMP of cationic conductance in plasma membrane of retinal rod outer segment. *Nature* 313:310-313.
- Fung, B.B.K., and L. Stryer 1980 Photolyzed rhodopsin catalyzes the exchange of GTP for bound GDP in retinal rod outer segments. *Proceedings of the National Academy of Sciences, USA*, 77:2500-2504.
- Hagins, W.A., R.D. Penn, and S. Yoshikami 1970 Dark current and photocurrent in retinal rods. *Biophysical Journal* 10:380-412.

- Hodgkin, A.L., and B.J. Nunn 1988 Control of light-sensitive current in salamander rods. *Journal of Physiology* 403:439-471.
- Koch, K.-W., and L. Stryer 1988 Highly cooperative feedback control of retinal rod guanylate cyclase by calcium ions. *Nature* 334:64-66.
- Lamb, T.D. 1986 Transduction in vertebrate photoreceptors: the roles of cyclic GMP and calcium. *Trends in Neurosciences* 9:224-228.
- Lamb, T.D., P.A. McNaughton, and K.W. Yau 1981 Spatial spread of activation and background desensitization in toad rod outer segments. *Journal of Physiology* 319:463-496.
- Lamb, T.D., H.R. Matthews, and V.R. Torre 1986 Incorporation of calcium buffer into salamander retinal rods: a rejection of the calcium hypothesis of phototransduction. *Journal of Physiology* 372:315-349.
- Liebman, P.A., and E.N. Pugh, Jr. 1979 The control of phosphodiesterase in rod disk membranes: kinetics, possible mechanisms and significance for vision. *Vision Research* 19:375-380.
- 1981 Control of rod disk membrane phosphodiesterase and a model for visual transduction. Pp. 157-170 in *Molecular Mechanisms of Visual Transduction*, W.H. Miller, ed. New York: Academic Press.
- 1982 Gain, speed and sensitivity of GTP binding vs. PDE activation in visual excitation. *Vision Research* 22:1475-1480.
- Liebman, P.A., K.R. Parker, and E.A. Dratz 1987 The molecular mechanism of visual excitation and its relation to the structure and composition of the rod outer segment. *Annual Review of Physiology* 49:765-791.
- MacLeish, P.R., A.E. Schwartz, and M. Tachibana 1984 Control of the generator current in solitary rods of the *Ambystoma tigrinum* retina. *Journal of Physiology* 348:645-664.
- Matthews, G. 1986 Comparison of the light-sensitive and cyclic GMP sensitive conductances of the rod photoreceptor: noise characteristics. *Journal of Neuroscience* 6:2521-2526.
- 1987 Single-channel recordings demonstrate that cGMP opens the light-sensitive ion channel of the rod photoreceptor. *Proceedings of the National Academy of Sciences, USA*, 84:582-585.
- Matthews, H.R., V. Torre, and T. Lamb 1985 Effects on the photoresponse of calcium buffers and cyclic GMP incorporated into the cytoplasm of retinal rods. *Nature* 313:582-585.
- Matthews, H.R., R.L.W. Murphy, G.L. Fain, and T.D. Lamb 1988 Photoreceptor light adaptation is mediated by cytoplasmic calcium concentration. *Nature* 334:67-69.
- McNaughton, P.A., L. Cervetto, and B.J. Nunn 1986 Measurement of the intracellular free calcium concentration in salamander rods. *Nature* 322:261-263.
- Miller, W.H., ed. 1981 *Molecular Mechanisms of Photoreceptor Transduction*. New York: Academic.

- Miller, W.H. 1982 Physiological evidence that light-mediated decrease in cyclic GMP is an intermediary process in retinal rod transduction. *Journal of General Physiology* 80:102-123.
- Miller, W.H., and G.D. Nicol 1979 Evidence that cyclic GMP regulates membrane potential in retinal rod photoreceptors. *Nature* 280:64-66.
- Nakatani, K., and K.W. Yau 1988 Calcium and light adaptation in retinal rods and cones. *Nature* 334:69-71.
- Nicol, G.D., and W.H. Miller 1978 cGMP injected into retinal rod outer segments increases latency and amplitude of response to illumination. *Proceedings of the National Academy of Sciences, USA*, 75:5217-5220.
- Penn, R.D., and W.A. Hagins 1972 Kinetics of the photocurrent of retinal rods. *Biophysical Journal* 12:1073-1094.
- Pugh, E.N., Jr., and J. Altman 1988 A role for calcium in adaptation. *Nature* 334:16-17.
- Pugh, E.N., Jr., and W.H. Cobbs 1986 Visual transduction in vertebrate rods and cones: a tale of two transmitters, calcium and cyclic GMP. *Vision Research* 26:1613-1643.
- Pugh, E.N., Jr., and W.J. Miller 1987 Phototransduction in vertebrates. *Annual Review of Physiology* 49:711-714.
- Stirling, C.E., and A. Lee 1980 [³H]ouabain autoradiology of frog retina. *Journal of Cell Biology* 85:313-324.
- Stryer, L. 1986 Cyclic GMP cascade theory of vision. *Annual Review of Neuroscience* 9:87-119.
- Szabo, A., K. Schulten, and Z. Schulten 1980 First passage time approach to diffusion-controlled reactions. *Journal of Chemical Physics* 72:4350-4357.
- Vuong, T.M., M. Chabre, and L. Stryer 1984 Millisecond activation of transduction in the cyclic nucleotide cascade of vision. *Nature* 311:659-661.
- Yau, K.W., and K. Nakatani 1984 Electrogenic Na-Ca exchange in retinal rod outer segment. *Nature* 311:661-663.
- 1985 Light suppressible, cyclic GMP-sensitive conductance in the plasma membrane of a truncated rod outer segment. *Nature* 317:252-255.
- Yau, K.W., L.W. Haynes, and K. Nakatani 1986 Roles of calcium and cyclic GMP in visual transduction. In *Membrane Control of Cellular Activity*, H.C. Luttgau, ed. Stuttgart: Gustav Fischer.
- Yoshikami, S., and W.A. Hagins 1970 Ionic basis of dark current and photocurrent of retinal rods. *Biophysical Journal* 10:60a.
- Zimmerman, A.L., and D.A. Baylor 1986 Cyclic GMP-sensitive conductance of retinal rods consists of aqueous pores. *Nature* 321:70-72.

Partitioning Visual Processes

Walter Makous

IDENTIFYING AND SEQUENCING OPERATIONS

Method and Rationale

The work reported here is based on the idea that vision consists of the transformation of spatiotemporal patterns of light into behavior and experience. It follows that a visual scientist's goal is to understand the operations that accomplish this transformation and the paths that signals take through those operations. In human vision the pursuit of this goal is complicated by the fact that the operations and their organization usually must be inferred from the relationship between input and output (i.e., from psychophysics). If the visual system were entirely linear, any given relationship between input and output could be produced by many different combinations of operations and choosing among them would be either guesswork or a matter of taste. So nonlinear processes are what allow us to pare down the alternative representations of the system and, when they have been pared down enough, to know what goes on inside.

The way nonlinearities constrain the alternatives can be seen from a simple example. Consider a system with two inputs, a and b , that are combined to determine a single output, c . If there is a nonlinearity in the system, such that $c \propto \log a$ and $c \propto \log b$, one can establish whether the inputs sum before the log transformation [i.e., $c \propto \log (a + b)$] or after it [i.e., $c \propto (\log a) + (\log b)$] by varying both a and b while holding their sum constant ($a + b = k$). If the inputs sum before the log transformation, the variations of a and b have no effect on the output; if they sum after the log transformation, variations of a and b do affect the output. This is true of

any nonlinear transformation, not just the log transformation used in this example.

The method my co-workers and I used is directly analogous to this example. The two inputs, a and b , correspond to the illumination of rows of adjacent cones. Under some conditions, differential illumination of adjacent cones, such as that produced by a 60 c deg^{-1} grating, cannot be discriminated from a homogeneous field; that is, the input from adjacent cones is summed to produce a single-valued output, c . To determine whether this summation occurs before or after any nonlinear transformations, we systematically vary the relative illumination of adjacent cones so that in some regions of the retina adjacent cones are equally illuminated ($a = b$), and in other regions they receive very different illumination ($a \neq b$); but the sum is always held constant ($a + b = k$).

The necessary conditions of illumination are produced by two laser interferometers. The first was invented by Dave Williams. As its design has been published (Williams, 1985a), I shall not describe it here except to say that it allows precise and facile control over the contrast, orientation, and spatial frequency of gratings created by optical interference, which can exceed 200 c deg^{-1} . As the changes of contrast are rapid and afford no extraneous cues, the instrument supports forced choice observations.

The second interferometer, of simpler design (MacLeod et al., 1985), allows superimposition of two separate interference gratings on the retina, such as was done first by Burton (1973). The result of such superposition is illustrated in [Figure 1](#).

The first two lines represent the profiles of the two sine wave gratings created by optical interference. When they are superimposed, the retinal illuminances at each point add, producing the profile of "beats" shown in the third line, where modulation is very high in some regions and nearly absent in others.

If the period of the gratings is about twice the diameter of a cone, adjacent cones receive very different illumination ($a \neq b$) in the places where modulation of the beat pattern is high and very nearly the same illumination ($a \approx b$) in regions where modulation is low, but the sum of the illumination of adjacent stripes is always the same ($a + b = k$).

If the excitation at each point is a nonlinear function of illumination, such as is produced by a rectifier, clipper, or abrupt saturation (as illustrated by the function graphed in the inset between the third and fourth lines in [Figure 1](#)), the excitation has the profile shown in the fourth line. The top of the curve is clipped off, and of course the local space average tends to be more negative in certain regions than in others. These variations of negativity show up in the Fourier spectrum as a new component (shown by the smooth curve) that has been introduced by the nonlinear distortion; its frequency is equal to the difference between the frequencies of the two

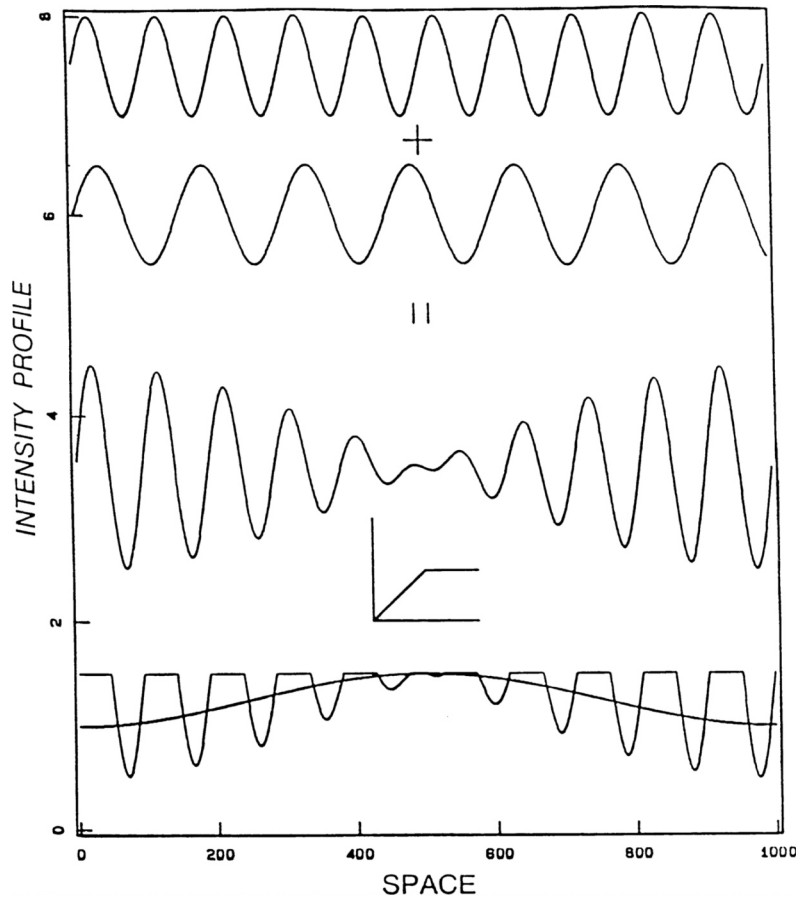


Figure 1
Formation of a difference frequency distortion product. The third curve is the sum of corresponding points in the first two. The bottom curve is this sum, transformed by the nonlinearity depicted by the graph between the two: vertical fluctuations of the third curve are the input, on the horizontal axis of the graph; and output, on the vertical axis of the graph, determines the vertical fluctuations of the last curve. Hence, variation of the input beyond the midpoint produces no variation in the output. The smooth curve superimposed on the output shows the frequency and phase of the distortion product that is a Fourier component of this output curve, on which it is superimposed.

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

superimposed gratings. Any nonlinear distortion produces this difference frequency; different nonlinearities simply produce different amplitudes of the difference grating.

Figure 2 shows how we use this distortion product. It depicts the information in two gratings of 60 and 70 c/deg entering a nonlinear stage. Coming out of the nonlinear stage are signals of the original frequencies, plus those of many other frequencies that are the products of the nonlinear distortion. In the human visual system these signals next encounter a low-pass filter that blocks everything but that of the lowest frequency, 10 c/deg. Subjectively, the observer sees only the low-frequency distortion product without seeing the high-frequency gratings that produced it. The essence of the technique lies in the fact that signals of only two input frequencies exist *before* the nonlinear stage and only the low-frequency distortion product exists *after* the nonlinear stage.

If one increases or decreases the frequencies of the two gratings while keeping the difference between their frequencies and everything else constant, the distortion product must remain unchanged. If this shift of frequencies does change the amplitude of the distortion product, it can only be due to the changes of the amplitudes of the signals elicited by the gratings at the site of the nonlinear stage. So the detectability of the distortion product can be used to measure the attenuation of different frequencies by whatever precedes the nonlinear stage.

Conversely, by changing the difference between the frequencies of input gratings without changing their mean frequency, the frequency of the distortion product can be varied with little effect on the signal passing through the stages that precede the nonlinearity. As the input pattern changes little, as there is little change in the signal reaching the nonlinear stage, and as the distortion product exists only after the nonlinearity, any effect on detectability of the distortion product can be caused only by attenuation of the signal *after* the nonlinear stage.

In practice, instead of using gratings of different spatial frequencies, we use gratings of equal spatial frequencies and create a distortion product by rotating the gratings slightly in opposite directions. This produces a pattern similar to a moiré pattern that has a distortion product corresponding to a sinusoidal grating oriented nearly perpendicular to the gratings producing it. The spatial frequency of the distortion product is changed by slightly changing the relative orientations of the gratings. Thus, there is no change of spatial frequency preceding the nonlinear process, and if the stages preceding the nonlinear process are reasonably isotropic, this procedure affects nothing except those processes that depend on the spatial frequency of the distortion product.

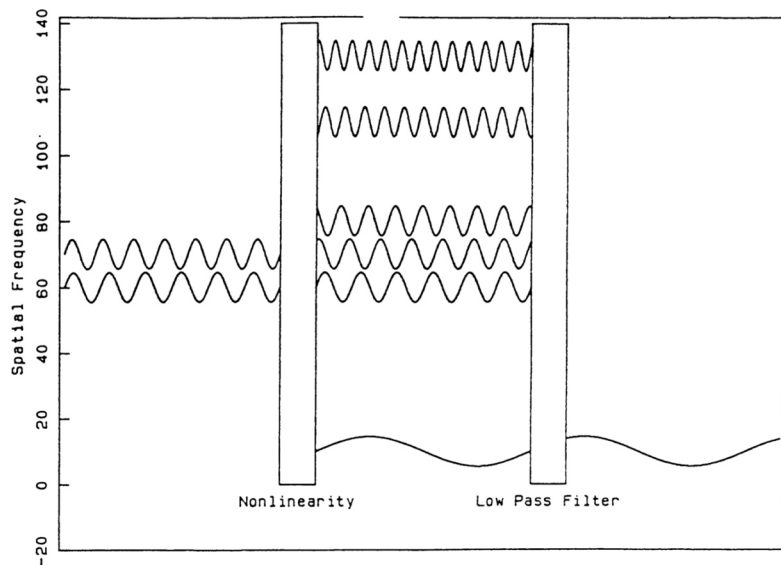


Figure 2

Frequencies present at successive stages of the visual system. Two sine waves enter, and no other signals are present preceding the nonlinear stage. So processes preceding the nonlinear stage are tested only at these input frequencies. The nonlinearity creates a variety of distortion products, one of which is the difference frequency. A subsequent low-pass filter blocks the input signals as well as all the distortion products except the lowest frequency, the difference frequency. The entering sine waves create the signal that is detected (the difference frequency) at the stage of the nonlinearity, and whatever happens to them subsequently is irrelevant. Only the difference frequency can be detected at the end of the system. The difference frequency exists only following the nonlinear stage, and so cannot be affected by what precedes the nonlinear stage (except indirectly through the effects on the input signals). So the processes that precede the nonlinearity are tested at the frequencies of the input signals, and the processes that follow the nonlinearity are tested at the frequency of the difference between the input frequencies.

Illustrative Results

Basic Partitions

Figure 3 shows schematically how this interferometric technique partitions visual processes. Interference gratings are unaffected by aberrations of the eye's optics and for many purposes must be treated as though they originate on the retina. This allows one to separate the effects of the eye's optics from other influences on the spatial information in visual stimuli. Campbell and Green (1965) used this to measure the losses of contrast attributable to the eye's optics. This they did by comparing the contrast

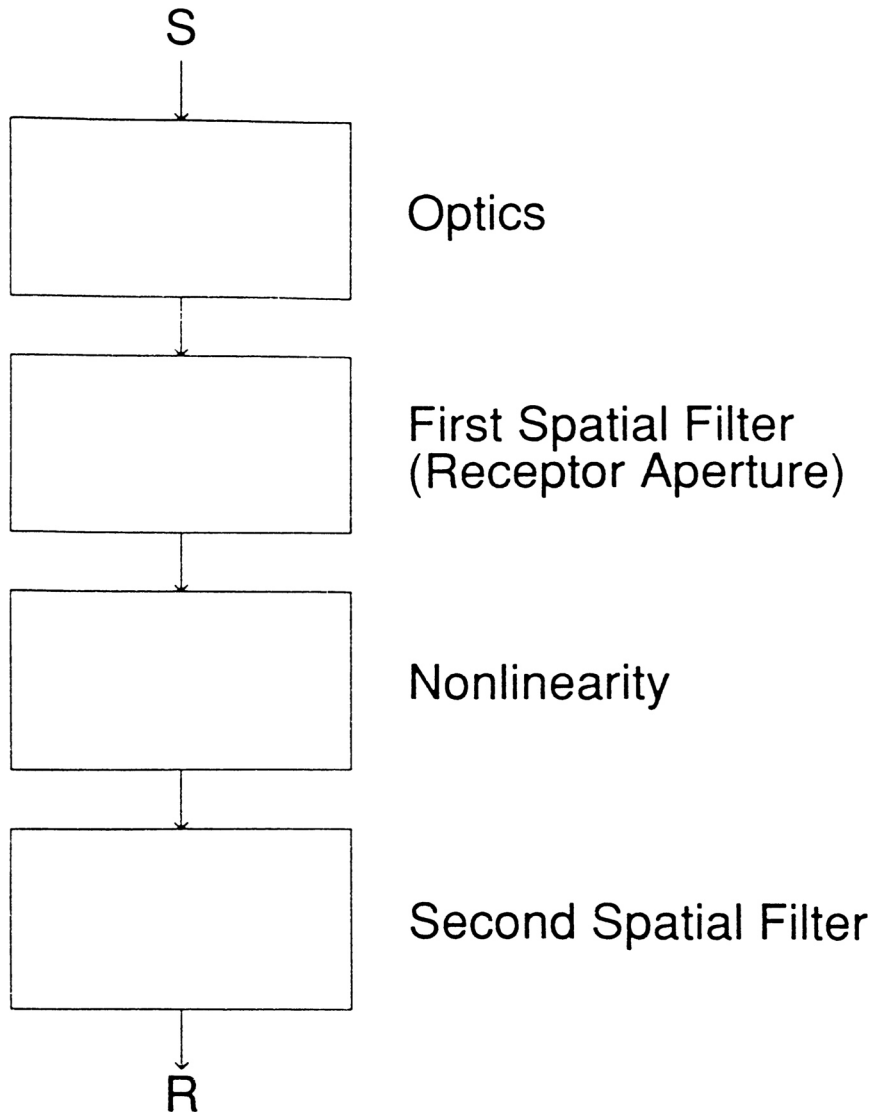


Figure 3
The four compartments into which the visual system is partitioned by the interferometer and the nonlinearity.

