Early vision and focal attention

Bela Julesz

Laboratory of Vision Research, Rutgers University—Kilmer Campus, New Brunswick, New Jersey 08903, and Division of Biology, California Institute of Technology, Pasadena, California 91125

At the thirty-year anniversary of the introduction of the technique of computer-generated random-dot stereograms and random-dot cinematograms into psychology, the impact of the technique on brain research and on the study of artificial intelligence is reviewed. The main finding—that stereoscopic depth perception (stereopsis), motion perception, and preattentive texture discrimination are basically bottom-up processes, which occur without the help of the top-down processes of cognition and semantic memory—greatly simplifies the study of these processes of early vision and permits the linking of human perception with monkey neurophysiology. Particularly interesting are the unexpected findings that stereopsis (assumed to be local) is a global process, while texture discrimination (assumed to be a global process, governed by statistics) is local, based on some conspicuous local features (textons). It is shown that the top-down process of "shape (depth) from shading" does not affect stereopsis, and some of the models of machine vision are evaluated. The asymmetry effect of human texture discrimination is discussed, together with recent nonlinear spatial filter models and a novel extension of the texton theory that can cope with the asymmetry problem. This didactic review attempts to introduce the physicist to the field of psychobiology and its problems—including metascientific problems of brain research, problems of scientific creativity, the state of artificial intelligence research (including connectionist neural networks) aimed at modeling brain activity, and the fundamental role of focal attention in mental events.

CONTENTS

SOME INTRODUCTORY READING FOR PHYSICISTS WHO ARE NOVICES IN BRAIN RESEARCH

This review is intended for the physicist who is interested in vision research. It is written by an experimental psychologist who works at the interface between psychophysics and neurophysiology and who has a background in physics. It is hoped that the essential concepts in this review can be followed by the physicist without any prior knowledge of psychobiology, and that some of the problems raised, and their solutions, will appeal to physicists.

As will be explained in the Introduction and Conclusion, this review is highly partisan and is restricted to a subfield of vision research called *early vision*. The advantages of studying early vision are twofold. First, it is possible to avoid the daunting problems of semantics and of the higher symbolic processes by concentrating on stereoscopic depth perception (stereopsis), motion perception, and effortless texture segmentation, which can be elicited by computer-generated stochastic displays (dynamic noise) of well specified geometrical, topological, and statistical properties. Second, these psychological phenomena can now be neurophysiologically studied in the input stages of the monkey cortex, for which, most importantly, the early neural processing stages of the visual system are practically identical to that of *homo sapiens*. The many dozens of higher cortical stages devoted to visual processing in the monkey are not well known, and their relationships to perception and cognition in the monkey, as well as in humans, are still a mystery. It is these enigmatic phenomena of form recognition, Gestalt organization, and the many unsolved problems of mental states (which are at present metascientific questions, as will be explained in the Introduction) that the naive reader would like most to read about, but that, unfortunately, will be skipped in this review. Although there are interesting speculations—even by Nobel laureates—about the nature of dreams, free will, consciousness, and so on, early vision can be studied without knowing what these concepts are, and very few references will be provided that relate to these subjective ideas.

The main reason why early vision, with its bottom-up processes, is still an interesting subfield of vision research is the inclusion of *focal attention.* Focal attention is a top-down process that permits us both to inspect different areas of the visual field without eye movement and to scrutinize different cortical areas. While the parallel processes of preattentive vision detect feature and texture gradients and tell the attentive system *where* to inspect, the searchlight of attention can scrutinize these gradients in order to identify *what* the features are (Sagi and Julesz, 1985). Focal attention superficially resembles some phenomena of consciousness, but is a simpler mental phenomenon and luckily can be scientifically studied with the available psychological and neurophysiological methods! In this review, much space is devoted to focal attention and to the preattentive-attentive dichotomy in general. For an up-to-date review on focal attention, see the article by neurophysiologists Desimone and Ungerleider (1989) and the review by Heilman *et al.* (1987).