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

sensitivity of directly viewed gratings, which are degraded by the optics of the eye, to the sensitivity to interference fringes, which are not. The procedures used here allow further separation of the causes of losses of contrast—that is, separation of those between the optics and the nonlinear stage (labeled here the first filter or the receptor aperture) from those that follow the nonlinear stage (here called the second filter).

Losses Preceding the Nonlinear Stage

When distortion products of fixed frequency are created by superimposed interference gratings of variable frequency, the detectability of the distortion products decreases with increasing spatial frequency of the gratings (MacLeod et al., 1985; Chen et al., 1988). Data from one observer are shown by the filled triangles in [Figure 4](#) (MacLeod et al., 1985). Therefore, a stage of spatial filtering, separate from aberrations of the eye's optics, precedes the nonlinear stage. However, detectability of the distortions does not decrease greatly until the spatial frequencies of the gratings approach very high values. This means that spatial filtering preceding the nonlinear stage is not great.

Information in retinal images must enter the visual system through a set of individual receptors that for many purposes can be considered identical parallel channels. The optical apertures of these receptors constitute low-pass filters. The attenuations of contrast inferred from the detections are just what would be produced by filtering the gratings through an optical aperture of 24" of arc (mean of four observers, with a standard deviation of 6"). This is close to 80 percent of the anatomical diameter of foveal cones (Curcio et al., 1987), as predicted by Miller and Bernard (1983). This allows for no additional filtering. Since all the spatial information entering the visual system must pass through the aperture of a receptor, and since filtering accounts for all that is measured, we take it as the average optical aperture for the foveal cones of these observers. Consequently, the box in [Figure 3](#) representing the spatial filtering following the optics of the eye and preceding the nonlinear stage is labeled receptor aperture.

The psychologically measured aperture 3.8 deg from the fovea is 52.4" of arc (mean of four observers, with a standard deviation of 11") (MacLeod et al., 1985; Chen et al., 1988). The difference in the psychophysically measured apertures is in proportion to the difference in cone sizes (Curcio et al., 1987).

Losses Following the Nonlinear Stage

The octagons in [Figure 4](#) show the attenuations of varying spatial frequencies by those processes that *follow* the nonlinear stage (Chen et al.,

1988). This combines in a single curve all the influences on transmission of spatial information following the nonlinear stage, whether a direct consequence of the receptive fields of visual neurons or such things as spatial uncertainty (Pelli, 1985), probability summation (Graham, 1977), or the consequences of photon noise (Banks et al., 1987). The low-pass filtering performed by these processes reduces the width of the band of spatial frequencies passed to about a third of that reaching the nonlinear process.

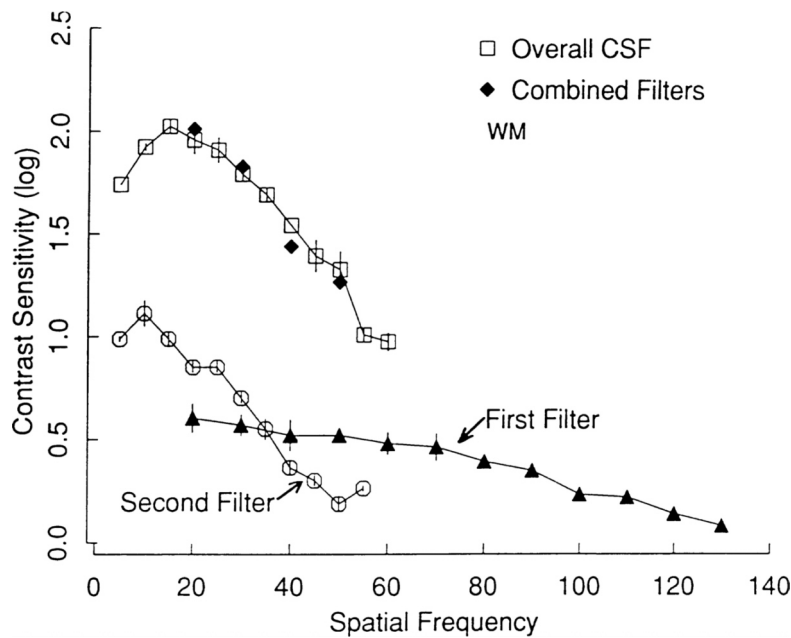


Figure 4
 Log attenuation of signals of varying frequency ($c \text{ deg}^{-1}$) measured by sensitivity to the contrast of interference fringes. Unfilled squares show the negative log contrast of interference fringes at threshold when the fringes are presented alone. The octagons show the negative log contrast of a $60\text{-}c \text{ deg}^{-1}$ interference fringe superimposed on another $60\text{-}c \text{ deg}^{-1}$ fringe of 100 percent contrast. The angle between the two fringes was adjusted to produce distortion products at the frequencies shown on the horizontal axis. Only the distortion product determined threshold. The triangles represent half the negative log contrast of an interference fringe superimposed on another fringe of 100 percent contrast, where the frequency of the distortion product was 10 c deg^{-1} , and the frequency of the fringes is represented by the position of the points on the horizontal axis. Only the distortion product determined threshold. (Division by 2 is necessary to allow for the fact that both fringes are attenuated.) The filled diamonds are the sum of the octagons and triangles, displaced vertically as a set to fit the unfilled squares.

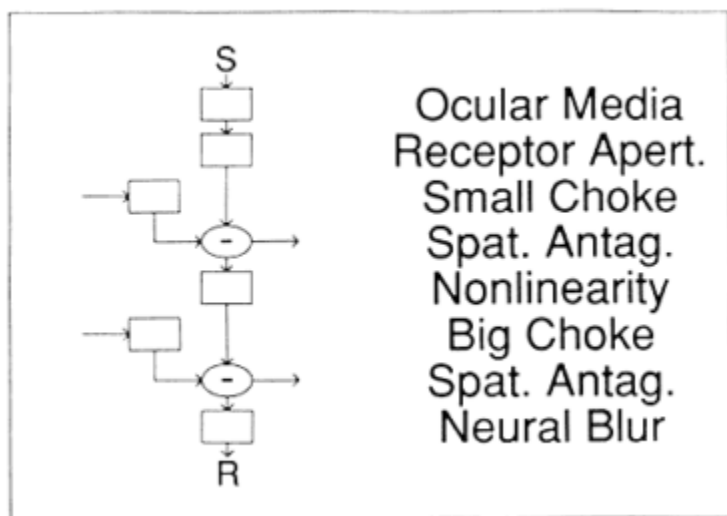


Figure 5
Elaboration of Figure 3 to show the spatial antagonism and temporal filters that form parts of the first and second filters.

The octagons in Figure 4 also show that the signals elicited by gratings of 5 c/deg are attenuated more than those elicited by gratings of 10 c/deg. Although this is but a single point, the difference is highly reliable and both observers tested show it. An increasing attenuation of signals elicited by gratings of decreasing spatial frequency requires an antagonistic interaction between signals originating in different receptors. This is a refinement not represented by the four operations of Figure 3. Consequently, in Figure 5 the representation of the filters preceding and following the nonlinear stage has been expanded to more detail. The last stage of Figure 3, labeled second filter, now consists of a low-pass filter, labeled neural blur, and a junction where antagonistic signals from different receptors interact. The negative sign represents mutual antagonism without specifically denoting algebraic subtraction.

A Test for Serial Filter Model

According to the logic expressed by Figure 3, the signals elicited by the gratings are attenuated by two serial filters, one before the nonlinear stage and one after. Then the contrast sensitivity to interference fringes (the total attenuation of the signals by the visual system) should be the

product of the attenuations by these two filters, that is, the sum of the triangles and octagons of Figure 4 (which have logarithmic ordinates). The filled diamonds in Figure 4 represent that sum, and the squares represent the contrast sensitivity to interference fringes (thus avoiding attenuation by the optics of the eye). As we cannot determine the absolute size of the distortion products, the set of diamonds has been adjusted vertically for the best fit. Insofar as the shapes of the curves defined by the two sets of points are similar over the range where the test can be made, the validity of partitioning these attenuations is supported. (Control experiments also show that the properties of the second spatial filter do not depend on the spatial frequencies of the fringes used to produce the distortion gratings.)

Spatial Antagonism Preceding the Nonlinearity

Our technique does not allow direct measurement of the attenuation preceding the nonlinear stage at low spatial frequencies. However, if the sum of the effects represented by triangles and octagons equals the total effects determining the shape of the contrast sensitivity curve represented by the squares, the *difference* between the squares and octagons should reflect the shape of the curve represented by the triangles. This provides a means of estimating where the triangles might lie if we could measure the function at low spatial frequencies. When this is done, the results show a progressive loss of sensitivity as spatial frequency decreases, so that 5-c/deg gratings are attenuated more than twice as much as 30-c/deg gratings. As stated above, this requires spatially antagonistic interactions between signals from different cones, and so in Figure 5 the connections representing such lateral interactions are introduced before the nonlinear stage.

Spatiotemporal Filters

Both antagonistic interactions (but *only* the antagonistic interactions) are reduced by decreasing the duration of the test grating from 500 to 50 msec (Chen et al., 1988); so the spatially antagonistic signals pass through low-pass temporal filters. These also have been introduced in Figure 5 in the pathway of signals entering into the spatial antagonism, where they have been given the admittedly passé label (to save space) *choke*.

Decreasing the duration of the test grating affects the low-frequency drop following the nonlinear stage more than that preceding the nonlinearity. Thus, the size or effectiveness of the two chokes differs, and this is represented by calling one *big* and the other *small*.

Cumulative Spatial Filtering

The results above show that spatial filtering by the visual system can be separated into the serial stages of [Figure 3](#), and the unfilled squares in [Figure 4](#) show the accumulated attenuations by two of these stages. However, the attenuations attributable to the optics are not shown, and the transfer functions of [Figure 4](#) do not serve intuition as well as other representations, such as spread functions. Consequently, the cumulative spreads of signal from a point source following each of the three stages are shown in [Figure 6](#). The solid line is the optical point-spread function computed from Westheimer's formula (1986), which is based on Campbell and Gubisch's data (1966). The dashed line is the convolution of the optical spread function with the spread function of the first filter, computed from triangles of [Figure 4](#). Finally, the dotted line is the convolution of the dashed line with the spread function of the second filter, computed from the octagons of [Figure 4](#). As the last filter follows a nonlinearity, this is not a true spread function. Nevertheless, it gives an intuitive grasp of the relative losses of fine detail as the information in an image passes through the visual system, confirming the general belief that the optics of the eye account for most but certainly not all of it. The figure also puts the psychophysical contribution of center-surround antagonism into perspective.

Summary of Results

The preceding section shows how the interferometer and nonlinear process can be used. [Figure 7](#) summarizes the results this approach has produced so far.

This diagram is not intended as a model but as a way to summarize the results, organize them, and show how they relate to one another. With the exceptions specified below, all the features of the diagram are dictated by the data. As a consequence, the figure shows only those operations required by these data, and some operations in the diagram represent a multitude of processes that the experiments do not separate.

Each stage, then, represents an operation the visual system must, according to our observations, perform on the information in its stimuli, and the arrows depict the flow of information through these operations. The labels are chosen simply to carry the most descriptive information in the space available. The solid lines and bold labels denote the stages where we have quantified the relationship between what goes into the designated operation and what comes out.

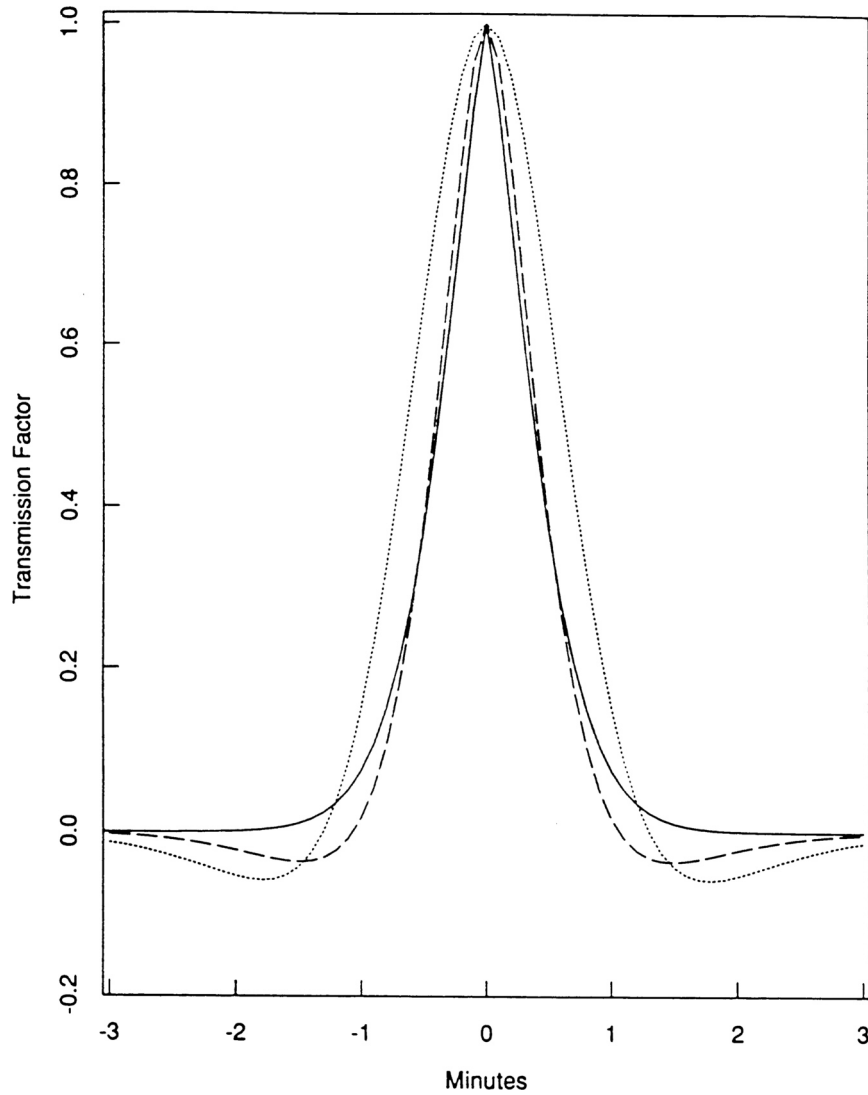


Figure 6
Point-spread function after the information has passed through the eye's optics (solid line), through the filtering preceding the nonlinear stage (dashed line), and through the entire visual system (dotted line).

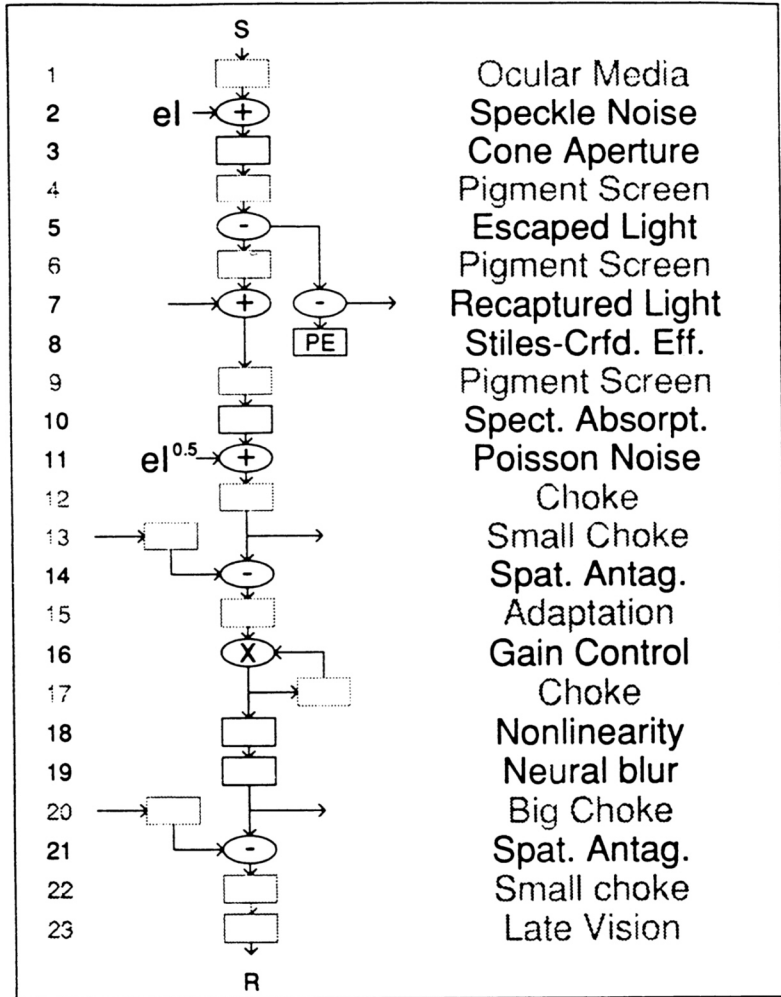


Figure 7
 Schematic diagram of the processes identified by use of the interferometer and nonlinearity. Operations where such experiments have quantified the relationship between input and output are emboldened.

About this PDF file: This new digital representation of the original work has been reproduced from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

Optics (Stages 1 to 3)

The first operation performed on the entering light is spatial filtering by the optics of the eye (stage 1). This we have experimentally bypassed so far, and so we have not measured it quantitatively, although Campbell and Green (1965) have.

As the coherent light of lasers passes through the ocular media, small inhomogeneities cause local interference patterns referred to as laser speckle. This is a source of spatial noise that masks the gratings (Williams, 1985b). Although this is not normally part of vision, it affects most of the observations made in these experiments and thus is included in the sequence of operations (stage 2).

The magnitude of masking caused by laser speckle is evaluated by mixing incoherent light, which is not subject to such speckle, with the coherent light from the laser. The elevation of threshold it produces can approach a factor of 10, but it varies with spatial frequency and observer (Williams, 1985b; Makous et al., 1985; Chen et al., 1988) but not the level of adaptation (Makous et al., 1985).

Light that enters the cone is spatially filtered by the receptor aperture (stage 3), as discussed above.

Phenomena Related to Light Absorption (Stages 4 to 11)

The Stiles-Crawford effect (1933) shows that more light is needed to have a given effect if it falls on the cones at an oblique angle than if it is incident axially. This is generally attributed to the escape of obliquely incident light from the cones. Some of the light that escapes the cones is ultimately absorbed by inert pigments in the pigment epithelium and is entirely lost to the visual system. This is part of the escaped light that is reflected by the Stiles-Crawford effect. The amount of light that must be added to obliquely incident light so as to have the same effect as axially incident light must equal the amount lost. However, if some light that escapes one cone re-enters and is absorbed by another cone (i.e., is *recaptured* by the other cone), it is not lost but has the same effect on most measures of the Stiles-Crawford effect as if it had been captured by the first cone it struck. Such light that escapes from one cone only to be recaptured by its neighbor must decrease the contrast of interference fringes cast on the retina by oblique light. The loss of contrast sensitivity to such fringes, then, provides a way to measure the recaptured light.

We have so measured the lost light and the recaptured light in two observers (Chen and Makous, 1988) and introduced symbols in [Figure 7](#) to represent them. The total amount of light that escapes the cones (stage 5) is the sum of the light lost (stage 8), measured by the Stiles-Crawford effect,

and the recaptured light (stage 7), measured by comparison of contrast sensitivity to interference fringes at oblique incidence with that at normal incidence. Some 70 to 90 percent of the light absorbed by the cones when it passes through the center of the pupil (axial incidence) is lost (stage 8) to the visual system if it passes through the margin of the dilated pupil; over half of what is ultimately absorbed (at stage 10) is recaptured light (stage 7). However, escaped light spreads little more one than cone before recapture. This recaptured light, then, causes optical cross talk among receptors.

Self-screening. Sensitivity to interference fringes is reduced by oblique incidence even when the stripes run parallel to the plane of incidence. This is taken as evidence that the incident light is temporarily captured by the first cone it encounters, before escaping. As light passes through the pigment contained by the cones, its spectrum changes, owing to preferential absorption of certain wavelengths by the photopigment (pigment screening). So there may be some self-screening of the light before it escapes. To recognize this possibility, stage 4 is introduced in Figure 7. Recaptured light may also undergo some self-screening after recapture and before absorption; stage 9 is introduced to recognize this possibility. Finally, the light captured by the first cone it strikes and ultimately absorbed by the same cone may pass through some pigment that neither escaped light nor recaptured light sees; stage 6 is introduced to cover that possibility.

None of these individual stages of self-screening is definitely established, but the phenomenon of self-screening in cones is generally accepted, and present evidence is consistent with its occurrence at any or all of these stages.

Action spectrum. As self-screening is the complement of the action spectrum, absorption of course occurs wherever pigment screening occurs. However, absorption and self-screening are shown as separate stages in Figure 7 to indicate that, although we have no quantitative information on these separate components of self-screening (stages 4, 6, and 9), we have identified the absorption spectrum (stage 10).

The Michelson contrast of interference fringes (Born and Wolf, 1970) is: $C = 2(I_1 I_2)^{1/2} / (I_1 + I_2)$. In the case where the light is filtered through a spectrally tuned filter (i.e., a visual photopigment), I is the retinal irradiance times the spectral absorption of the pigment at the wavelength used. Note that if the sum of I_1 and I_2 is held constant the equation defines a parabola with a peak at the point where the absorptions of the two wavelengths are equal. We have complete data on one observer and preliminary data on another showing that the threshold for distortions produced by superimposing 40-c/deg fringes of 632.8 and 514.5 nm (forming

heterochromatic beats) follows the equation above (Chen et al., 1986 unpublished observations) when I is weighted according to a single pigment with the spectral absorption a Smith and Pokorny fundamental (Pokorny and Smith, 1986). Which fundamental it is depends on the adaptive state of the eye.

Poisson noise. After the effects of the speckle noise (stage 2) are taken into account, sensitivity varies inversely with the square root of the mean retinal illuminance (Makous et al., 1985). This is attributed to a source of noise that increases proportionally; that source is introduced in Figure 7. Since some unknown part of this must be shot noise associated with the quantal absorptions, and because the sources of any other such noise can be localized only between the site of absorption and that of the nonlinear process, representation of all such sources of noise is positioned (stage 11) close to the spectral filter (stage 10), which also is associated with absorption of light. These two sources of noise, laser speckle and Poisson noise, account for 98 percent of the variance of thresholds measured at varying levels of adaptation.

Neural Processes Preceding the Nonlinear Process

Low-pass temporal filter (stage 12). Since there is a limit to the rate of flicker that can be detected, the visual system has a temporal filter. However, when two homogeneous superimposed fields flicker at slightly different rates, no flicker at the difference frequency can be detected unless the flicker of the fields also is detectable (Makous and Mandler, unpublished observations, 1984). This means that temporal filtering after the nonlinear stage does not attenuate the flicker of the superimposed fields below that of the distortion product. This is true even when the superimposed fields are flickered at rates that are much harder to detect than that of the difference frequency. This would not occur if all or even most of the attenuation of temporal fluctuations followed the nonlinear process. Hence, at least some of the attenuation of temporal frequencies must precede the nonlinear process, and so stage 12 (called a *choke* to save space) is introduced to represent it in Figure 7. (No significance should be attached to its position here relative to stages 13 to 17, for our present observations do not allow us to localize it in the sequence of these processes.) We cannot say whether additional temporal filtering follows the nonlinear stage, but at present we have no evidence of it. To provide for that possibility, however, stage 22 is introduced.

Surround antagonism (stages 13 and 14). This is discussed above.

Aftereffects of light exposure (adaptation) (stage 15). If one of the two interference fringes used to create these distortion products is presented as an intense 10-msec flash, subsequent presentation of the other fringe alone produces distortions similar to those observed when the two are presented simultaneously (MacLeod et al., 1985). Since the contrast sensitivity and hence the spatial spread of the aftereffect of the flashed grating are the same as those measured by steadily presented gratings, the residual physiological change of the adaptation produced by and outlasting the flashed grating is localized in the pathway between the receptor aperture and the site of the nonlinear stage (i.e., before the site of summation of signals from multiple cones). We have no conclusive evidence on whether it is subject to spatial antagonism (lies before or after stage 14); nor do we know whether the aftereffect sums with the effects of the steadily presented grating or attenuates it, say, by changing the gain of stage 16 or some other gain control. However, Hayhoe's experiments on sensitivity to test flashes against backgrounds of varying size (Hayhoe, 1979; Buss et al., 1982) lead us to place it tentatively at stage 15, after spatial antagonism (stage 14) and before the gain control mechanism (stage 16).

Adaptive gain control with temporal filter. Contrast sensitivity for the distortion products is constant as mean intensity varies from 240 to 10,000 Td, and if the effects of noise (stages 2 and 11) are taken into account, the sizes of the distortion products are inferred to be constant down to 1.7 Td (Makous et al., 1985). This invariance of contrast sensitivity we attribute to a gain control mechanism that keeps the time-average input to the nonlinear stage constant, and stage 16 is introduced to represent it in Figure 7. The gain control adjustment cannot be instantaneous, or no signal would be passed on the rest of the system. Hence, the low-pass temporal filter is introduced at stage 17. The gain control mechanism is tentatively positioned following spatial antagonism because of Hayhoe's demonstration (Hayhoe, 1979; Buss et al., 1982) of spatial antagonism preceding a gain of control mechanism that might be the same as that responsible for the effects observed here.

The Nonlinearity

As discussed above, the distortions would not be seen if there were no nonlinear process (Burton, 1973), which is designated as stage 18 (Figure 7). The form of the nonlinearity is described by a third-order polynomial, the coefficients of which are determined from three measurements (Makous et al., 1985): sensitivity to single gratings, which almost exclusively determines the linear parameter (a , in the equation below); sensitivity to distortions produced by two superimposed gratings of the same frequency, which

determines the value of the quadratic parameter (b below); and sensitivity to distortions produced by two superimposed gratings, one with a frequency twice that of the other, which determines the cubic component (c below). As no higher-order distortion products can be detected, no higher-order terms are required in the polynomial. Computations confirm that this polynomial produces distortions of equal size under the three threshold conditions listed above. This function is shown in Figure 8, the equation being: $R = aI + bI^2 + cI^3$, where $a = 65.0$, $b = -57.0$, and $c = 82.2$.

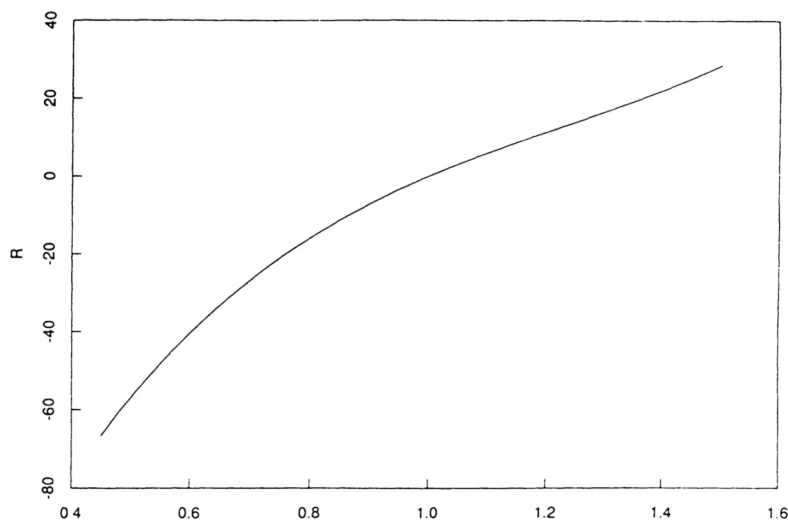


Figure 8

The form of the nonlinearity. I is scaled so that the mean intensity (1000 Td) is unity. R is scaled in arbitrary units, with zero response set at the mean intensity. 0: WM.

The third-order polynomial is the only function we have found that does not produce the distortions we observe. Its shape is independent of spatial frequency where we (Chen and Makous, unpublished observations, 1987) have been able to test it (above 30 c deg^{-1}), and it is independent of mean intensity above 1000 Td. Control tests show that the amplitudes of the distortions are proportional to the product of the contrasts of the superimposed beams, as required by theory (MacLeod et al., 1985).

Processes Following the Nonlinear Stage

The stages (19 to 23) following the nonlinear stage are in arbitrary sequence, and as each of these stages almost certainly represents many different processes, their sequences may be interlaced, parallel, and/or recursive. The first three have been discussed above. The last is taken up briefly in the next paragraph.

Generalizing the Approach

The results summarized by [Figure 7](#) suggest that use of the Rochester interferometer to exploit a nonlinear process has been rewarding, and working out the remaining unknowns in this system promises continuing rewards, but these successes also stimulate an interest in testing the power of analogous techniques elsewhere in the visual system. This would require a method for bypassing the present nonlinear process (stage 18) and identification of a different nonlinear process susceptible to such exploitation. Neither obstacle seems insurmountable.

Relations to Anatomy and Physiology

Inferences about the organization underlying behavior gain explanatory and predictive power if the inferences can be related to anatomical structures and physiological processes. Although [Figure 7](#) is essentially a way to organize a set of psychophysical observations, it also aids the effort to relate these observations to anatomy and physiology.

The physical referents for stages 1 and 2, the locus of temporal filtering (stages 12 and 22), and the stages following the nonlinearity are discussed above. Stages 3 through 11 patently describe the properties of a large population of cones. Although there are three classes of cones within the retinal region tested, the experiments represented by stage 10 show that in these experiments all processes preceding the nonlinear stage represent the properties of the *L* cones.

Horizontal cells mediate spatial antagonism (Baylor et al., 1971; Naka, 1971), such as is represented by stages 13 and 14, and the spatial extent of the antagonism observed here is close to the size of H1 horizontal cells (Boycott et al., 1987). However, an additional contribution from more proximal cells, such as amacrine cells, cannot be excluded at present.

Little can be said about the locus of bleaching adaptation (stage 15) except that it precedes the nonlinear stage (stage 18). A gain change of 5000 to 1 (stages 16 and 17) also precedes the nonlinear stage. This exceeds by more than 100-fold that observed in the membrane potentials (Valeton and Van Norron, 1983) and membrane currents (Schnapf, 1988) of primate

cones. Whatever accounts for this gain change, it occurs too far proximal to affect the electrical properties of cone membranes. However, as the release neurotransmitter from a cone can be influenced without affecting the electrical properties of the cone (Sarantis et al., 1988), one cannot exclude the site on cones where neurotransmitter is released from the locus of the gain change.

Locus of the Nonlinearity

The nonlinear process is the point of reference in the sequence of operations, and identification of its anatomical locus is necessary to determine the locus of many of the other processes. Once this locus is determined, correspondences between the electrophysiologically observed properties of neurons and psychophysically observed operations can be established, and the information present or lost in the two domains can be compared. If the two correspond closely enough, the functional organization of the human visual system can be determined by comparisons across species, and parameters specific to the human species can be determined by psychophysics. The locus of the nonlinear process is not yet narrowly constrained.

As the amplitude of every physiological process is limited, all must enter a nonlinear stage at least when driven hard. This makes it hard to conceive that the signal passes very far without encountering a nonlinear stage. As discussed above, the particular nonlinearity producing the distortions we observe under these conditions (stage 18) is preceded by spatial antagonism (stage 14), adaptation, temporal filtering, and gain adjustment, but this excludes only sites distal to cone neurotransmitter release.

The small aperture preceding the nonlinear stage, corresponding to that of a single cone (stage 3), means that no pooling of signals from other cones precedes the nonlinearity. However, this is true of the midget or parvocellular system of the fovea at least as far as the lateral geniculate nucleus and probably at the cortex as well. So if these observations can be mediated by the parvocellular system, this hardly constrains the locus of the nonlinear process at all.

Although the small aperture (stage 3) fails to constrain the locus of the nonlinear process, it does limit applicability of observations on processes preceding the nonlinear stage to individual cones and the parvocellular system. Even the parvocellular system, however, is heterogeneous, including, for example, both on and off pathways. If both on and off systems contribute to observations of processes preceding the nonlinearity, their respective contributions are not known.

Form of the Nonlinear Process

As the distortions observed psychophysically presumably depend on a physiological process with nonlinear properties, the simplest assumptions suggest that the very distortions observed should be produced by transforming each point in the stimulus pattern according to the mathematical function that describes the physiological process. However, none of the functions that have been used to describe physiological processes so serve. Evidently, the assumptions underlying the expectation that they would serve are too simple. Complications arise (1) from the multiplicity of dissimilar but interrelated nonlinear physiological processes; (2) from the differences between the response of individual members of a population and the statistical properties of the population as a whole; and (3) from subsequent processing, such as those that transform continuous physiological processes into discrete behavioral responses.

Multiplicity of processes. Each cell has several observable properties that contain visual information. Nearly all such processes enter a nonlinear phase as they approach saturating limits, and some have nonlinearities near threshold, but no two need necessarily be simply related to one another.

The input to a bipolar cell, for example, is a change of concentration of a neurochemical, but there may be several such neurochemicals in the outer plexiform layer that affect the cell. The concentration of any one may be a nonlinear function of stimulus intensity, though not necessarily the same function.

Each neurochemical affects the properties of ionic channels in the postsynaptic membrane, and perhaps the properties of the presynaptic membrane as well, but different transmitters have different effects on different channels, and a given channel has different effects on the conductance of different ions. In aggregate these ionic conductances affect membrane current and membrane potential at the presynaptic membrane of the bipolar cell in the inner plexiform layer. This presynaptic potential is determined by a weighted average of currents through different types of channels, in different regions of the cell, over different courses of time; by electrical coupling among cells; and by all the influences described for the outer plexiform layer. So the concentrations of neurochemicals and the conductance, current, and potential of the synaptic membrane each play their respective roles in carrying the visual information. Although these are all interrelated, their interrelations are seldom linear, and so each typically is related to stimulus intensity by a different function. It seems probable that each of these participate in carrying the visual information through the system and that none is related to it in a simple way.

Population statistics. An individual distortion grating is created by the concerted action of many individual nonlinear elements. Unless each follows the identical nonlinear function, the amplitude and detectability of the grating may be different from the contribution of any individual element. For example, if each element had an all-or-none response but different thresholds, the average response would be graded. So the individual physiological functions that create the distortions can be quite different from their collective effect.

Subsequent processing. If there were a particular physiological process related to the psychophysically observed nonlinearity, it would in any case be a continuous, time-varying signal. But psychophysical thresholds are discrete: yes or no; alternative 1 or 2. Then what aspect of the time-varying signal determines the psychophysical response? Whatever is chosen—whether the favorite, *peak* response, or something else, such as integrated response—subsequent temporal filtering has different effects on responses that do not have the same time course. Normann and Werblin (1974) reported, for example, that the exponent of the Naka-Rushton equation (Naka and Rushton, 1966) describing peak responses of mudpuppy receptors is 0.7 for long flashes and 1.0 for short flashes.

So responses to stimuli that are both at threshold may differ, according to *any* measure, before the final stage of temporal filtering (cf. Zacks, 1970), and one cannot know how to relate physiological responses to psychophysics without knowing how the signals are processed between the site observed and the site where the response is determined.

It is worth adding here that this has nothing to do with psychophysical linking laws. However valuable the concept may have proved elsewhere, here it is a red herring. When one associates the squeeze of the trigger of a rifle with the emission of hot gases from the barrel, it is not necessary to invoke chemophysical linking laws; it is satisfactory to describe the process as a causal sequence. The same is true of the experiments discussed here.

ON METAPHOR IN VISION

An essay such as this perhaps allows one to indulge in broader commentary than might otherwise be permissible, and I would like to use this platform to comment on the metaphors that visual scientists use. Those I have used here are borrowed from electronics and communication theory. I think they are the best we have for this work, and they serve because the visual system and the artificial systems for which they were developed are subject to similar demands. Yet natural and artificial systems do differ, and so in some ways the metaphors are inapt.

For example, functional diagrams such as [Figure 7](#), and others that visual scientists use to represent neural circuits, show discrete elements interrelated by discrete connections. I wonder if they do justice to the variety of cells represented and the complexity and variations in the interconnections among them and the systems they form.

To some extent the schematic diagrams we use arise from the idea that individual neurons are the appropriate units of analysis in the visual system. However, the ubiquitous gap junctions spread electrical potentials through the barriers between cells that guard their independence, so that the cooperative actions of these extensively interconnected neurons generate fields that tend to vary continuously in space, perhaps more in accordance with the rules of field theory than those of the neuron doctrine. When 90 percent of the signal issuing from the individual rod originates in other rods (Fain, 1975), the idea of individual cells as independent elements is strained. On the other hand, metaphors derived entirely from field theory go too far toward ignoring the discontinuities produced by the membranes that *are* there.

Block diagrams such as [Figure 7](#) are also clumsy in representing multiple local interactions and feedback. They do not adroitly treat the cooperative action of large interacting populations of diverse cells.

This is not to say that these things cannot or are not being done; it is to say that they are done neither adroitly nor with grace; and when they are done, the concepts are adopted from other sciences, not born of the peculiarities of the phenomena we study.

I end this essay with comments on an altogether different kind of metaphor, such as Pugh's handsome and informative portrait of the transduction process inside photoreceptors (Pugh, this volume). It deftly encapsulates a complex and dynamic set of interrelationships. Yet these are, and for the present must remain, static representations of a dynamic system. And, for once, I think they make the possibilities of electronic journals attractive, for in a journal viewed through a computer, the processes and cycles now imitated by arrows could be set into motion. When such illustrations incorporate the models themselves, the possibilities enlarge. We shall be able to play with the variables and watch gears turn as the model grinds out the results. Perhaps then models will be fun for readers as well as modelers, and the distinction between the two may blur.

ACKNOWLEDGMENTS

Preparation of this paper was supported by Public Health Service grants EY-4885 and EY-1319.

References

- Banks, M.S., W.S. Geisler, and P.J. Bennett 1987 The physical limits of grating visibility. *Vision Research* 27:1915-1924.
- Baylor, D.A., M.G.F. Fuortes, and P.M. O'Bryan 1971 Receptive fields of cones in the retina of the turtle. *Journal of Physiology* (London) 214:265-294.
- Born, M., and E. Wolf 1970 *Principles of Optics*, 4th edition. New York: Pergamon Press.
- Boycott, B.B., J.M. Hopkins, and H.G. Sperling 1987 Cone connections of the horizontal cells of the rhesus monkey's retina. *Proceedings of the Royal Society of London B* 229:345-379.
- Burton, G.J. 1973 Evidence for non-linear response processes in the human visual system from measurements on the threshold of spatial beat frequencies. *Vision Research* 13:1211-1225.
- Buss, C.M., M.M. Hayhoe, and C.F. Stromeyer III 1982 Lateral interactions in the control of visual sensitivity. *Vision Research* 22:693-709.
- Campbell, F.W., and D.G. Green 1965 Optical and retinal factors affecting visual resolution. *Journal of Physiology* 181:576-593.
- Campbell, F.W., and R.W. Gubisch 1966 Optical quality of the human eye. *Journal of Physiology* 186:558-578.
- Chen, B., and W. Makous 1987 Retinal losses of contrast with oblique incidence. *Investigative Ophthalmology and Visual Science* 28(Suppl.):357.
- 1988 Light capture by cones. *Journal of Physiology*, submitted.
- Chen, B., W. Makous, and D.R. Williams 1988 Serial spatial filters in vision. *Investigative Ophthalmology and Visual Science* 29(Suppl.):58.
- Curcio, C.A., K.R. Sloan, Jr., O. Packer, A.E. Hendrickson, and R.E. Kalina 1987 Distribution of cones in human and monkey retina: individual and radial asymmetry. *Science* 236:579-582.
- Fain, G.L. 1975 Quantum sensitivity of rods in the toad retina. *Science* 187:838-841.
- Graham, N. 1977 Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Research* 17:637-652.
- Hayhoe, M. 1979 After-effects of small adapting fields. *Journal of Physiology* 296:141-158.
- MacLeod, D.A.I., D.R. Williams, and W. Makous 1985 Difference frequency gratings above the resolution limit. *Investigative Ophthalmology and Visual Science* 26(Suppl.):11.
- Makous, W., D.R. Williams, and D.A.I. MacLeod 1985 Nonlinear transformation in human vision. *Journal of the Optical Society of America A* 2:P80.
- Miller, W.H., and G. Bernard 1983 Averaging over the foveal receptor aperture curtails aliasing. *Vision Research* 23:1365-1369.
- Naka, K.I. 1971 Receptive field mechanism in the vertebrate retina. *Science* 171:691-693.