Early vision exhibits many interesting mental phenomena, from cooperative interactions and order-disorder transitions to learning and memory. It is the author's belief that it serves as an excellent introduction to the vast field of psychobiology. Because of the link between psychology and neurophysiology, the reader has to be somewhat familiar with some rudimentary concepts of brain research. Much effort has been expanded to make this article self-contained and to define new concepts as they are first mentioned; for more details the relevant references are cited.

I should like to offer here some suggestions for introductory reading for the novice in psychobiology. The less ambitious reader might skip to the Introduction. Of course, there is no royal road to psychobiology, as there is no such road to physics. One approach might be to read this review with an encyclopedia of brain research nearby, such as *The Oxford Companion to The Mind* (Gregory, 1987). Another approach might be to read this review and then to browse through the recent monograph by Resnikoff (1987a) entitled *The Illusion of Reality,* which introduces a large variety of concepts and facts on biological information processing, such as spatial filters, the analysis of subjective contours, the duality between aliasing and noise, visual illusions, and even an early version of my texton theory. After reading Resnikoff, the reader might find my review clearer, the more so since we cover similar topics and I start where Resnikoff ends. The reader unfamiliar with the rudimentary concepts of brains and neurons should turn to one of the pioneers in the field and read the brief and clear work by David Hubel (1988), *Eye, Brain, Vision.* Patricia Churchland (1986) gives a good introduction to elementary neuroscience, although the author is not a practicing neurophysiologist, but a philosopher. Those interested in the principles of neuroscience could benefit from the excellent textbook *Principles of Neural Science,* edited by Eric Kandel and James Schwartz (1985). Richard Gregory (1990) in *Eye and Brain* and Irvin Rock (1984) in *Perception* give a colorful introduction to perceptual phenomena of vision, although from a cognitive point of view. The introductory textbook *Perception,* particularly on visual perception, by Sekuler and Blake (1985), is more neurophysiologically inclined and is highly recommended. A remarkable textbook, *The Senses,* on sensory and perceptual processes, edited and written by Horace Barlow and John Mollon (1982) gives a balanced introduction to early vision and to the other modalities. (The first author is a pioneer in visual neurophysiology, the second an expert in color perception.) Those interested in parallel computers and their possible analogies to the brain will benefit from the monograph by Daniel Hillis (1985), *The Connection Machine;* the author is the creator of a massive parallel computer with the same name.

There are many other interesting introductions to vision. The reader could benefit from the 25th anniversary issue of *Vision Research* (a leading journal of brain researchers working in vision), also published as a book edited by Boynton (1986), in which a dozen pioneers give a historical review of their specialties, including one by this author. A somewhat earlier, but still relevant, collection of articles on perception and visual adaptation phenomena, that benefitted considerably from the careful editing of Charles Harris (1980), *Visual Coding and Adaptability,* contains many introductory articles in advanced perception, including one by this author. A similar collection, but with emphasis on cortical neurophysiology, is the informative volume *Dynamic Aspects of Neocortical Function,* edited by Edelman *et al.* (1984). For more advanced texts, many references are given in this article. I highly recommend the monumental *Handbook of Physiological Optics* by the physicist-physiologist Hermann von Helmholtz (1896; the 1924 English edition by the Optical Society of America was reprinted by Dover in 1962). It is somewhat outdated, yet surprisingly, after a century, still fresh and stimulating. It is interesting to compare it with the *Handbook of Perception and Human Performance,* edited by Boff *et al.* (1986) a century later. The progress in knowledge is remarkable, but the many specialists in the later work cannot convey the unified and stimulating outlook of Helmholtz. The posthumous monograph of David Marr (1982), *Vision,* gives a unified and stimulating account of problems in computational vision from a leading figure of artificial intelligence. For a followup on Marr's legacy, the collection *Natural Computation,* edited by Whitman Richards (1988), is highly recommended. While Marr's view of computational vision is basically bottom-up, a recent short monograph by Roger Watt (1988), *Visual Processing: Computational, Psychophysical, and Cognitive*

Research, argues for algorithms in early vision that are under the control of high-level processes and memory. Finally, the reader who has difficulties in fusion random-dot stereograms (RDS) and would like to view more than are shown in this review can turn to Julesz (1971) *Foundations of Cyclopean Perception,* which contains many RDS printed as red-green anaglyphs; these can be easily fused with the provided red-green goggles.