- Naka, K.I., and W.A.H. Rushton 1966 S-potentials from colour units in the retina of fish (*Cyprinidae*). *Journal of Physiology* (London) 185:536-555.
- Normann, R.A., and F.S. Werblin 1974 Control of retinal sensitivity: I. Light and dark adaptation of vertebrate rods and cones. *Journal of General Physiology* 63:37-61.
- Pelli, D.G. 1985 Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A* 2:1508-1532.
- Pokorny, J., and Smith, V.C. 1986 Colorimetry and color discrimination. Pp. 8-1 to 8-51 in *Handbook of Perception and Human Performance*, K.R. Boff, L. Kaufman, and J.P. Thomas, eds. New York: Wiley.
- Sarantis, M., K. Everett, and D. Attwell 1988 A presynaptic action of glytamate at the cone output synapse. *Nature* 332:451-453.
- Stiles, W.S., and B.H. Crawford 1933 The luminous efficiency of rays entering the pupil at different points. *Proceedings of the Royal Society* (London) B112:428-450.
- Valeton, J.M., and D. Van Norron 1983 Light adaptation of primate cones: an analysis based on extracellular data. *Vision Research* 23:1539-1547.
- Westheimer, G. 1986 The eye as an optical instrument. Pp. 4-1 to 4-20 in *Handbook of Perception and Human Performance*, K.R. Boff, L. Kaufman, and J.P. Thomas, eds. New York: Wiley.
- Williams, D.R. 1985a Aliasing in human foveal vision. *Vision Research* 25:195-205.
- 1985b Visibility of interference fringes near the resolution limit. *Journal of the Optical Society of America A* 2:1087-1093.
- Zacks, J.L. 1970 Temporal summation phenomena at absolute threshold: their relation to visual mechanisms. *Science* 170:197-199.

SPATIAL EFFECTS OF PHOTORECEPTOR MOSAIC

The Distribution of Cones in the Primate Retina

V. Hugh Perry

Every student of vision is aware that the sampling of an image by the photoreceptors imposes important limitations on the information that is available at every subsequent stage of processing along the visual pathways. Despite the importance of knowing the receptor distribution, the actual data base for the primate retina has been remarkably small until the recent publication of a number of papers that have sought to rectify this and that have brought to light new and interesting points. This paper will bring together some of the main points, but no attempt is made to give all the quantitative values since these are readily available in the primary sources.

TISSUE PREPARATION

Assessing the distribution of photoreceptors across the retina would appear to be a relatively trivial problem. However, it must be recognized that the preparation of any tissue for microscopic examination usually involves some sort of processing, and this processing commonly results in shrinkage of the tissue. Unless the shrinkage is taken into account by making appropriate measurements at the various stages of preparation, the value of receptor counts from such material is limited. The fact that the receptor density changes dramatically with eccentricity also requires that the sampling eccentricity be exactly known. The use of whole-mounted retinas greatly reduces many of the problems associated with using sectioned retinas and allows accurate estimates of retinal shrinkage. The identification of rods and cones in the primate retina poses few problems (Figure 1).

The classic study of Osterberg (1935) described the distribution of rods and cones over a large part of a single human retina. Polyak (1957) described the density of cones in monkey (*M. mulatta*) and human retinas

but largely restricted his observations to the fovea and perifovea. More recent studies on a small number of retinas have also been restricted to the fovea or central few degrees (Miller, 1979; de Monasterio et al., 1985). The foveal cone density has attracted much attention, in no small part because of the interest of many visual scientists in measurements of threshold vision. What was required was a study of cone density distribution in a sufficient number of retinas to give an idea of the variability between individuals. Rolls and Cowey (1970) examined a number of retinas from *M. mulatta* and *S. sciureus* along the horizontal meridian, but they used vertical sections of retina and were well aware of the problems in extracting accurate estimates from such material.

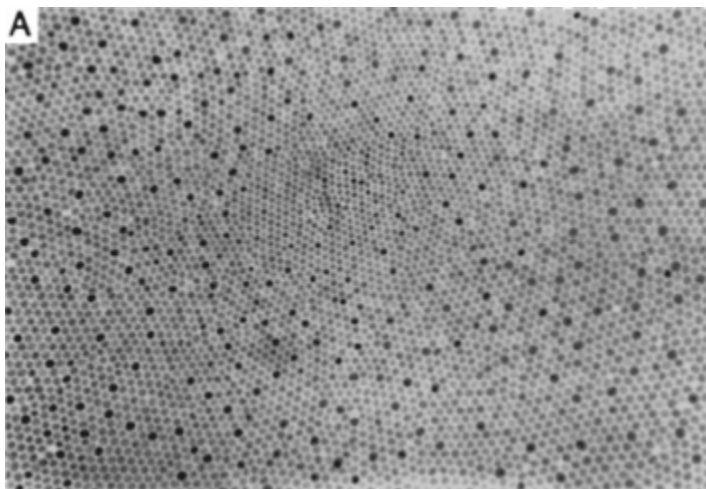


Figure 1A
Photomicrograph of foveal cones in a horizontal section through the inner segments. The retina is from an *M. fascicularis* monkey. The field is a little more than 1 deg across (201 $\mu\text{m}/\text{degree}$ from Perry and Cowey, 1985).

My co-workers and I examined, in whole-mounted retinas, the distribution of cones along the horizontal and vertical meridians in four retinas from *M. mulatta* (Perry and Cowey, 1985). In these preparations the shrinkage was minimal. Estimates of the foveal cone density were made from horizontal sections that passed through the inner segments. The shrinkage of the sections was known. In one case the entire retina was sampled except for the fovea itself. A number of important points came out of this study. The cone density changes very rapidly over the central 1 deg of the visual field (Figure 1A) and continues to decline rapidly over the central

5 degrees (Figures 1B and 1C). The change in the size of the cones and the increasing number of rods are easily seen in these photomicrographs. There was a clear nasotemporal asymmetry in the cone distribution outside the central 10 deg or so of the retina; the cone density declined more slowly along the nasal horizontal meridian of the retina than along the other axes. The asymmetry in cone density was somewhat smaller than that seen for the ganglion cells (Perry et al., 1984). This asymmetry is consistent with psychophysical observations of acuity and sensitivity being superior in the temporal field. In addition, we found differences in foveal cone density between species (see also Rolls and Cowey, 1970; Borwein et al., 1980). While these data provide a useful basis for comparison with physiological and psychophysical studies on monkeys, it remained to be shown how the cone density varied in the primate for which we have so much psychophysical data—the human.

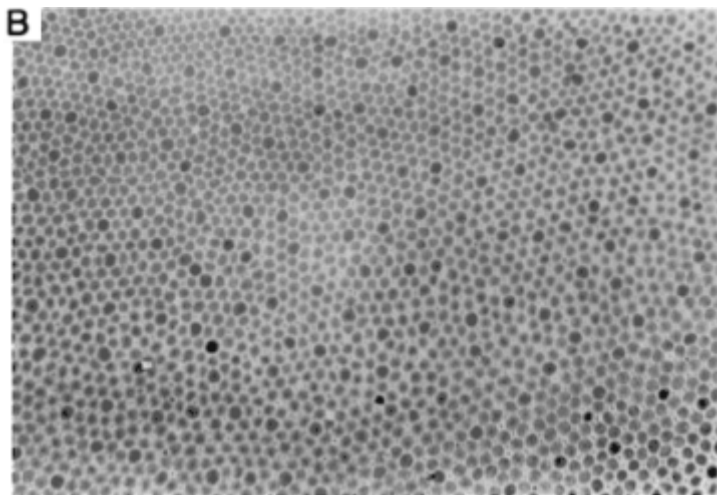


Figure 1B
A field centered on 1 deg from the center of the fovea (M. fascicularis).

CONE DENSITY

Curcio et al. (1987b) have provided the first complete maps of the retinal cone distribution in humans. The same laboratory has also shown that the density of cones varies with age (Yuodelis and Hendrickson, 1986). Bearing in mind that the number of retinas examined is small, Yuodelis

and Hendrickson have shown that the foveal cone density in the newborn is relatively low; density then increases for at least the next 4 or more years. The density may fall again in aged subjects. It is believed that the increase in cone density in the postnatal period is accomplished by the migration of cones toward the center of the fovea (Hendrickson and Kupfer, 1976; Yuodelis and Hendrickson, 1986). The loss of cones in the aging eye may be a result of cell death.

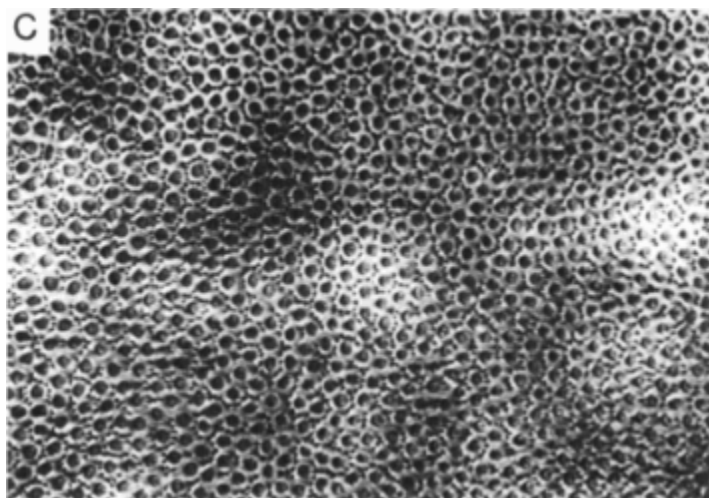


Figure 1C

A field centered on a point 5 deg from the center of the fovea (*M. mulatta*). The estimated linear shrinkage of A and B during processing was about 15 percent; no correction has been applied to the scaling. Scale bar = 25 μ m.

Not only is there a change in cone density with age, Curcio et al. (1987b) have shown that the foveal density may vary by as much as a factor of 3 in normal adult human eyes. A sample of four retinas from subjects aged 27 to 44 years was studied. There was considerably less variation in density outside the central 1 deg. The cone distribution in the human eye has a similar but not as pronounced nasotemporal asymmetry as described for the monkey with the higher densities in the nasal retina.

The large variation in foveal cone density cannot be readily attributed to differential tissue shrinkage. The method of clearing the tissue and viewing it under Normaski optics greatly reduced processing distortion (Curcio et al., 1987a) and the authors estimate the shrinkage as varying from 2 to 12 percent. It could be argued, however, that the large variation in cone density was due to the fact that human eyes are not always obtained

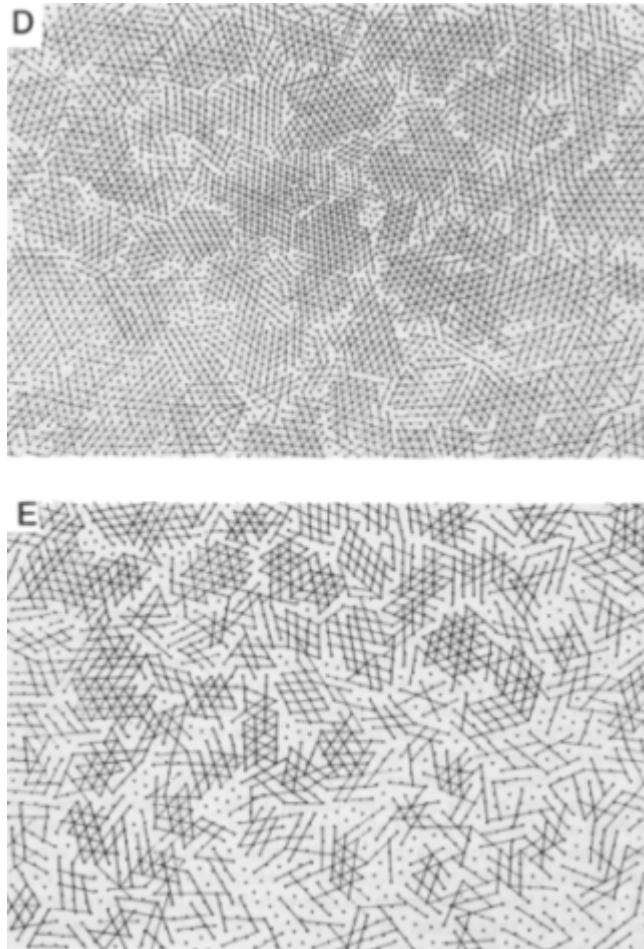


Figure 1D, E
Analysis of cone mosaics to demonstrate the local triangular lattice at the fovea
and how it degenerates with eccentricity. See text for details.

under conditions for tissue histology, and it is generally agreed that the fovea is particularly vulnerable to postmortem artefacts. We therefore undertook a study of eyes from *M. fascicularis*, where the eyes could all be obtained immediately postmortem following fixation in situ by perfusion (Hawken et al., 1988). Each eye was subsequently processed in as near identical fashion as possible following the method of Curcio et al. (1987a). Counts of the cone density were made throughout the central degree and at selected intervals out to the periphery. There would appear to be no a priori reason why the size of the sampling area chosen for counting should be of any particular dimensions. In [Figure 1A](#) it is clear that the density is changing very rapidly across the central 1 deg. The sample area used by Perry and Cowey (1985) of about 0.45×0.45 deg was almost certainly slightly too large. Some of the variability among different studies can, in part, be accounted for by differences in the sampling area. Pokorny (1968) has examined how grating acuity changes with the size of the target viewed, and his results suggest that an area about 0.125×0.125 deg would be appropriate. This was the area we used.

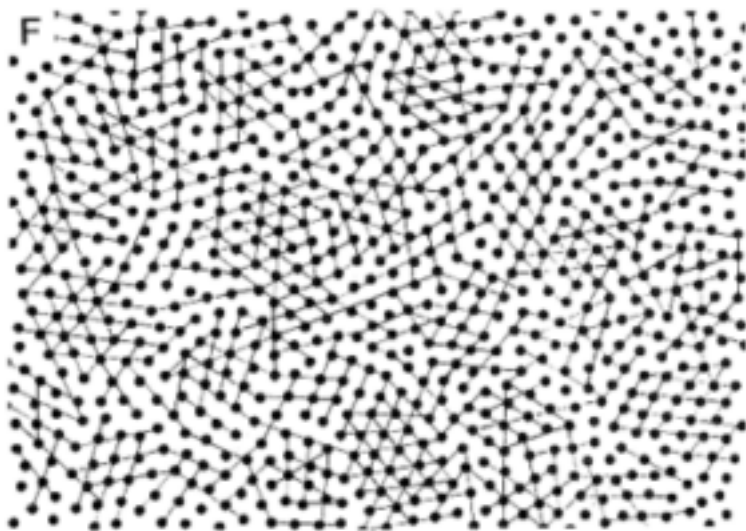


Figure 1F
Analysis of cone mosaics to demonstrate the local triangular lattice at the fovea and how it degenerates with eccentricity. See text for details.

A sample of our data is shown in [Figure 2](#), where the cone density along the temporal horizontal meridian in eight retinas is plotted. In six of the retinas there is a rapid decline in cone density with eccentricity, as we

expected from previous studies. The variation is similar in these six retinas across the eccentricities we have examined. In contrast, two of the retinas (\blacktriangle and \diamond) showed quite clear reductions in density within the central 0.5 deg. In these animals we could observe no pathology or distortion of the tissue, and indeed it was clear that the inner segments of the cones were themselves larger than in the retinas with higher densities. If these low counts resulted from distortion of the tissue, we might expect to see anomalies at further eccentricities; this was not the case. Despite having low densities in the foveal center, these two animals would still theoretically have a cone density at 0.5 deg, sufficient to resolve about 30 cycles/degree. The highest and lowest foveal counts give about a sixfold range in peak density. It would be interesting to know how these anatomical results compared with behavioral measures of acuity in *M. fascicularis*, but unfortunately no large sample is available. However, behavioral measures of acuity in *M. mulatta*, which

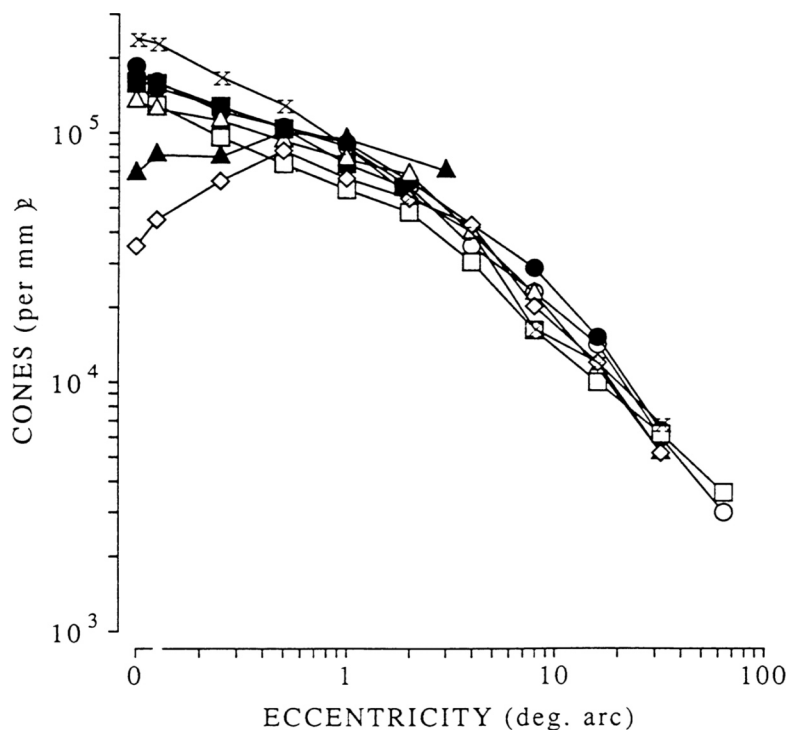


Figure 2
Density of cones at the fovea and along the temporal horizontal meridian in retinas from *M. fascicularis* monkeys.

were refracted and optically corrected, showed considerable variation (A. Cowey, personal communication).

Thus, we are able to confirm the results of Curcio et al.—namely, that cone density varies substantially at the fovea between individual eyes. The reason that some animals have a low density is unclear, but it may well be that the late migration of cones into the foveal region (Hendrickson and Kupfer, 1976; Yuodelis and Hendrickson, 1986) is particularly susceptible to environmental, nutritional, and mechanical factors. How the variation in the foveal cone density is related to the refractive state of the eye was, unfortunately, not examined in this study but certainly merits further study.

The anatomical data suggest that we can give no single value as the cone spacing at the human fovea. Ideally, we would like to look at the cone separation in the living eye and compare the results obtained with those found in whole-mount preparations. Using laser interferometric methods, it is possible to do just this (Williams, 1985, 1988). Using this technique the data suggest that there is rather little variability in cone spacing. The discrepancy between the anatomical and psychophysical results may be more apparent than real at this time, since only small numbers of subjects have been examined in both instances. Williams (1988) does report that at least one subject found the moiré patterns produced by the interference fringes difficult to see, which may reflect some anomaly in this subject's cone packing.

IMAGE SAMPLING

The way in which cones are packed at the fovea is also a matter of some interest since positional disorder in the elements sampling an image will degrade the high frequencies in an image (French et al., 1977; Hirsch and Hylton, 1984). Once again, for this analysis we chose only those preparations where the distortion was minimized. Figure 1A shows that the cones are regularly packed but gives the impression of short-range order and little long-range order. This is simply demonstrated in Figure 1D. A spot was placed in the center of each cone shown in Figure 1, and a straight line was drawn through as many spots as could be joined. If the cone mosaic was perfectly regular, the lines would cross the entire field. This is clearly not the case; the cone mosaic divides into "islands" in which the cones are arranged in a triangular lattice with fractures between the islands. The largest islands are about 10×20 cones. We adopted the same procedure for cones centered on a region 1 and 5 deg from the center of the fovea. Larger spots were used at 5 deg to allow for the fact that the cones themselves were substantially larger than at the fovea. The regularity of the mosaic clearly degenerates with eccentricity. The islands are less obvious, and the number of cones contributing to each of them is greatly

reduced. Williams (1988) has shown that in the human eye, again using interference fringes, the orientations of the moiré patterns observed are of a form consistent with triangular packing of the cones with local but not long-range order. The extent of the regular packing predicted by his studies is in good agreement with the size of the largest islands and the rapid degeneration of the regularity with distance from the foveal center.

If the cones are arranged in a crystalline lattice in the fovea, it would be of interest to know whether the different types of cones long-, medium-, and short-wavelength sensitive (here referred to as red, green, and blue cones) are also distributed in regular patterns. The blue cones, which can be identified by staining with Procion dyes, are organized in a regular lattice (de Monasterio et al., 1985). Unfortunately, there is no simple anatomical method for discriminating between red and green cones in the primate. The only study that successfully differentially labeled the red and green cones in a primate was in the baboon retina (Marc and Sperling, 1977). Their results showed that the two types of cones were randomly arrayed.