There are several perceptual problems that seem elementary, do not require semantic memory, can be solved effortlessly, and yet are beyond our understanding. For example, knowing whether a point is inside or outside a bounded region—with rather meandering boundary lines—is such a problem, even though humans can solve it instantaneously. In this review, these problems of "visual cognition" will not be covered, and the interested reader should consult the insightful work of Shimon Ullman (1984). I can also recommend the two textbooks by Bruce and co-workers, *Visual Perception* by Bruce and Green (1985) and *Visual Cognition* by Humphreys and Bruce (1989). I wish to draw attention to the interesting work of Steven Grossberg and his collaborators, the more so because it has been widely ignored by the artificial intelligence community, although it is based on sophisticated mathematical tools. A good sample is Grossberg (1988), *Neural Networks and Natural Intelligence.* The physicist D. J. Amit's monograph *Modeling Brain Function* (1989) discusses connections between neural networks, spin glasses, Ising models, solid-state physics, etc., and might be of interest to physicists. There are many other excellent introductory texts to psychobiology, and the reader can always turn to friends for further advice.

Here I only define the important concepts of a neurophysiological "receptive field" and of a "cortical column." In current neurophysiology, a single microelectrode is placed near a neuron in order to record electrical activity from the neuron. The cat or monkey is anesthetized and immobilized, and some visual stimulus (e.g., an oriented bar) is moved across the visual field until the selected neuron is activated. [Recently monkeys with a sealed opening in their skulls (installed weeks prior to experimentation) have been trained to fixate on a marker. Since the brain is devoid of pain detectors, the neurophysiologists can penetrate the brain with their microelectrodes hundreds of times (until the dura hardens) and measure neural activity in the *alert* monkey.] The specific local area of the luminance gradient with proper orientation and aspect ratio that elicits the largest activity is the receptive field of the selected neuron. Some neurons require quite complex trigger features to become activated, and the spatial concept of the receptive field is inadequate to describe the behavior of these complex and hypercomplex feature analyzers. Nevertheless, as a first approximation, a receptive field is an important concept. After all, who might have believed before the epoch-making findings of Barlow, Hartline, Hubel and Wiesel, Kuffler, or Lettvin and Maturana, to name a few pioneers of modern neurophysiology, that one could look for some specific feature extractors at a well defined locale in the brain and find neurons that would respond selectively to local stimulus features, such as the color, orientation, velocity, direction, flicker rate, etc., in the image. The psychologically defined local features (and the textons, to be defined later) are in most cases similar to the neurophysiological trigger features, and this correspondence between fundamental units of mind and brain makes the field of early vision so exciting.

The cortical column, first discovered by Vernon Mountcastle in 1957 (for details, see Mountcastle, 1979) in the somatic sensory cortex of the cat, consists of narrow columns (slabs) that are perpendicular to the surface of the cortex and made up of neurons that perform similar local computations. For instance, neurons in the visual cortex of the monkey that are tuned to similar (but not identical) orientations of the same retinal location form such a column. Another example are those aggregates of neurons that are activated by left or right eye stimulations, respectively. These slabs are the columns of binocular dominance and can be made visible by radioactive dioxyglucose stains (Sokoloff, 1984); they can also be directly seen *in vivo* (after electronic enhancement) under the microscope (Ts'o *et al.*, 1990). For the concept of the "hypercolumn," see Hubel and Wiesel (1974).