It has been suggested that not only are the red and green cones randomly arrayed but also that the cells of the inner retinal layers are not able to distinguish between the two cone types during development (Shapley and Perry, 1986). How then are color-opponent ganglion cells generated? We hypothesize that a midget bipolar cell contacting a single cone and connected to a single one of the smallest P ganglion cells (so-called midget ganglion cells; Polyak, 1941) would provide the central input to a cell with color-opponent receptive field. The surround of this cell might receive inputs from both red and green cones and, depending on the relative proportions of them, would appear as a more or less color-opponent cell. Across the whole population of color-opponent P ganglion cells, the neutral point to these cells would vary, as is found to be the case (Zrenner, 1983). It will be of considerable interest to know just how and in what proportions these two cone types are distributed in the retina and whether this is consistent with this the "hit-or-miss" hypothesis.

At the fovea the cells of the inner nuclear and ganglion cell layers are laterally displaced; in order that the cones might contact the cells of the inner nuclear layer, there is a long process, known as a fiber of Henle, that courses in the outer fiber layer and terminates in a pedicle to make synaptic contact in the outer plexiform layer (Polyak, 1941). The length of these fibers and their arrangement around the fovea will not only be important for the preservation of the sampled image but will also determine the numerical relationship between the number of cones and ganglion cells sampling a given point of visual space.

The fibers of Henle of foveal cones in the monkey retina are at least 400 μm in length (Boycott et al., 1987; Perry and Cowey, 1988). Cones immediately adjacent to the central foveal cones have slightly shorter fibers

and so on, such that the length of the fibers decreases in a systematic fashion from the fovea out to about 15 deg eccentricity. The fibers radiate out from the center like the spokes from a wheel, and it is clear that adjacent cones have very similar lengths (Perry and Cowey, 1988), so the topography of the cone mosaic at the level of the inner segments is preserved at the pedicles. There is, however, an interesting problem for those cones in the most central part of the fovea since the fibers of these cones project to all four quadrants around the fovea. Initially the cells to which these fibers are connected lay over the foveal center and migrated laterally during development; thus, it would appear that the topography was preserved. However, while this is in large part correct, there remains the problem that some fibers pass to either side of the vertical meridian. Ganglion cells on either side of the meridian send their axons to opposite sides of the brain. This poses a problem as to how the image is then retrieved, particularly with high spatial frequencies since these are poorly transmitted across the corpus callosum (Bernardi et al., 1987). Williams (1988) has found that in some observers the cone spacing is larger in the horizontal direction than in the vertical direction. It seems possible that this may be related to the behavior of the fibers of Henle.

GANGLION/FOVEAL CELL RATIO

From our knowledge of the length of the fibers of Henle or by taking into account the fact that cone pedicles are at least twice the diameter of the cone inner segment (Boycott et al., 1987), it is possible to estimate the numerical relationship of ganglion cells to cones for the foveal retina. The present estimates show that the ratio of ganglion cells to cones is slightly greater than 2:1 for the central 1 deg (Perry and Cowey, 1988; Schein, 1988). It has been shown that each cone pedicle is contacted by at least two bipolar cells, which themselves terminate at different levels in the inner plexiform layer (Kolb et al., 1969). From what is known of the organization of other retinas, it seems likely that one of these bipolar cells contacts an ON-center ganglion cell, the other an OFF-center cell. A 2:1 ratio may be of little advantage with respect to the spatial sampling but may be an advantage for producing a system with a large dynamic range.

The foveal cones have a fiber of Henle of the order of 400 μm long and about 2 to 3 μm in diameter. The conduction of a potential along such a process will result in both a decrement and temporal smearing of the signal. This is not likely to be an advantage. Polyak (1941) pointed out that Muller's fibers, the radial glia of the retina, appear to partially wrap themselves around the fibers of Henle. We have recently shown that adjacent Muller's fibers and fibers of Henle have very similar lengths at the same retinal eccentricities. It would appear that the Muller's fibers may act

as ensheathing glia and thus help to prevent the decrement of the signal as it passes from the cell body to the pedicle.

A description of the distribution of the cones in the monkey and human eye is now available. This information should prove useful in showing the extent to which cone density limits visual performance across the visual field. The next important task is to discover from anatomical and physiological experiments how the information from the cones is used to construct the receptive field characteristics of the different types of ganglion cells.

References

- Bernardi, N., S. Bisti, and L. Maffei 1987 The transfer of visual information across the corpus callosum: spatial characteristics. *Journal of Physiology* (London) 384:619-632.
- Borwein, B., D. Borwein, J. Medeiros, and J.W. McGowan 1980 The ultrastructure of monkey fovea photoreceptors, with special reference to the structure, shape, size and spacing of foveal cones. *American Journal of Anatomy* 159:125-146.
- Boycott, B.B., J.M. Hopkins, and H.G. Sperling 1987 Cone connections of the horizontal cells of the rhesus monkey's retina. *Proceedings of the Royal Society B* 229:345-379.
- Curcio, C.A., O. Packer, and R.E. Kalina 1987a A whole mount method for sequential analysis of photoreceptor and ganglion cell topography in a single retina. *Vision Research* 27:9-15.
- Curcio, C.A., K.R. Sloan, Jr., O. Packer, A.E. Hendrickson, and R.E. Kalina 1987b Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* (New York) 236:597-582.
- de Monasterio, F.M., E.P. McCrane, J.K. Newlander, and J.S. Schein 1985 Density profile of blue sensitive cones along the horizontal meridian of macaque retina. *Investigative Ophthalmology and Visual Science* 26:289-302.
- French, A.S., A.W. Snyder, and D.G. Stavenga 1977 Image degradation by an irregular retinal mosaic. *Biological Cybernetics* 27:229-233.
- Hawken, M.J., V.H. Perry, and A.J. Parker 1988 Structural relationships of photoreceptors to VI receptive fields in primate. *Invest. Ophthalm. Vis. Sci.* 29(Supp.):297.
- Hendrickson, A., and C. Kupfer 1976 The histogenesis of the fovea in the macaque monkey. *Investigative Ophthalmology* 15:746-756.
- Hirsch, J., and R. Hylton 1984 Quality of the primate receptor lattice and limits of spatial vision. *Vision Research* 24:347-356.
- Kolb, H., B.B. Boycott, and J.E. Dowling 1969 A second type of midget bipolar cell in the primate retina. *Philosophical Transactions of the Royal Society B* 225:177-184.
- Marc, R.E., and H.G. Sperling 1977 Chromatic organization of primate cones. *Science* 196:454-456.

- Miller, W.H. 1979 Ocular optica filtering. In *Handbook of Sensory Physiology*, vol. V11/6A. Berlin: Springer.
- Osterberg, G. 1935 Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica* Supplement 6:1-103.
- Perry, V.H., and A. Cowey 1985 The ganglion cell and cone distributions in the monkey's retina; implications for central magnification factors. *Vision Research* 25:1795-1810.
- 1988 The lengths of the fibres of Henle in the retina of macaque monkeys: implications for vision. *Neuroscience* 25:225-236.
- Perry, V.H., R. Oehler, and A. Cowey 1984 Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. *Neuroscience* 12:1101-1023.
- Pokorny, J. 1968 The effect of target area on grating acuity. *Vision Research* 8:543-554.
- Polyak, S. 1941 *The Retina*. Chicago, Ill.: University of Chicago Press.
- 1957 *The Vertebrate Visual System*. Chicago, Ill.: University of Chicago Press.
- Rolls, E.T., and A. Cowey 1970 Topography of the retina and striate cortex and its relationship to visual acuity in rhesus monkeys and squirrel monkeys. *Experimental Brain Research* 10:198-311.
- Schein, S.J. 1988 Anatomy of the macaque fovea and spatial densities of neurons in foveal representation. *Journal of Comparative Neurology* 269:479-505.
- Shapley, R.M., and V.H. Perry 1986 Cat and monkey retinal ganglion cells and their visual functional roles. *Trends in Neuroscience* 9:229-235.
- Williams, D.R. 1985 Aliasing in human foveal vision. *Vision Research* 25:195-205
- 1988 Topography of the foveal cone mosaic in the living human eye. *Vision Research* 28:433-454.
- Yuodelis, C., and A.E. Hendrickson 1986 A qualitative and quantitative analysis of the human fovea during development. *Vision Research* 26:847-855.
- Zrenner, E. 1983 Neurophysiological aspects of colour vision in primates. In *Studies of Brain Function*, vol. 9. Berlin: Springer-Verlag.

The Photoreceptor Mosaic as an Image Sampling Device

John I. Yellott

The fact that vision begins with a spatial sampling of the retinal image by discrete photoreceptors was first recognized in the 1850s, when Heinrich Müller combined psychophysics and anatomy to show that the light-sensitive elements of the retina must be the rods and cones (Brindley, 1970). That recognition touched off a debate about the perceptual implications of "mosaic vision," in particular its implications for spatial resolution. Apparently Müller himself initially argued that for two points to be seen as separate it was sufficient for them simply to stimulate two different cones (L.N. Thibos, personal communication, 1988). But his rival, Bergmann (1858), pointed out that in that case a continuous line would be indistinguishable from a row of dots! To Bergmann, and to Helmholtz (1860), it was obvious that for two points to be resolved there must be a third unstimulated cone between two stimulated ones. And on that basis Helmholtz was satisfied that the limits of visual acuity were consistent with the anatomical dimensions of the receptor mosaic: at his own resolution limit for gratings, 60 cycles/degree, alternate light and dark bars would stimulate alternate rows of cones.

Once this point was settled, the visual consequences of receptor sampling seem to have attracted surprisingly little attention for quite a long time. But in this century the general topic of signal sampling has assumed great practical importance, and mathematicians and engineers have developed sophisticated tools for analyzing the information transmission properties of arbitrary sampling schemes. Recently visual scientists have begun to apply these tools to the photoreceptor mosaic to analyze its properties as an image sampling device. This paper reviews some of what has been learned from this analysis.

For an engineer setting out to build an image sampling device, the natural starting point is Shannon's (1949) sampling theorem. I will start by

reviewing the assumptions and consequences of that theorem and then will describe the image sampling parameters of the cones and point out how they violate Shannon's assumptions. Work by anatomists and psychophysicists over the past decade has shown that the cone mosaic is quite unlike the kind of sampling device envisioned by Shannon—from his perspective its construction seems much too sloppy. I think this suggests that the receptor mosaic is not really designed with Shannon's theorem in mind (so to speak), but instead follows a different blueprint based on some other approach to recovering continuous images from discrete samples. Part of this paper deals with image recovery algorithms that do not require the kind of precise architecture assumed by Shannon's theorem and therefore seem better suited to the kind of image sampling found in the primate retina.

SHANNON'S THEOREM

We are concerned with the sampling of two-dimensional signals—retinal images—but to illustrate the basic ideas of sampling theory it is convenient to begin with one-dimensional signals, such as temporal waveforms (see [Figure 1](#)). Shannon (1949) originally stated his theorem for such signals: "If a function $f(t)$ contains no frequencies higher than W cycles/sec it is completely determined by giving its ordinates at a series of points spaced $1/2W$ seconds apart."

In other words, no information is lost by discrete sampling, provided we sample periodically (as in [Figure 1A](#)) and the highest frequency in the signal is at most one-half the sampling rate. That frequency cutoff, half the sampling rate, is called the Nyquist limit of the sampling array. The Nyquist limit is a key concept—in a moment we will see what it means to violate it.

"Completely determined" in the theorem means that the continuous input signal can be exactly reconstructed from its sample values. Shannon showed that this can be achieved by convolving the sample values with a function of the form $\sin 2\pi Wt/2\pi Wt$ —the so-called sinc function. I will refer to this process as sinc interpolation. The sinc function arises here because it produces perfect low-pass filtering in the frequency domain: sinc interpolation kills all frequencies above the Nyquist limit and passes all lower frequencies intact. So another way to describe the reconstruction process is to say that the input signal is recovered by low-pass filtering of the sample values.

To visualize the process of image reconstruction by sinc interpolation, one can imagine that each sample point creates a rippling point-spread function whose value equals the signal at that point and zero at all the other sample points, as illustrated in [Figure 1B](#). If we add up all these point-spread functions, we get the original signal.

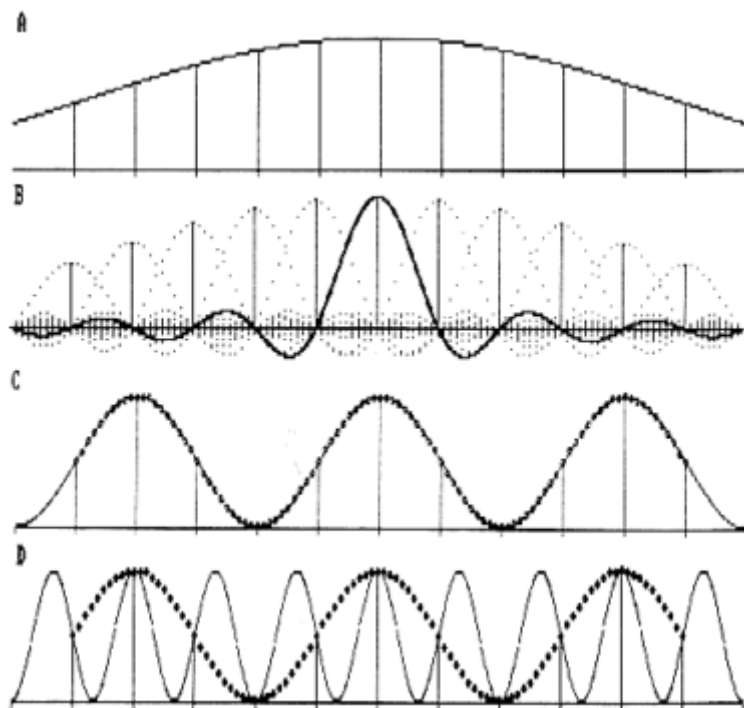


Figure 1

Shannon's sampling theorem in one dimension. NOTE: (A) Continuous signal sampled at equally spaced points. (B) Sinc interpolation functions around the sample points. (For clarity only one function is shown as a continuous curve.) (C) Postsampling reconstruction of a sinusoid at half the Nyquist frequency. The input signal is shown as a continuous curve; its reconstruction is shown as a dotted curve. (D) Aliasing: the input signal is a sinusoid whose frequency is 1.5 times the Nyquist limit of the sampling array. Sinc interpolation yields a reconstruction (dotted curve) whose frequency is one-half the Nyquist limit.

Shannon's theorem deals strictly with bandlimited signals and infinite sampling arrays. In the real world, of course, we never encounter either one. (In principle, a band-limited signal must have infinite duration.) So it is natural to wonder how well sinc interpolation works for finite-duration signals sampled by finite sets of points. Figure 1C shows a sinusoidal signal at half the Nyquist frequency that has been sampled at 11 points and reconstructed by sinc interpolation. The results here are essentially perfect,

and that is quite characteristic: Shannon's reconstruction algorithm works very well on a local scale for signals below the Nyquist limit.

Figure 1D shows what happens when the input signal exceeds the Nyquist limit. Here a sinusoid at 1.5 times the Nyquist frequency has been sampled and reconstructed by low-pass filtering. The result is a sinusoid whose frequency is half the Nyquist limit. This is an example of aliasing distortion: an input frequency above the Nyquist limit creates a spurious low frequency in the postsampling reconstruction. Here it is obvious from the figure why this occurs: the supra-Nyquist sinusoid and its sub-Nyquist alias create exactly the same sample values, so low-pass filtering of those values must produce the same result in both cases. The same would be true for any operation we performed on the samples.

Shannon's theorem generalizes readily to two dimensions—in other words, to image sampling (e.g., Goodman, 1968). There it assumes that the sampling array is a perfect lattice and tells us that any image can be reconstructed from its sample values provided it contains no spatial frequencies higher than a cutoff—a Nyquist limit—that depend on the spacing of the sample points. Roughly speaking, that limit again is half the sampling rate. And here again the input signal is reconstructed by low-pass filtering. In this case that filtering is accomplished by convolving the sample values with a two-dimensional analog of the sinc function—a Mexican sombrero with multiple ripples in its brim. And just as in one dimension, if we sample a spatial frequency higher than the Nyquist limit, we get sample values identical to those of some lower-frequency alias, when we reconstruct by low-pass filtering it is the alias that survives.

The only new wrinkle in two dimensions is that the alias generally has a different orientation than the input as well as a different frequency. Figure 2 illustrates this by showing a grating seen through a hexagonal lattice of holes whose Nyquist limit is less than the grating frequency. A point to note for future reference is the perfect regularity of the alias pattern.

Our concern is with violations of Shannon's assumptions by the cone mosaic. To understand the effects of those violations, one needs a general idea of the proof of the theorem, especially of the role played in that proof by the Fourier spectrum of the sampling array. Figure 3 illustrates the key points. Panel A represents a two-dimensional array of sample points—a forest of delta functions. Image sampling is a matter of multiplying the image times such an array, and in the spectral domain that means we are convolving the spectrum of the image (Figure 3B) with that of the array. When the sampling array is a perfect lattice, its spectrum is also a lattice, and the spacing of the spectral lattice points is inversely proportional to the spacing of the sample points (i.e., it is proportional to the sampling rate). Convolution creates multiple replicas of the image spectrum, one centered at each point of the array spectrum (Figure 3C). When the sampling rate

is sufficiently high these replicas do not overlap, and the one centered on zero frequency is a perfect copy of the image spectrum. So we can isolate it by low-pass filtering and recover the image by Fourier inversion. But if the sampling rate is too low, the replicas do overlap, and no clean copy can be isolated by spatial filtering.

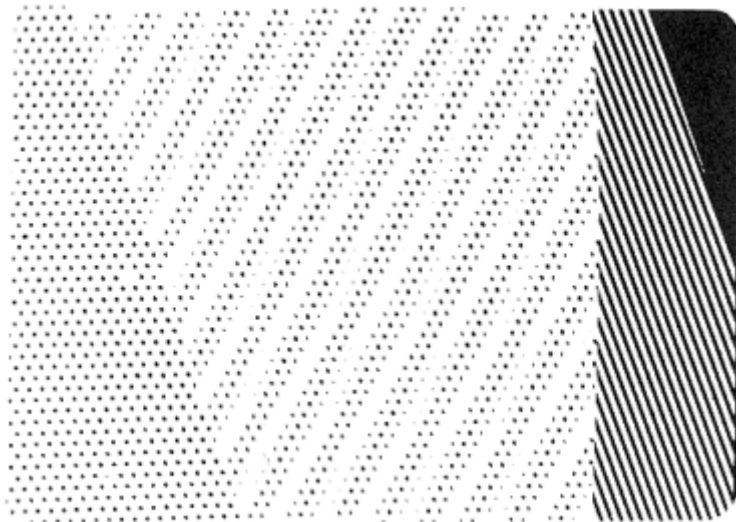


Figure 2

Aliasing in two dimensions. NOTE: A hexagonal array (seen alone on the left) samples a grating (seen alone on the right) whose frequency exceeds its Nyquist limit. After sampling (middle) the grating appears to have a lower frequency and a different orientation. (The figure is printed in reverse contrast.)

So the critical assumption of Shannon's theorem is really the spatial regularity of the sampling array—or, equivalently, the lattice-like nature of its spectrum. It is this regularity that creates the discrete spectral replicas needed for perfect image reconstruction by low-pass filtering.

SAMPLING PARAMETERS OF THE FOVEAL CONE MOSAIC

Now we turn to the parameters of retinal image sampling by the cones. I will deal first with the fovea and then with the rest of the retina. Our concerns are (1) the spatial frequency bandwidth of the image, (2) the Nyquist limit implied by cone density, and (3) the spatial regularity of the mosaic.

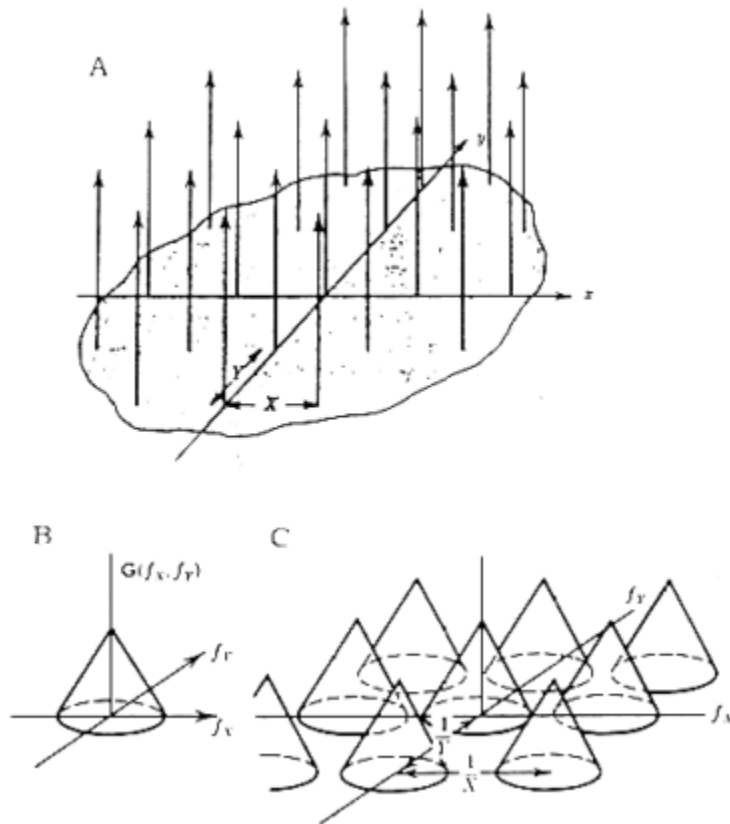


Figure 3

Key concepts in the proof of Shannon's theorem for two-dimensional sampling. NOTE: (A) Two-dimensional lattice of sample points represented as delta functions. Here the lattice is rectangular, with x, y spacings X, Y . The Fourier spectrum of this lattice is another lattice with reciprocal spacings $1/X, 1/Y$. (B) $G(f_x, f_y)$ is the spectrum of some two-dimensional signal $g(x, y)$. (C) The spectrum of $g(x, y)$ after sampling consists of multiple replicas of G , one centered at each point in the spectrum of the sampling array.

SOURCE: Adapted from Goodman (1968).

For the fovea, Campbell and Gubisch's (1966) line-spread measurements indicate that the spatial bandwidth of normal retinal images is on the order of 60 cycles/degree.

Osterberg's (1935) classic measurements of cone density in a single human eye implied a Nyquist limit for the foveal center of almost exactly 60 cycles/degree. Recently, Curcio and her colleagues examined several eyes and found foveal Nyquist limits ranging from 50 cycles all the way up to 85 (Curcio et al., 1987)! But their average value is 65 cycles, so we can still say that the Nyquist limit of the foveal cones is on the order of 60 cycles/degree.

Now if the foveal cones formed a perfect spatial lattice, Shannon's theorem would allow us to say that in principle no information is lost there by receptor sampling. Figure 4 shows that foveal cones in the primate retina can achieve a high degree of spatial regularity over distances on the order of one-tenth of a degree. The top panel shows the center-point positions of roughly 100 cones in the center of the fovea of a macaque monkey, carefully measured by Hirsch and Hylton (1984). The cones were sectioned at the inner-segment level, near their optical entrance aperture (Miller and Bernard, 1983), so the picture accurately represents the effective sampling regularity of the mosaic.

The bottom panel of Figure 4 shows the Fourier spectrum of this array of cones (Ahumada and Yellott, 1985). We see that the spectrum is quite lattice-like out to very high frequencies. So over small regions containing a 100 or so cones, the lattice assumption of Shannon's theorem seems to be well satisfied by the foveal receptor mosaic.

But this high degree of spatial regularity is not maintained over larger distances. Figure 5 (panels C and D) shows the aliasing patterns created when a 1-deg section of monkey foveal cones is made to sample gratings at frequencies above its nominal Nyquist limit (Williams, 1985). The section itself (shown in Figure 5A) is the same one from which Hirsch and Hylton selected the 0.1-deg patch of cones whose positions are shown in Figure 4. We see that there is very conspicuous aliasing here, but it does not have the long-range periodic character that would be produced by a perfect lattice (cf. Figure 2). And when the normal 60-cycle bandwidth of the eye is bypassed by interferometry, as Williams (1985) has done, the aliasing patterns one perceives have the same ragged quality as those we see here. So both anatomy and psychophysics indicate that spatial regularity in the human foveal cone mosaic is preserved only over distances on the order of one-tenth of a degree.

In the fovea then the spatial bandwidth of the retinal image matches the Nyquist limit implied by overall cone density, and the cone mosaic is locally regular enough to create narrowband aliasing of super-Nyquist

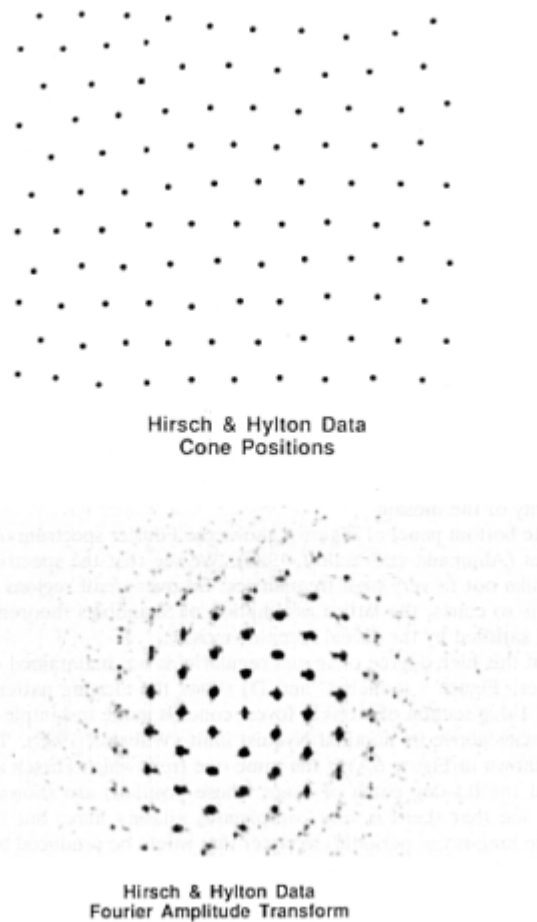


Figure 4

Spatial regularity of the central foveal receptor lattice. NOTE: Top: dots mark the positions (center-points) of cone inner segments in the central fovea of a monkey (*Macaca fascicularis*) as measured by Hirsch and Hylton (1984). The mean interpoint distance is $3 \mu\text{m}$. Section width corresponds to about 0.1 deg on the human retina. Bottom: Fourier amplitude spectrum of the point array on the left.

gratings. But the mosaic overall certainly does not have the kind of perfect lattice structure assumed by Shannon's theorem.

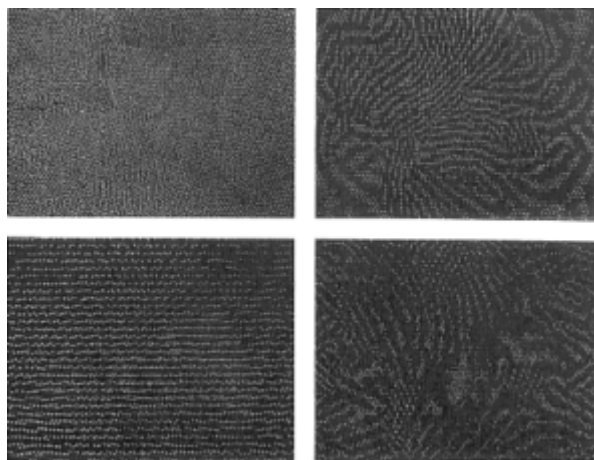


Figure 5

Gratings sampled by the cones in a monkey fovea. NOTE: (A) Cone inner-segment positions (at the external limiting membrane) in the fovea of *Macaca fascicularis*. Section width corresponds to 1 deg on the human retina. (B) Foveal cones sampling at 40 cycles/degree (i.e., sub-Nyquist) square wave. (C) 80 cycles/degree. (D) 110 cycles/degree. SOURCE: Adapted from Williams (1985).

IMAGE RECOVERY FROM IRREGULAR SAMPLES: YEN'S THEOREM

Does this lack of spatial regularity imply that foveal receptor sampling necessarily loses information? In principle, the answer is no—provided the actual positions of the sample points are known. Perhaps the best-known mathematical result in this connection is Yen's theorem (1956, Theorem I), which provides an explicit reconstruction algorithm for the case in which any finite number of sample points have been arbitrarily displaced away from their lattice positions. Yen's theorem assumes that the input signal contains no frequencies higher than the nominal Nyquist limit (i.e., the Nyquist limit implied by the average sampling rate).

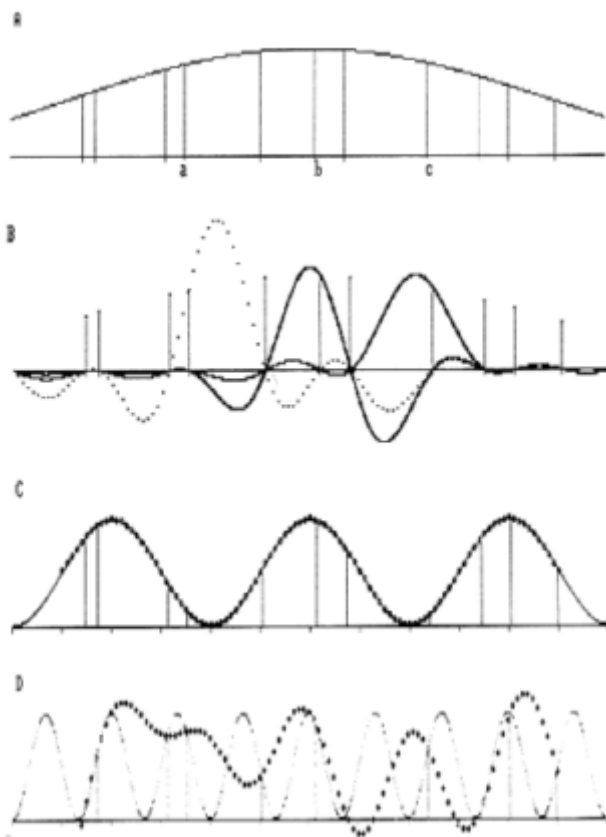


Figure 6
Yen's theorem in one dimension. NOTE: (A) Continuous signal sampled at irregularly spaced points. (B) Yen interpolation functions for sample points 3, 6, and 9. (The curve for point 3 is dotted for visibility.) (C) Yen reconstruction of a sinusoid at one-half the nominal Nyquist frequency. (D) Yen reconstruction of a sinusoid at 1.5 times the Nyquist frequency.

Yen's sampling theorem, like Shannon's, assumes an infinite number of sample points and strictly bandlimited signals, so we need assurance that his algorithm actually works under realistic conditions. [Figure 6](#) demonstrates that it does and also illustrates the mechanics of Yen reconstruction. Panel A shows 11 irregularly spaced sample points. Like Shannon reconstruction, Yen reconstruction is based on interpolation functions centered at the sample points. But instead of using the same sinc function for every point, Yen's algorithm requires each point to have its own idiosyncratic interpolation function—one that depends on the distances between that point and all the others. Panel B shows the interpolation functions for three of the points from panel A. One can see that Yen interpolation functions look like distorted sinc functions drawn by a careless draftsman: they are often distinctly asymmetric, and no two are exactly alike.

Like Shannon's algorithm, Yen reconstruction consists of weighting each sample point's interpolation function by the sample value (i.e., the input signal value) at that point and then summing all the interpolation functions. [Figure 6C](#) shows the results for a sinusoid whose frequency is one-half the nominal Nyquist limit. The results are as good as the comparable Shannon reconstruction shown in [Figure 1C](#) (i.e., essentially perfect).

[Figure 6D](#) shows what happens when Yen's algorithm is applied to a sinusoid whose frequency exceeds the nominal Nyquist limit. Here the results are quite unlike those produced by regularly spaced sample points (e.g., in [Figure 1D](#)): instead of aliasing creating a perfect low-frequency (sub-Nyquist) sinusoid, the reconstructed signal looks like a sinusoid corrupted by low-frequency noise. This is a general feature of Yen reconstruction: as the spacing of the sample points becomes more irregular, the less the aliases of supra-Nyquist sinusoids look like perfect low-frequency sinusoids and the more they look like broadband noise. This property of Yen reconstruction is at least qualitatively consistent with what one sees when supra-Nyquist gratings are imaged on the fovea by interferometry.

In the vision literature Yen's algorithm has sometimes been treated as one that could not be readily implemented by a visual system, and this has been used as an argument that irregular image sampling must have very deleterious consequences for vision (French et al., 1977). This idea may have been prompted in part by the fact that in print the analytic expressions for Yen interpolation functions look horribly complicated. But in reality the theorem is based on a simple idea, and the computations involved are all straightforward linear operations.

The idea is that if we had sampled an appropriately bandlimited signal at regularly spaced points, we could have used sinc interpolation to find its values at any other set of points. So given the signal values at the points where we actually did sample, we can interpolate backward to find the

values that must have occurred at the lattice points and then use them to reconstruct the entire signal by Shannon's method. This process turns out to be equivalent to erecting a tailor-made interpolation function around each of the actual sample points—a function that is a weighted sum of sinc functions. The only trick is to find the proper weights, and these can be readily calculated (by a matrix inversion) if the actual positions of the sample points are known.

It is important to note that these weights need to be computed only once for any given sampling array. Thereafter, Yen reconstruction is computationally just as simple as Shannon reconstruction. So Yen's algorithm seems quite compatible with a visual system whose sampling elements do not change position from moment to moment, like blades of grass in the wind, but that form a stable irregular mosaic that lasts a lifetime.

EXTRAFOVEAL IMAGE SAMPLING

Yen's theorem allows us to say that if the visual system can learn the actual positions of its foveal cones, then despite spatial disorder in the cone mosaic, receptor sampling of normal foveal images need not impose any loss of information. The critical point is that in the fovea the average sampling rate matches the spatial frequency bandwidth imposed by the optics of the eye.

Outside the fovea, however, there is a significant mismatch: the cones undersample the retinal image. Osterberg (1935) found that cone density decreased very rapidly with eccentricity in the single retina he examined, and the recent work of Curcio et al. (1987) indicates that Osterberg's density versus eccentricity curve is generally valid across individuals for eccentricities beyond about 1 deg. Thus, in a typical human retina the nominal Nyquist limit of the cones at 4-deg eccentricity has dropped from its foveal value of 60 cycles/degree to about 20. But the available evidence indicates that the spatial bandwidth of the retinal image has not decreased by anything like the same amount. In fact the line-spread measurements of Jennings and Charman (1981) suggest that at 4 deg the bandwidth is still around 60 cycles/degree, so in that region the cones undersample the retinal image by a factor of 3!

That estimate may be somewhat extreme, because increased retinal thickness outside the fovea complicates the problem of estimating the true spatial bandwidth of the image at the level of the receptors. But there seems to be little doubt that image bandwidth decreases more slowly than cone density, so that the retinal image is significantly undersampled outside the fovea. Recent psychophysical results support this conclusion: Still and Thibos (1987) found that at 20 deg in the periphery the gratings on a CRT screen can be discriminated from uniform fields up to 22 cycles/degree,

while the nominal Nyquist limit at that eccentricity is about 10 cycles. In other words, the retinal image bandwidth at 20 deg must be at least twice the local Nyquist limit of the cones.

Unlike cone density, which decreases continuously with retinal eccentricity, the spatial regularity of the cones declines only over the first two to three degrees and then appears to reach a stable state that prevails throughout the rest of the retina. This fact is easy to appreciate by examining the Fourier spectra of small sections of cones from different eccentricities (Yellott, 1983a) or from comparable statistical analyses (Hirsch and Miller, 1987). [Figure 7](#) shows, on the left, the cone positions in a small section of monkey retina at 2.5-deg eccentricity (from Hirsch and Miller, 1987) and, on the right, the Fourier spectrum of this cone array. Clearly this spectrum looks nothing like the perfect lattice of deltas required by Shannon's theorem. Instead it contains a single delta at the origin, surrounded by a circular island of empty space that ends abruptly in a sea of noise. This kind of "desert island" spectrum characterizes local sections of the cone mosaic at all eccentricities beyond about 2.5 deg. As cone density decreases beyond that point, the radius of the desert island decreases, but it always has a value that is approximately twice the nominal Nyquist frequency implied by the local density. [A desert island spectrum indicates that the extrafoveal cones are packed essentially at random but are subject to a constraint on the minimum cone-cone distance (Yellott, 1983b).]

[Figure 8](#) illustrates the spectral consequences of desert island sampling for a sinusoidal grating. The top represents the spectrum of the sampling array and the spectrum of the grating and simply reminds us that the spectrum of the grating after sampling will be the convolution of the two. On the bottom we see the postsampling spectrum for cases in which the grating frequency falls below or above the nominal Nyquist limit. Three points can be observed here: (1) For sub-Nyquist frequencies the postsampling spectrum contains deltas that fall in a relatively noise-free region of frequency space. In other words, sub-Nyquist images do not mask themselves. (2) For frequencies above the Nyquist limit there is no concentration of aliased energy at any single low frequency. Instead the aliased energy is scattered out into broadband noise at all orientations. Consequently, aliasing here will not take the form of periodic moiré patterns but instead will look like broadband noise. (3) The postsampling spectrum retains a strong concentration of spectral power at the original input frequency, even when that frequency falls above the Nyquist limit. This means that an irregular sampling array offers the potential for vision beyond the Nyquist frequency.

[Figure 9](#) illustrates the last two points using an actual array of extrafoveal cones. At the top the cones are shown sampling horizontal and vertical gratings at 1.25 times the Nyquist limit. Below are the power

spectra of the sampled images. We see that there is no hint of the kind of narrowband aliasing (i.e., moiré patterns) produced by the foveal cones (cf. Figure 5). Aliasing here takes the form of broadband noise. And we see in the postsampling spectra that there is easily enough energy concentrated at the input frequencies to allow us to determine which spectrum is which.

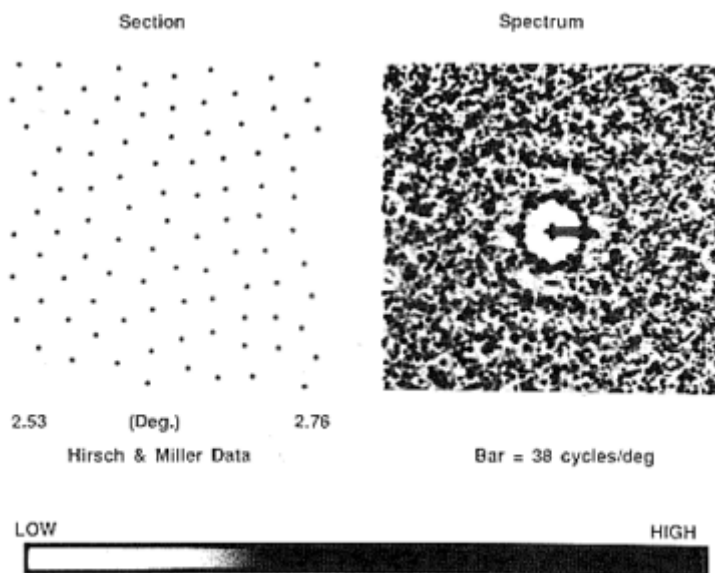


Figure 7

Spatial disorder in the extrafoveal cone mosaic. NOTE: Left side: dots mark the positions of extrafoveal cone inner segments in a monkey retina at 2.5 deg eccentricity (as measured by Hirsch and Miller, 1987). Right side: Fourier spectrum of the cone array (computed by A.J. Ahumada). The horizontal bar across the desert island portion of the spectrum marks a distance equal to twice the Nyquist frequency implied by local cone density.

Of course for the visual system to make such a discrimination, it must have orientation-sensitive mechanisms tuned to spatial frequencies that exceed the local Nyquist limits of its cones. Apparently our visual system does have such mechanisms, because Williams and Coletta (1987) have recently shown that observers can discriminate grating orientation at frequencies up to 1.5 times the nominal Nyquist limit out to 10-deg eccentricity. (At higher frequencies grating orientation can no longer be

identified, but spatial contrast can still be detected in the form of broadband noise.)