I. INTRODUCTION

A. On strategic and metascientific problems in psychobiology and the doctrine of "specific nerve energies"

During the spring and summer of 1990, while I was preparing this manuscript for physicists, by coincidence I worked on and finished three other manuscripts, one written for experimental psychologists, that is, colleagues in my own field, the second for philosophers interested in visual perception, and the third for neurophysiologists. In the first, just published in *Trends and Tacit Assumptions in Vision Research* (Julesz, 1991a), I pose about 40 strategic questions in vision research, in addition to half a dozen metascientific questions. Although this paper is written for active workers in psychobiology, it might give the physicist reader additional glimpses of my field. The second article is an open peer review in which I answer, among others, the philosopher John Searle on his recent idea that the brain cannot have unconscious processes. This paper (Julesz, 1991b), will be published in the journal *Behavioral and Brain Research,* and I shall quote from it in a few places here. The third article (Julesz, 1990c) intended for neurophysiologists, will appear soon in the volume *Symposium #55: The Brain,* celebrating the 100th anniversary of the Cold Spring Harbor Laboratory. These three articles, together with this one, span a large audience belonging to four different disciplines, and while working on them I was forced to ponder the state of psychobiology from four different perspectives. The exercise confirmed yet again the broad appeal of problems in perception and their accessibility to workers in many disciplines.

The human brain, the most complex structure in the known universe, with its visual cortex is many orders of magnitude more complex than the objects customarily studied by physicists and chemists. Therefore the psychobiology of vision appears to be in a premature state comparable to that of physics prior to Galileo, or molecular biology prior to the discovery of the double helical structure of DNA by Watson and Crick. Indeed, workers in visual psychophysics, neurophysiology, and neuroanatomy cannot agree on strategic problems; what is more, some would even deny that we have adequate knowledge at present to formulate a strategic question. Just take the problem of "sensations" (also called "qualia"). Wherever the neurophysiologists probe the brain with microelectrodes, they seem to record similar histograms of neural spike activity, regardless of whether the corresponding sensations are brightness, color, pitch, itch, temperature, pain, pleasure, anxiety, hunger, or contentment. Johannes Müller (1844), in his doctrine of "specific nerve energies," was already aware of the problem of sensations and stated that *specific sensations arise in specific brain areas.* While this doctrine might seem somewhat unhelpful, I challenge any philosopher or brain researcher to say more with certainty at present than Johannes Müller said 150 years ago. [The interested reader is referred to a debate between John Searle and the author (Julesz, 1991b, in press).]

Sensations are only one of the many metascientific problems in brain research. The problems of "what is consciousness?," "is there free will?," "how are memories encoded?," "why do higher organisms sleep?," "what is the purpose of dreams?"—to name a few intriguing ones—cannot be tackled by the scientific methods currently available to us. Obviously these metascientific questions are not all equally complex. Perhaps the riddle of short- and long-term memory will be solved in the not-too-distant future. However, it could be that "consciousness" belongs to a Gödel-like problem, which might exclude neural nets to inspect (solve) certain complex states of their own. ["Is the human brain equivalent to, or more powerful than, a Universal Turing Machine (UTM)?" seems to be another metascientific problem, and the fact that the relatively simple "halting problem" cannot be solved (computed) by a UTM should instill in us a certain humility about tackling the really tough metascientific problems of our mental states. Returning to the question of the human brain's being more powerful than a machine, Penrose (1989) thinks he can discuss this question scientifically by claiming that the conscious human mind can grasp (create and solve) mathematical problems that are not algorithmic, and even states that the essence of mathematics is not described by algorithms.]

B. Bottom-up versus top-down processes, parallel versus serial processes, early vision, and focal attention

In spite of this somewhat pessimistic introduction, I have witnessed the extraordinary progress in vision research during the last 30 years, and at the Editors' request I shall try to review some of the progress that was made in my own specialty, "early vision," a subfield of the experimental psychology of visual perception. I shall also glance at visual neurophysiology and neuroanatomy. Before I define the discipline of early vision, I have to explain some of my assumptions:

I regard as "scientific psychology" only those subfields of psychology which emulate thermodynamics in the sense that higher (level I) phenomena can be explained by lower (level [I-1]) phenomena. In analogy to the phenomena (concepts) of temperature, pressure, entropy, enthalpy, etc., that arise from interactions between atoms and molecules as they collide with each other and the wall of the container, I want to define psychological phenomena (percepts) of depth and motion perception, textural segmentation (discrimination), focal attention, etc., as excitatory and inhibitory interactions between pools of neurons tuned to specific features (i.e., binocular disparity, motion disparity, texture gradients, etc.). This "structuralist" (reductionist) quest can be pursued only in vision (and even there within limits) and not in speech perception, as I shall now explain. Obviously, human speech perception cannot be explored by the current neurophysiological methods used on animals. (Although noninvasive methods of studying brain activity already exists, nevertheless the spatial and temporal resolution of PET scans and similar techniques are still inadequate.) Furthermore, the clinical literature on speech defects in stroke victims—though intriguing—by its very nature is more anecdotal than scientific. Because only *homo sapiens* possesses speech, my quest for a scientific psychology can only be pursued in vision. The visual system of the monkey is practically identical to ours and is intensively studied by neurophysiologists and neuroanatomists. At present only the first stages of the central nervous system of the monkey are studied with microelectrodes, and we have to restrict our stimuli so as to stimulate only the first retinal and cortical input stages (also called "bottom-up" processing stages). Thus we have to reduce or eliminate the influence of the higher (and enigmatic) cortical stages of semantic memory and symbolic processing (also called "top-down" stages).

Figure 1 depicts such a bottom-up/top-down view of the central nervous system. Of course, both the bottom-up and the top-down processes might have subprocesses with feedback loops, but in general the two main information streams flow as depicted. The reader should note that focal attention is regarded as a separate mechanism that can inspect many processing stages at will. Indeed, the reader, while ruminating over my last sentence, could easily turn his attention to countless thoughts and

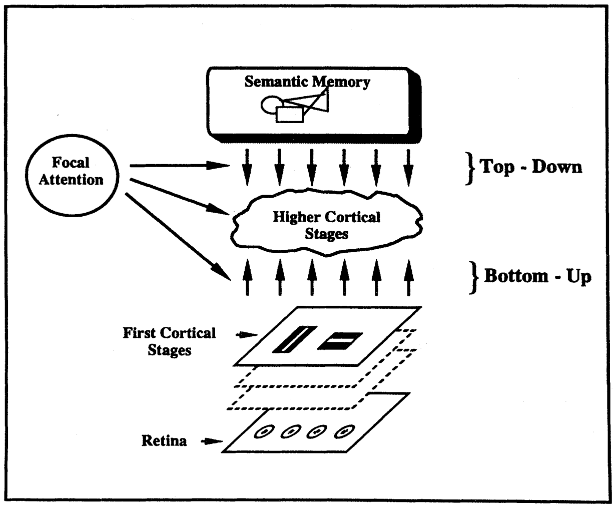


FIG. 1. Bottom-up vs top-down processes in the visual system and focal attention. After Julesz (1990b).

without effort could return to my next sentence. However, no introspection would permit the reader to discriminate a yellow color from a matched red and green color. That is, according to the theory of trichromacy, known since George Palmer (who stated it in 1777) and later rediscovered by Thomas Young in 1802 (whose publications are reprinted in a historical collection by MacAdam, 1970), any color can be matched perceptually to three primary colors of given weights. While focal attention may inspect certain cortical stages, such as V4 (conjectured to be the color center by the neurophysiologist Semir Zeki, 1980), it does not have individual access to the red, green, and blue retinal receptors. In summary, focal attention can be drawn to many perceptual stages, but not to the earliest ones, and probably even most of the higher perceptual stages and states cannot be attended.

The fundamental problem of attention, particularly focal attention, means different psychological phenomena to different schools of psychologists and neurophysiologists. Here I stress the metaphor of focal attention as a shifting searchlight, as illustrated by Fig. 2. Here, two orthogonal line segments form elements of textural arrays. The X's pop out from the L's as a parallel process, i.e., almost instantaneously, independent of the number of elements. However, it requires time-consuming element-by-element scrutiny to detect an array composed of T's among the L's. These serial shifts of focal attention underlying scrutiny can take place rapidly without eye movements, as first observed by Helmholtz (1896) and elegantly studied by Posner (1978). [An English translation of this pioneering contribution by Helmholtz is provided by Nakayama and Mackenben (1989) from the second German edition after a century delay, because all existing English editions are based on the first German edition and Helmholtz added this important observation only to the second edition.] While evidence is in-