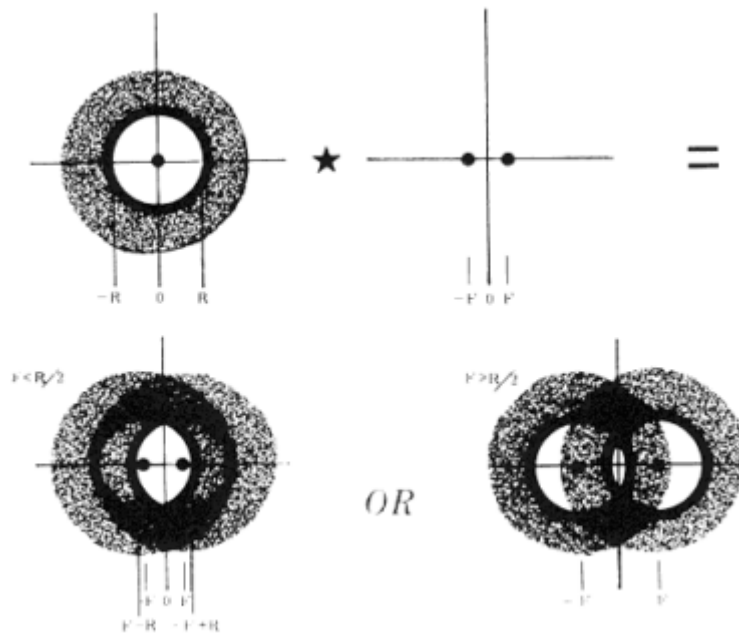


Figure 8
 Schematic illustration of the effects of sampling a grating with a point array whose spectrum has the desert island form found in the primate extrafoveal cone mosaic. NOTE: Top left: idealized desert island spectrum whose noise-free island has radius R ($R/2 =$ nominal Nyquist frequency). Top right: spectrum of a sinusoid with frequency F . Bottom left: spectrum of the sampled grating when $F < R/2$. Here the grating energy at $\pm F$ escapes masking by the sampling noise, all of which falls outside the Nyquist region. Bottom right: postsampling spectrum when $F > R/2$. Aliased energy in the form of broadband noise is widely scattered over the Nyquist region, with no concentration at any single sub-Nyquist frequency, and sharp spikes remain at the input frequency points $\pm F$.

How can the visual system make effective use of the information provided by a highly disordered receptor mosaic that undersamples the retinal image by a factor of 2 or more? Yen's theorem provides no guidance here, but recently Chen and Allebach (1987) showed that for bandlimited images undersampled by irregular point arrays, the least-squares reconstruction algorithm under quite general assumptions is very similar to Yen's. Like Yen reconstruction, Chen-Allebach reconstruction is based on tailor-made interpolation functions for each sample point (quite like those in Figure 6B),

and the construction of those functions requires knowledge of the actual sampling position. But here again, once the interpolation functions have been created, they are good for life, and the reconstruction of any given input image is computationally just as simple as it would be for spatially regular sampling below the Nyquist limit. It is too early to say how well the Chen-Allebach algorithm with mesh with the facts of extrafoveal vision, but I think it is a promising direction for exploration. A.J. Ahumada and I are currently studying mechanisms by which higher visual centers could learn

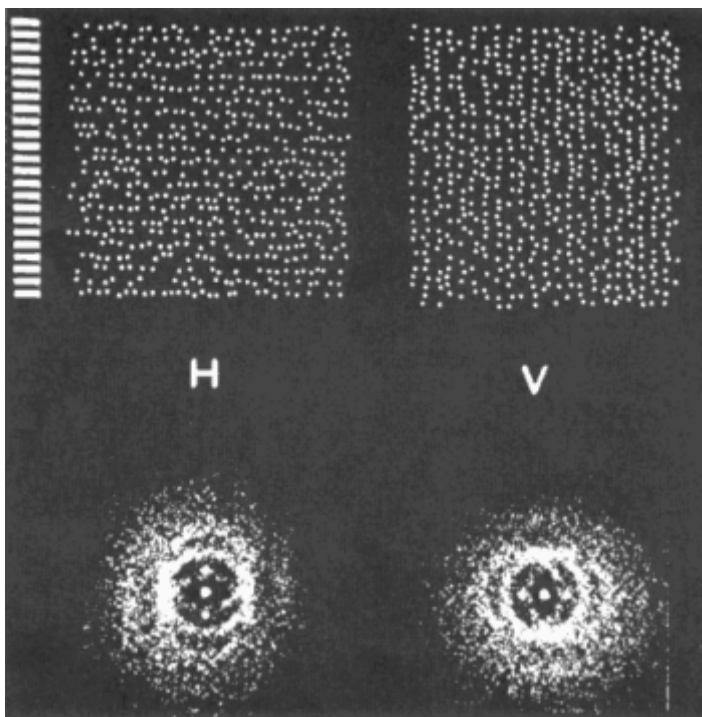


Figure 9
Visibility of spatial frequencies beyond the Nyquist limit in extrafoveal retina.
NOTE: Top: one degree section of extrafoveal rhesus monkey cones (3.8-deg eccentricity) sampling horizontal and vertical gratings whose frequency is 1.25 times the nominal Nyquist frequency of the cone array. Bottom: Fourier power spectra (optical transforms) of the sampled gratings. SOURCE: Adapted from Williams and Coletta (1987).

the receptor positions and compute the appropriate interpolation functions (Ahumada and Yellott, 1988).

ACKNOWLEDGEMENTS

I thank L.N. Thibos for sharing information on the work of Bergmann (1858), M. D'Zmura for providing his English translation of Bergmann's paper, D.R. Williams for permission to reproduce Figures 5 and 9, and A.J. Ahumada for his long-term collaboration. This research was supported in part by the National Aeronautics and Space Administration under joint research interchange NCA2-5.

References

- Ahumada, A.J., and J.I. Yellott 1985 A model for foveal photoreceptor placing. *Investigative Ophthalmology and Visual Science* 26 (Suppl.):11 (abstract).
- 1988 A connectionist model for learning receptor positions. *Investigative Ophthalmology and Visual Science* 29 (Suppl.):58 (abstract).
- Bergmann, C. 1858 Anatomisches und physiologisches über die netzhaut des auges. *Zeitschrift für rationelle Medicine* 2:83-108.
- Brindley, G.S. 1970 *Physiology of the Retina and Visual Pathway*. Baltimore, Md.: Williams & Wilkins.
- Campbell, F.W., and R.W. Gubisch 1966 Optical quality of the human eye. *Journal of Physiology* (London) 186:558-578.
- Chen, D.S., and J.P. Allebach 1987 Analysis of error in reconstruction of two-dimensional signals from irregularly spaced points. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 35:173-180.
- Curcio, C.A., K.R. Sloan, O. Packer, A.E. Hendrickson, and R.E. Kalina 1987 Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* 236:579-582.
- French, A.S., A.W. Snyder, and D.G. Stavenga 1977 Image degradation by an irregular retinal mosaic. *Biological Cybernetics* 27:229-233.
- Goodman, J.W. 1968 *Introduction to Fourier Optics*. New York: McGraw-Hill.
- Helmholtz, H. 1860 *Handbuch der Physiologischen Optik*, vol. II. (English edition republished by Dover Publications, New York, 1962).
- Hirsch, J., and R. Hylton 1984 Quality of the primate photoreceptor lattice and limits of spatial vision. *Vision Research* 24:347-356.

- Hirsch, J., and W.H. Miller 1987 Does cone positional disorder limit near-foveal acuity? *Journal of the Optical Society of America A* 2:1481-1942.
- Jennings, J.A.M., and W.N. Charman 1981 Off-axis image quality in the human eye. *Vision Research* 21:445-455.
- Miller, W.H., and G.D. Bernard 1983 Averaging over the foveal receptor aperture curtails aliasing. *Vision Research* 23:1365-1369.
- Osterberg, G. 1935 Topology of the layer of rods and cones in the human retina. *Acta Ophthalmologica* 6(Suppl.):1-103.
- Shannon, C.E. 1949 Communication in the presence of noise. *Proceedings of the IRE* 37:10-21.
- Still, D.L., and L.N. Thibos 1987 Detection of peripheral aliasing for gratings seen in the Newtonian view. *Journal of the Optical Society America A*. 4:P79-P80 (abstract).
- Williams, D.R. 1985 Aliasing in human foveal vision. *Vision Research* 24: 195-205.
- Williams, D.R. and N.J. Coletta 1987 Cone spacing and the visual resolution limit. *Journal of the Optical Society of America A* 4:1514-1523.
- Yellott, J.I. 1983a Spectral consequences of photoreceptor sampling in the rhesus retina. *Science* 221:382-385.
- 1983b Nonhomogeneous Poisson disks model the photoreceptor mosaic. *Investigative Ophthalmology and Visual Science* 24 (Suppl.): 147 (abstract).
- Yen, J.L. 1956 On the nonuniform sampling of bandwidth limited signals. *IRE Trans. Circuit Theory* CT-3:251-257.

The Invisible Cone Mosaic

David R. Williams

The last 5 years have seen a resurgence in interest among theoreticians, anatomists, and psychophysicists in the imaging properties of the cone mosaic. This recent activity might lead one to the conclusion that the cone mosaic represents a critical and important bottleneck for spatial vision. However, I argue that the cone mosaic is spatially "transparent" for most natural images we might encounter and does not pose a primary limitation on spatial vision. This may sound heretical at a meeting devoted to photoreceptors, but I do not mean to imply that the recent work on cone sampling has been for naught. On the contrary, this work is interesting because it has shown us how evolution has smoothly incorporated the cone mosaic into the visual system, leaving almost no perceptual trace. By considering the mosaic in the context of the visual stages that precede and follow it, the reasons for its invisibility have become apparent.

The geometry of the cone array gives rise to at least two optical properties that can, under the right laboratory conditions, affect spatial vision. These two properties are illustrated in [Figure 1](#). The first has to do with the spacing between receptors, r , which can lead to the ambiguity known as aliasing. The second has to do with the diameter of the cone aperture, d , which causes them to act as low-pass spatial filters. These two factors are potential sources of information loss that are quite different in nature, and I will discuss them in turn, explaining why neither has much impact on ordinary spatial vision.

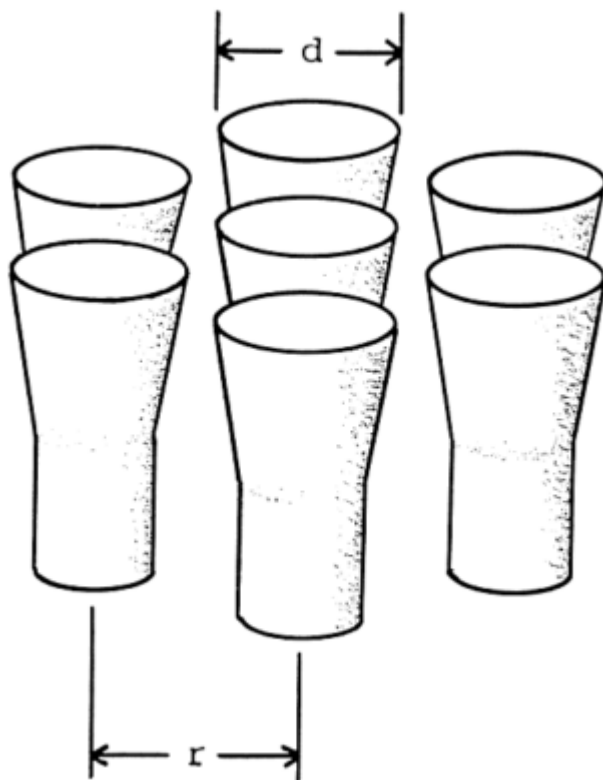


Figure 1
A psychophysicist's view of the foveal cone mosaic, characterized by the spacing between rows of cones, r , and cone diameter, d .

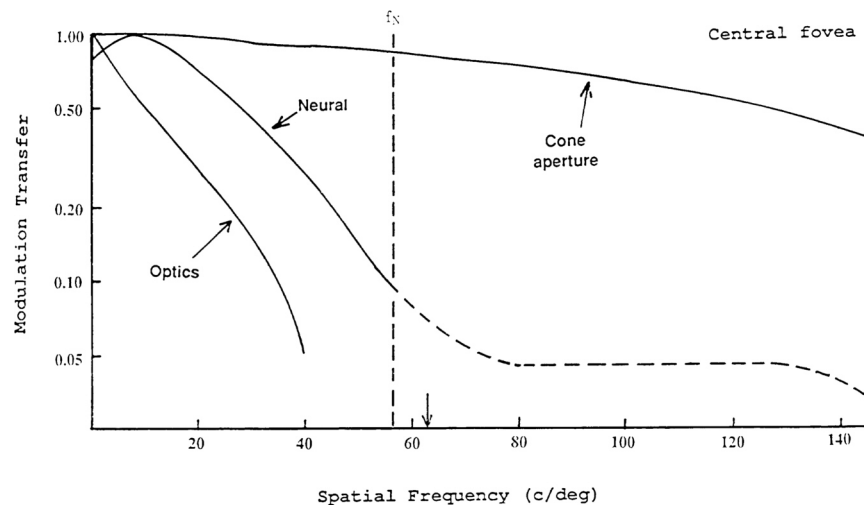


Figure 2
Modulation transfer functions for the foveal cone aperture and the optics of the eye (2-mm pupil). The foveal cone Nyquist frequency, f_N is shown as a vertical dotted line. The curve labeled "Neural" is a contrast sensitivity function arbitrarily normalized to unity at its peak. It represents foveal contrast sensitivity to interference fringes (Williams, 1985). The dotted portion of this curve indicates spatial frequencies that produce aliasing.

CONE SPACING

It has been known for some time that under normal viewing conditions blurring by the optics of the eye protects the fovea from aliasing. The vertical dotted line in Figure 2 shows the Nyquist frequency at the very center of the fovea, 56 cycles/degree. This estimate comes from psychophysical observations of aliasing (Williams, 1985a, 1988, 1989) and is in reasonable agreement with anatomical estimates (Osterberg, 1935; Miller, 1979; Curcio et al., 1987). Spatial frequencies below the Nyquist frequency are represented veridically by the visual system, while those above it are subject to aliasing. The modulation transfer function (MTF) of the eye's optics chosen here is that of Campbell and Gubisch (1966) obtained with a 2-mm pupil. The diffraction limit dictated by this pupil size is 63 cycles/degree at 555 nm. Slightly larger pupils could produce slightly more contrast at higher spatial frequencies, but in general there is precious little modulation in the retinal image in the aliasing range.

The situation is considerably different in the parafovea and periphery, where the MTF of the eye probably exceeds the cone Nyquist frequency. Unfortunately, we have surprisingly little information about the off-axis image quality of the human eye. I am not aware of any measurements of the MTF obtained with a normal daylight pupil size. Jennings and Charman

(1981) are frequently cited for their claim that optical quality is relatively constant across the central 25 deg of the retina. However, their data were obtained with a fully dilated 7.5-mm pupil, so their line-spread functions, even at the fovea, are roughly four times broader than those of Campbell and Gubisch (1966), obtained with an optimum pupil size. Though it remains to be seen how good the off-axis optical quality of the eye is under optimal conditions, there is no doubt that the retinal image there can be modulated at spatial frequencies above the cone Nyquist frequency. Bergmann (1858) described effects observed near the grating resolution limit that may be attributable to aliasing under normal viewing conditions. Smith and Cass (1987) and Thibos and Still (1988) have confirmed this, also without the use of high-contrast interference fringes.

Some observers can see extrafoveal aliasing by viewing a unity contrast grating with the eccentric retina. A Ronchi ruling placed on a light table works reasonably well. It helps to orient the grating parallel to the retinal meridian in which the grating lies since ocular aberrations, such as lateral chromatic aberration, will reduce contrast less. Gratings with a spatial frequency above the cone Nyquist frequency (e.g., above roughly 15 cycles/degree at 10-deg eccentricity in the nasal retina) can take on the appearance of dynamic two-dimensional noise. Under these conditions there is little or no perceptual evidence for the original grating: the stripes cannot be resolved and their orientation cannot be discerned. However, the speckled appearance of the grating distinguishes it from a uniform field of the same luminance (such as a neutral density filter matched to the average transmittance of the Ronchi ruling).

Nonetheless, there are a number of reasons why cone aliasing is not a severe problem in the extrafoveal under normal viewing conditions (Williams, 1986). Off-axis aberrations and perhaps increased retinal scatter help to some extent. The power spectra of natural scenes typically decline with increasing frequency. The high spatial frequency, high-contrast, sinusoidal gratings used in the laboratory are rare events in natural scenes. Natural scenes are typically complex, so that weak aliasing effects may be masked by the predominant spatial frequencies in the image that lie below the Nyquist frequency. Furthermore, small refractive errors can produce dramatic losses in contrast at high frequencies, and it seems likely that the peripheral retina is typically less well refracted than the fovea. Even with laser interference fringes, the most visible aliasing effects we have been able to produce are never more than about five times contrast threshold.

Yellott (1982) has suggested that disorder in the cone mosaic plays a major role in eliminating aliasing in the periphery, since it smears aliasing energy into a broad range of spatial frequencies and orientations. However, contrast sensitivity for detecting aliasing in the extrafoveal retina where the mosaic is disordered is not very different from that for detecting aliasing

with the more regular foveal mosaic. To be sure, disorder in the mosaic prevents regular patterns from being aliased into regular moiré patterns. But such regular patterns are rare in nature, and it seems more likely that the difference in regularity of the cone mosaic in fovea and periphery is a consequence of some disruptive effect of rods outside the foveal center.

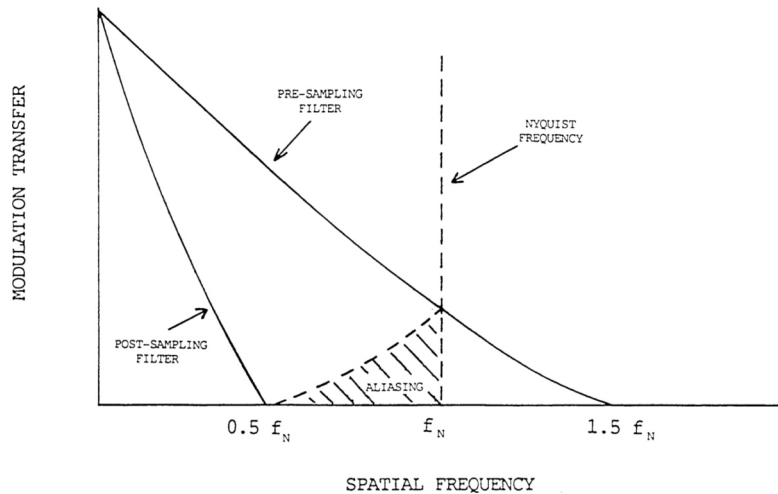


Figure 3
Hypothetical scheme, qualitatively similar to that found in peripheral retina, that allows low-pass filtering following cone sampling to eliminate aliasing under normal viewing conditions.

Hirsch and Hylton (1984) have pointed out that neural filtering following the sampling process can reduce aliasing distortion, though at the expense of overall spatial bandwidth. Consider the hypothetical example illustrated in Figure 3. Suppose the optical attenuation preceding sampling, labeled "pre-sampling filter" in the figure, has a cutoff at 1.5 times the Nyquist frequency, f_N . Aliasing corresponds to a reflection of signals above f_N to corresponding spatial frequencies below, as shown by the hatched area. However, if a postsampling filter, which might correspond, for example, to the centers of ganglion cell receptive fields, had a cutoff frequency as high as only $0.5 f_N$, the system as a whole would be protected from aliasing. Disorder in the cone mosaic would reduce the effectiveness of this scheme somewhat, but it could still greatly reduce the visibility of aliasing. A situation similar to this exists in the peripheral retina. Even with interference fringes, contrast sensitivity is clearly limited by postreceptoral factors in the peripheral retina, and acuity falls far short of the cone Nyquist frequency. So the scheme suggested by Hirsch and Hylton may

operate there. There are a large number of factors that mitigate against aliasing distortion outside the fovea, so that the granularity of the cone mosaic does not intrude in everyday vision.

CONE DIAMETER

The mosaic as a whole can never resolve a pattern that cannot be resolved by its individual elements. The cones in the human eye are typically smaller than the point-spread function of the eye's optics, so that the contrast attenuation produced by individual cones is negligible. [Figure 2](#) shows a psychophysical estimate of the aperture of foveal cones from MacLeod et al. (1985). Their estimate was obtained by measuring the contrast demodulation in the visual system prior to an intensive nonlinearity. This estimate should be considered as a lower bound on the spatial bandwidth of receptors because of uncertainty about the effect of additional sources of demodulation besides the receptor aperture, such as retinal scatter. The true curves might actually be shallower than this but are not likely to be steeper.