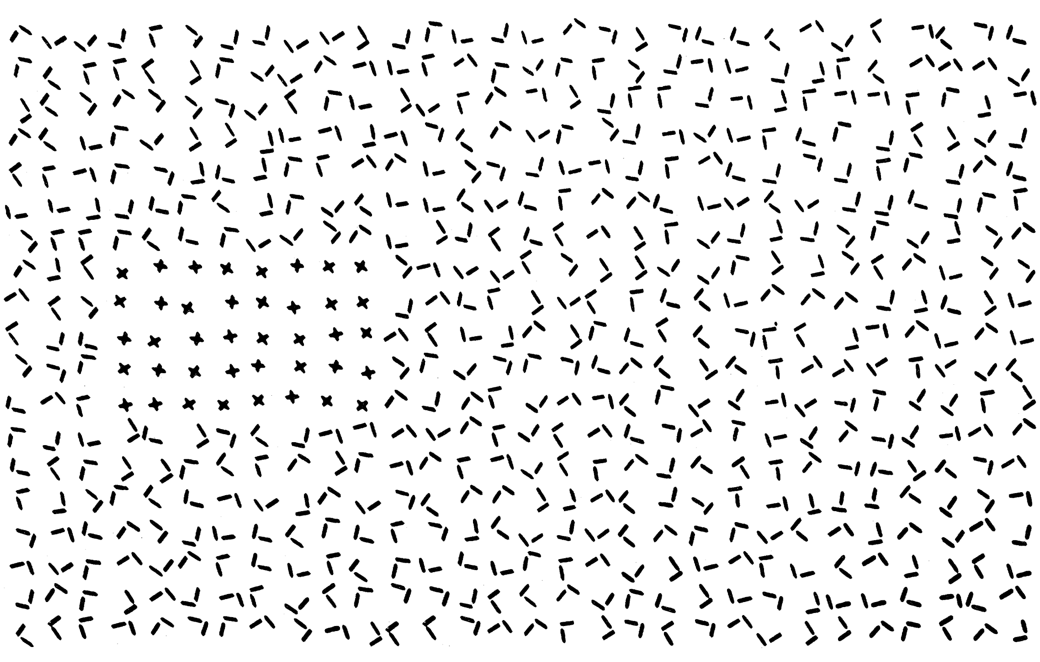


FIG. 2. Preattentive (parallel) texture discrimination vs serial scrutiny by focal attention. The X's among the L's pop out effortlessly, while finding the T's among the L's requires an element-by-element search. From Julesz and Bergen (1983).

direct, depending on which model is used to interpret the experimental results, the "searchlight of attention" scans about 30–60 msec/item (Sternberg, 1966; Treisman and Gelade, 1980; Bergen and Julesz, 1983; Weichselgartner and Sperling, 1987). Saarinen and Julesz (1991) have recently measured directly the scanning speed of focal attention by briefly presenting (with masking) numbers at random locations, and though observers had difficulty correctly reporting the order in the sequence, they could follow and identify as many as four consecutive numbers at rates of 30 msec/item with orders of magnitude above chance. Obviously this rate depends on the visibility of the texture gradients, and some parallel mechanism seems to facilitate serial search (Kröse and Julesz, 1989; Wolfe and Cave, 1990).

Because during our waking states we are bombarded by countless visual objects and patterns in a scene and can only make a few eye or limb movements in a given instance, much of the unwanted information must be filtered out centrally. Furthermore, we cannot store at a given moment more than a limited number of items (called "chunks") in memory, (usually 7 ± 2 chunks), which is another reason why we have to filter the amount of incoming information. It appears that, in order to inspect objects and events thoroughly, focal attention is needed. It seems as if there are only a very few processes (perhaps only one) of the highest level that can count as well as observe consciously and in great detail objects, patterns, and events, and it is focal attention that does the preselection and presents the selected patterns and events item-by-item to the highest levels. Indeed, we found (Sagi and Julesz, 1985) that observers were able to find the positions of texture gradients without scrutiny—they just popped out effortlessly (and did not depend on the number of elements in the arrays). However, to identify the features on the two sides of the texture boundaries (whose differences yielded the texture gradients) required serial search by focal attention (and the search increased monotonically with the number of elements). The question of what can we perceive preattentively while our focal attention is engaged (e.g., our ability to identify at some specific position a character and to identify as well some characteristic feature at a texture gradient) is a strategic problem, currently being studied by Braun and Sagi (1990). Preliminary results indicate that, for instance, some crude features of position can be perceived preattentively, but fine positional information requires scrutiny by focal attention.