This estimate is slightly shallower than that of Miller and Bernard (1983), which was based on anatomical measurements of inner-segment diameter in the primate. MacLeod, Williams, and Makous (1985) calculate an upper bound on the effective foveal cone diameter of about 1.8 microns, compared with their estimate of 2.4. But what is the relationship between anatomical inner-segment diameter and the functional light-gathering aperture of cones? How does the presence of neighboring cones affect the waveguide aperture? Our present knowledge of cone waveguide properties is insufficient to answer these questions.

This leaves us with another puzzle, since cones everywhere could benefit from being somewhat larger, which would allow them to catch more photons without producing appreciable contrast loss in the "visible" range of spatial frequencies. Why then are cones so small for a given spacing between them? In the fovea perhaps there are physical constraints that prevent cones from occupying more of the image plane. Is their size in the periphery limited by a requirement to make room for quantum-hungry rods? Or are there other factors that limit their size? In any case it is clear that neither the cone aperture nor the spacing between cones has an important impact on spatial vision (Williams, 1985b), the possible exception being a modest amount of aliasing in extrafoveal vision under ordinary viewing conditions.

SAMPLING BY THE THREE CONE MOSAICS

So far I have been treating the cone mosaic as a homogeneous population, ignoring the three submosaics of short (S), middle (M), and long wavelength-sensitive (L) cones that comprise it. The spatial sampling rate of each submosaic is of course lower than the sampling rate of the mosaic as a whole. One might suppose that this would make each submosaic susceptible to aliasing distortion. But as with the mosaic as a whole, the visual system has succeeded in hiding the three submosaics from subjective experience.

The S cone mosaic runs the greatest risk of undersampling because it accounts for less than 10 percent of the cone population. However, as Yellott et al. (1984) have argued, it is protected in part by axial chromatic aberration in the eye, which is most severe at the short wavelengths to which the B cones are sensitive. Under most conditions the short wavelengths to which the S cones are sensitive produce such a blurred image that the array of S cones is adequate to represent it. Williams and Collier (1983) (see also Williams et al., 1983) showed that aliasing with the S cone mosaic can be seen, but it required careful correction for chromatic aberration, and even then the phenomenon was difficult to detect.

Are the M and L cone mosaics susceptible to aliasing? One view would hold that it would be: if foveal resolution is limited by cone sampling, then visual acuity for a grating that modulates only a single cone type should be reduced as a result of an effective reduction in spatial sampling rate. For example, assume for simplicity that the loss of one cone type through chromatic adaptation reduces the number of effective sampling elements underlying the fringe by a factor of 2. (The actual value will be somewhat more or less depending on the ratio of M to L cones.) If the remaining cones were arranged in a perfectly regular lattice, the Nyquist frequency of the submosaic would be $1/\sqrt{2}$ or 71 percent of its value when M and L cones operate together. This simplistic application of sampling theory would then predict that visual resolution would be reduced correspondingly.

However, Green (1968) and Cavanaugh and Estevez (1975) measured contrast sensitivity at high spatial frequencies under conditions designed to isolate M or L cones. Neither study found a difference in performance at high spatial frequencies when only one submosaic mediated detection compared with both cone classes operating together. Nancy Coletta and I recently took another look at this issue, making measurements of visual acuity with laser interference fringes superimposed on chromatic backgrounds. We hoped the use of these high-contrast stimuli would improve our chances of seeing a difference. But so far we also have failed to find a difference between M cone acuity, L cone acuity, and M plus L cone acuity. Chromatic adaptation was used to attempt to isolate M or L cones

in the manner of Green (1968). Measurements were made for three kinds of interference fringe stimuli: (1) those of equal contrast for M and L cones produced by superimposing either a 488 nm or a 633 nm fringe on a background of the same wavelength, (2) fringes that favored M cone grating detection in which a unity contrast 488 nm grating was superimposed on a 660 nm uniform adapting field, and (3) fringes that favored L cone grating detection in which a unity contrast 633 nm grating was superimposed on a 460 nm background. The observer made two settings for each of these conditions of chromatic adaptation: (1) he adjusted the spatial frequency of the fringe to his resolution limit, defined as the highest spatial frequency at which he could perceive fine stripes at the appropriate orientation, and (2) he adjusted the spatial frequency of the interference fringe to produce the coarsest zebra stripes at the foveal center. That is, he identified a "moiré zero" (Williams, 1988) in which the period of the grating matches the spacing between receptors.

The mean settings for both these tasks are plotted in [Figure 4](#). The abscissa is an index of the cone isolation achieved by the three adaptation conditions, where 0 indicates perfect M cone isolation, 1 indicates perfect L cone isolation, and 0.5 indicates equal contrasts in both M and L cones. The contrasts in the M and L cones for each of these stimuli were calculated using Smith and Pokorny fundamentals (Boynton, 1979). These conditions were chosen, based on the results of contrast sensitivity measurements, so that both fringes near the resolution limit and zebra stripes would be above contrast threshold for the favored cone mechanism but well below threshold for the unfavored cone mechanism. The results show that neither resolution nor the moiré zero depends on the ratio of contrasts in the two cone types, even when the ratio approaches a factor of 10. Furthermore, the observer could detect no obvious subjective difference between the zebra stripes seen when both cone types operated together and when one or the other cone type was strongly favored.

Why should foveal resolution not decrease when only one submosaic is operating? The logic that leads to the prediction of a resolution loss under chromatic adaptation rests on the assumption that the remaining submosaic is perfectly regular. This is unlikely to be true. Williams and Coletta (1987) have shown that under some conditions observers can reliably identify the orientation of a grating even when its spatial frequency exceeds the Nyquist limit by 50 percent. That is, the invariance of visual resolution with chromatic adaptation is consistent with the supra-Nyquist visual resolution previously observed in the parafovea. Despite the silencing of sampling elements by chromatic adaptation, there are apparently sufficient samples remaining to provide the observer with enough information to recognize a grating. It seems that the visual system can sustain substantial loss of sampling elements and still retain high visual acuity. This observation is

relevant to the clinical assessment of retinal damage since visual acuity measures are not likely to be sensitive to random losses of large numbers of neural elements across the visual field.

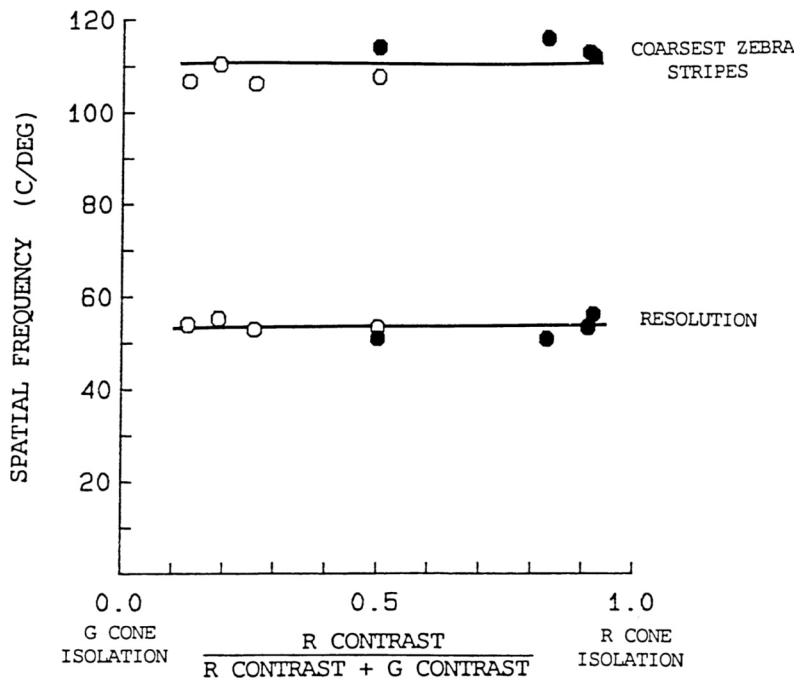


Figure 4
Measurements of interference fringe acuity and the coarsest aliasing patterns obtained under various conditions of chromatic adaptation.

Why does the moiré zero not change when only one submosaic is operating? The moiré zero should not change because the silencing of sampling elements from a mosaic, in either a random or regular fashion, does not eliminate the periodicity in the mosaic that is responsible for the moiré zero. This fact can be understood by considering Figure 5. The left half of the image shows a regular triangular lattice sampling a grating whose period nearly equals the sample spacing. That is, the grating frequency is near a moiré zero for the mosaic. The low-frequency diagonal grating is the resulting moiré or alias. On the right, two-thirds of the samples from the original array have been removed in random fashion. Clearly, the moiré is largely unaffected. Figure 6 shows that this is true even if the removal of cones is nonrandom. In this case, on the right, only every third row of

receptors has been retained, yet the moiré is the same. So there is in fact no theoretical expectation that the spatial frequency yielding the coarsest zebra stripes should change when one submosaic is effectively removed.

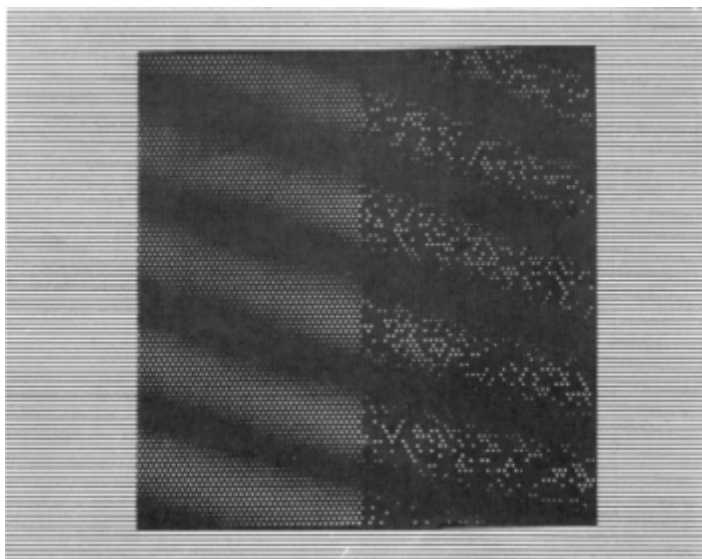


Figure 5

Square wave grating sampled by a regular triangular lattice on the left and the same lattice with two-thirds of the elements removed at random on the right (a submosaic). The grating frequency nearly equals twice the Nyquist frequency of the complete mosaic, producing an alias (low-frequency diagonal grating) that is conspicuous with the submosaic as well.

This does not imply that there are no differences in the aliasing effects predicted for a complete mosaic and the submosaics that comprise them. The submosaic in Figure 5 introduces aliasing noise at all input spatial frequencies, unlike the complete mosaic. More interesting effects arise when the submosaic has a regularity over and above that of the mosaic as a whole. For example, the submosaic in Figure 6 has a Nyquist frequency for horizontal gratings that is only a third that of the complete mosaic. This implies that there will be a second moiré zero for the submosaic at one-third the frequency of the moiré zero shared by both mosaics. That is, regular submosaics should show additional moiré effects at low frequencies over and above those seen with the complete sampling array.

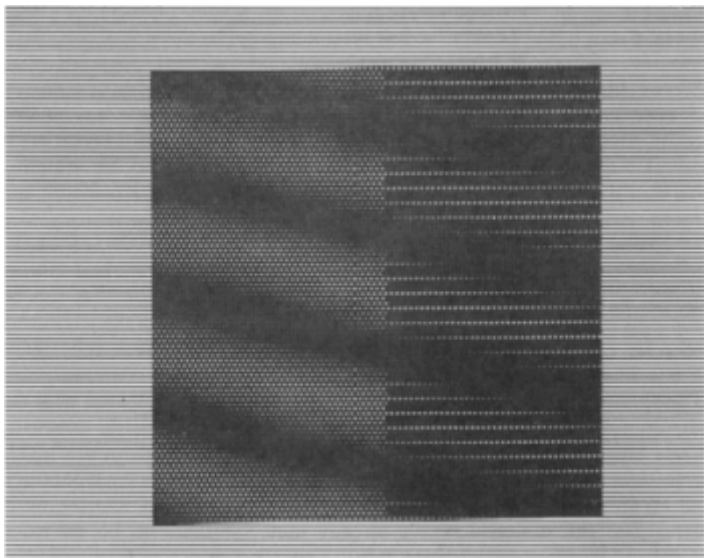


Figure 6

Square wave grating sampled by a regular triangular lattice on the left and the same lattice with two-thirds of the rows of cones removed on the right (a submosaic). The grating frequency nearly equals twice the Nyquist frequency of the complete mosaic, producing an alias (low-frequency diagonal grating) that is conspicuous with the submosaic as well.

Sekiguchi, Williams, and Packer (in press) have simulated the aliasing effects predicted by a variety of packing arrangements of the M and L cones. These simulated effects can be quite striking and yield "chromatic aliases." These aliases should vary in chromaticity as well as brightness even when viewing a single monochromatic interference fringe. We are presently engaged in a search for chromatic aliasing, since the spatial frequencies and orientations of the gratings that produce them would specify the packing arrangement of M and L cones. We have identified a red-green zebra stripe pattern when viewing interference fringes near the foveal Nyquist frequency (Sekiguchi et al., in press), but unfortunately this effect does not seem to be generated by chromatic aliasing. Indeed, it is interesting that chromatic aliasing is not more obvious, and this may indicate that the assignment rule for the M and L cones is not particularly regular. Completely random submosaics could produce two-dimensional

aliasing noise but no regular chromatic zebra stripes (Ahumada, 1986). Alternatively, it may be that the low spatiotemporal bandwidth of chromatic mechanisms makes it impossible to see chromatic aliasing. Our failure to observe chromatic aliasing with gratings so far is still something of a puzzle, however, since a quite different set of observations suggest that it should be possible to see it. The observation that the color appearance of foveal point sources fluctuates from flash to flash has been attributed to the sampling effects of the M and L cone mosaic (Krauskopf, 1964, 1978; Krauskopf and Srebro, 1965; Cicerone and Neger, 1989; Vimal et al., 1989).

In any case, the M and L cone submosaics, like the mosaic as a whole, are apparently well protected from aliasing distortion under normal viewing conditions. The extreme ratios of M to L cone stimulation that can be produced with appropriate choices of monochromatic light in the laboratory are a far cry from the effects of the smooth and continuous reflectance spectra of real scenes. These submosaics are presumably less susceptible to undersampling both because of their high densities—compared with the S cones, for example—and because the overlap in their spectral sensitivity curves ensures that the images in each of the two submosaics are highly correlated. This is not to say that effects cannot be found in the laboratory, and we are in the process of trying to develop more sensitive tests to reveal them. But the stimuli we have already used are likely to be far more selective than those encountered in ordinary viewing.

Since this meeting, evidence for aliasing by the M and L cone submosaics has been found. Williams, Sekiguchi, and Packer (1990) have argued that the red and green subjective colors seen when viewing fine achromatic gratings are caused by M and L cone aliasing.

But what about arrays of cells deeper in the visual system? Do they produce aliasing? A likely candidate would be the array of ganglion cells in the peripheral retina, whose density is six to nine times lower than that of cones. Coletta and Williams (1987a, 1987b, 1987c) have shown that the ability to detect interference fringes above the cone Nyquist frequency must be due in part to cone aliasing. But is it possible that some of this aliasing originates at sites deeper in the visual system? Or has the postreceptoral visual system protected itself from aliasing? One simple way of doing this is to make receptive field centers large enough that they demodulate spatial frequencies above the array's Nyquist frequency. Precisely how well the visual system succeeds at this is not yet known.

References

- Ahumada, A., Jr. 1986 Models for the arrangement of long and medium wavelength cones in the central fovea. *Journal of the Optical Society of America A* 3:P92.

- Bergmann, C. 1858 Anatomisches und Physiologisches über die Netzhaut des Auges. *Zeitschrift für rationelle Medizin* 2:83-108.
- Boynton, R.M. 1979 *Human Color Vision*. New York: Holt, Reinhart and Winston.
- Campbell, F.W., and R.W. Gubisch 1966 Optical quality of the human eye. *Journal of Physiology* 186:558-578.
- Cavonius, C.R., and O. Estevez 1975 Contrast sensitivity of individual colour mechanisms of human vision. *Journal of Physiology* 248:649-662.
- Cicerone, C.M., and J.L. Nerger 1989 The relative numbers of long-wavelength-sensitive to middle-wavelength-sensitive cones in the human fovea centralis. *Vision Research* 29:115-128.
- Coletta, N.J., and D.R. Williams 1987a Undersampling by cones reverses perceived direction of motion. *Investigative Ophthalmology and Vision Science* 28(Suppl.):232.
- 1987b Motion reversal in peripheral retina. *Journal of the Optical Society of America A* 4:P80.
- 1987c Psychophysical estimate of extrafoveal cone spacing. *Journal of the Optical Society of America A* 4:1503-1513.
- Curcio, C.A., K.R. Sloan, Jr., O. Packer, A.E. Hendrickson, and R.E. Kalina 1987 Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* 236:579-582.
- Green, D.G. 1968 The contrast sensitivity of the colour mechanisms of the human eye. *Journal of Physiology* 196:415-429.
- Hirsch, J., and R. Hylton 1984 Quality of the primate photoreceptor lattice and limits of spatial vision. *Vision Research* 24:347-355.
- Jennings, J.A.M., and W.N. Charman 1981 Off-axis image quality in the human eye. *Vision Research* 21:445-455.
- Krauskopf, J. 1964 Color appearance of small stimuli and the spatial distribution of color receptors. *Journal of the Optical Society of America* 54:1171.
- 1978 On identifying detectors. Pg. 283-285 in *Visual Psychophysics and Physiology*. New York: Academic Press.
- Krauskopf, J., and R. Srebro 1965 Spectral sensitivity of color mechanisms: derivation from fluctuations of color appearance near threshold. *Science* 150:1477-1479.
- MacLeod, D.I.A., D.R. Williams, and W. Makous 1985 Difference frequency gratings above the resolution limit. *Investigative Ophthalmology and Vision Science* 26(Suppl.):11.
- Miller, W.H. 1979 Ocular optical filtering. In *Handbook of Sensory Physiology*, vol. VII/6A. Berlin: Springer.
- Miller, W.H., and G.D. Bernard 1983 Averaging over the foveal receptor aperture curtails aliasing. *Vision Research* 23:1365-1369.
- Osterberg, G. 1935 Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica* 6(Suppl.):1-103.

- Sekiguchi, N., D.R. Williams, and O. Packer in press Distortion of Nyquist frequency gratings in foveal vision. *Vision Research*.
- Smith, R.A., and P.F. Cass 1987 Aliasing in the parafovea with incoherent light. *Journal of the Optical Society of America A* 4:1530-1534.
- Thibos, L.N., and D.L. Still 1988 What limits visual resolution in peripheral vision? *Investigative Ophthalmology and Vision Science* 29(Suppl.):138.
- Vimal, R.L.P., J. Pokorny, V.C. Smith, and S.K. Shevell 1989 Foveal cone thresholds. *Vision Research* 29:61-78.
- Williams, D.R. 1985 Aliasing in human foveal vision. *Vision Research* 25:195-205.
- 1986 Seeing through the photoreceptor mosaic. *Trends in Neuroscience* 9:193-198.
- 1988 Topography of the foveal cone mosaic in the living human eye. *Vision Research* 28:433-454.
- in press Photoreceptor sampling and aliasing in human vision. In *Tutorials in Optics*, Duncan Moore, ed. Number 1, 1987 OSA Annual Meeting. *Journal of the Optical Society of America*.
- Williams, D.R., and N.J. Coletta 1987 Cone spacing and the visual resolution limit. *Journal of the Optical Society of America A* 4:1514-1523.
- Williams, D.R., and R.J. Collier 1983 Consequences of spatial sampling by a human photoreceptor mosaic. *Science* 221:385-387.
- Williams, D.R., R.J. Collier, and B.J. Thompson 1983 Spatial resolution of the short-wavelength mechanism. Pp. 487-508 in *Colour Vision: Physiology and Psychophysics*, J.D. Mollon and L.T. Sharpe, eds. London: Academic Press.
- Williams, D.R., N. Sekiguchi, and O. Packer 1990 Spatial aliasing by chromatic mechanisms. *Investigative Ophthalmology and Vision Science* (Supp.) 31(4):494.
- Yellott, J.I., Jr. 1982 Spectral analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing. *Vision Research* 22:1205-1210.
- Yellott, J.I., Jr., B.A. Wandell, and T.N. Cornsweet 1984 The nervous system. Pp. 257-316 in *Handbook of Physiology*. Bethesda, Md.: American Physiological Society.