Besides focal attention, which scans objects about five times faster than one can do with eye movements, there is another way to increase information intake. This is by learning increasingly complex chunks. This kind of learning is a high-level process, beyond the scope of my interest; it is in the domain of cognitive psychologists to study this important mental phenomenon. Here I only explain the main idea of "chunking." If the reader is very briefly shown, say, five pebbles, probably their number can be guessed without error, while with more pebbles mistakes will be made. However if the pebbles are arranged in chunks of regular pentagons, the reader can count up to five pentagons without errors, and thus can count 25 pebbles. Whether one can extend this learning of pentagons to increasingly complex chunks of pentagons is a typical question cognitive psychologists like to study.

The literature on attention is vast. The previous paragraphs emphasize the *single* searchlight metaphor that originated with Helmholtz (1896) and was neglected afterwards by cognitive psychologists for some time, as I alluded to. Another prominent figure of psychology, the American psychologist William James (1890) asked ". . . how many ideas or things can we attend to at once, . . . how many entirely disconnected systems or processes of conceptions can go on simultaneously; the answer is, not easily more than one, unless the processes are very habitual, but then two, or even three, without very much oscillation of the attention." James's question opened up an entirely new field of inquiry called divided attention. This is characterized by the metaphor of finite resources (or limited capacity). According to this metaphor the resources can be regarded as a "fuel," and one can change the amount of fuel allocated to various multiple (usually double) tasks. In a way, the question of what resources are left for perceiving certain features when focal attention is engaged belongs to the problem of divided attention, but in a more concrete sense than usually asked by cognitive psychologists. (The Saarinen and Julesz experiment discussed above—though conceived in the spirit of the searchlight metaphor—can support equally well the metaphor of divided attention.) The interested reader might consult the book edited by LeDoux and Hirst (1986), which contains a critical debate between psychologists and neurophysiologists, with four articles devoted to attention. The end result is somewhat disappointing, since there is no real consensus between workers in the two disciplines. However, the articles review the many theories of attention, including the bottleneck and variable filter theories by Broadbent (1958) and his many followers. For a recent review on some outstanding findings in cortical neurophysiology of attention, see the article by Desimone and Ungerleider (1989).

C. Structuralist versus Gestalt theories

Having introduced the basic concepts above, we return to the definition of *early vision*. Conceptually defined, "early vision" should be identical to the pure bottom-up visual processes depicted in Fig. 1, without being influenced by the top-down stream of semantic information. Neurophysiologically defined, "early vision" should correspond to the first neural processing stages in the retina and the visual cortex. Psychologically defined, "early vision" should encompass a range of perceptual phenomena that can be experienced by humans in the absence of higher cognitive and semantic cues. In the next section, we shall see that with the techniques of computer-generated random-dot stereograms (RDS), random-dot cinematograms (RDC), and texture pairs with controlled statistical properties (Julesz, 1960, 1962), it is possible to show that stereoscopic depth perception, motion perception, and preattentive texture segregation can occur in humans without the mysterious cues of form and Gestalt. This in turn permits us to link these mental phenomena to the neurophysiological findings obtained in the last three decades by probing with microelectrodes individual neurons in the early processing stages of the monkey cortex.

My structuralist approach of treating early vision in a thermodynamic fashion does not mean that I regard human vision in its entirety as amenable to such treatment. In real-life situations, bottom-up and top-down processes are interwoven in intricate ways, and the slogan of the Gestalt psychologists that "the whole is more than the sum of its parts"—a negation of the structuralist view of science—is probably true. Indeed, in Fig. 3 it is obvious that in the right upside-down image the eyes and mouth of the face have been manipulated. Because we are not familiar with inverted faces, the original face (left side) and the right one appear quite similar. Turning the page upside-down, so that the faces are now correctly seen, we experience a dramatic difference; the untouched face appears normal while the manipulated face looks grotesque! Obviously the mouth, the eyes, etc., are not simple building blocks of a perceived face; instead some global and highly complex interactions between them exist, and the concatenations of these parts into a Gestalt make the study of form recognition so frustrating at present. What I am suggesting is not that Gestalt phenomena be overlooked, but instead that an entire subfield of vision—early vision—be experimentally isolated and studied by the proven structuralist methods of the physicists. [For an epistemological review on the limits and merits of structuralist models in psychobiology, see Uttal (1990).]

While I assure the reader that I stick to the structuralist (reductionist) paradigm throughout this review, I acknowledge the important contributions that Gestalt ("configuration") psychologists have made in the first part of this century, from the principles underlying figure-ground organizations to perceptual grouping. One reason why their popularity waned was that their belief in electric brain fields whose convergence toward minimum-energy states was not supported by results gained from single-microelectrode neurophysiology. However, some of the Gestalt ideas have resurfaced recently under the guise of connectionist neural networks;

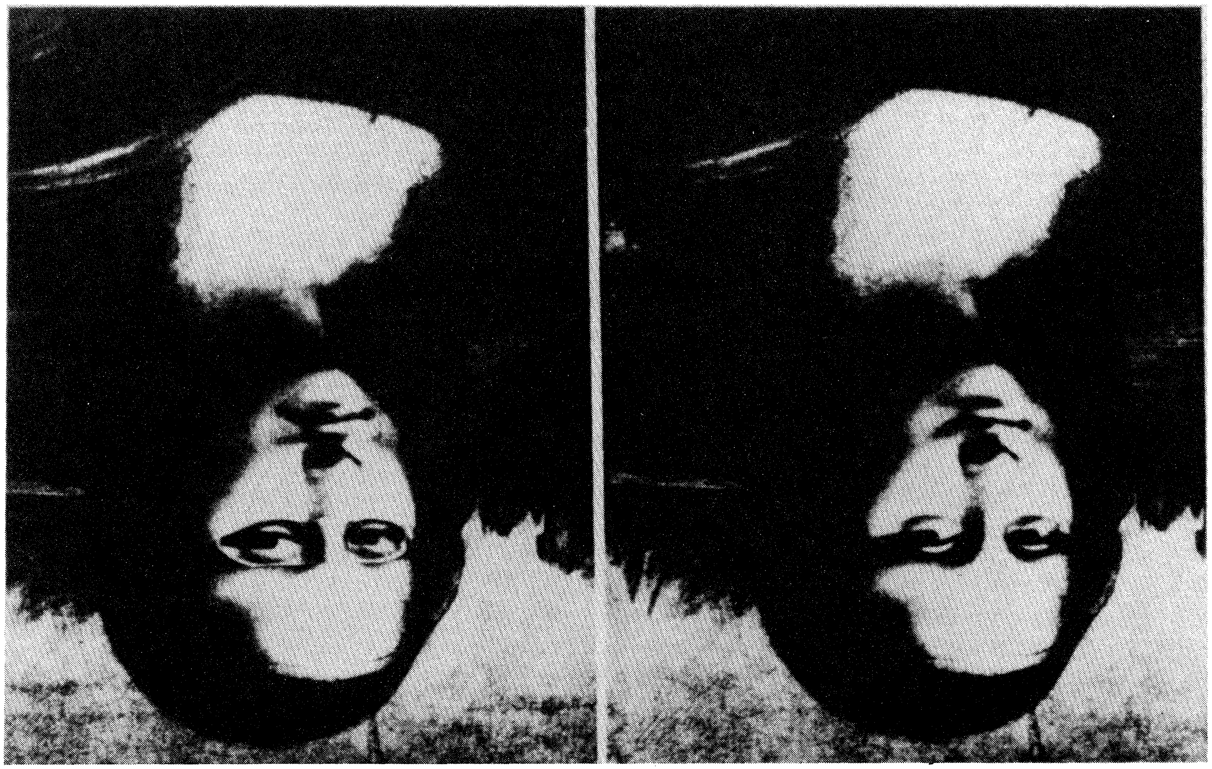


FIG. 3. Demonstration of Gestalt. The upside-down pictures appear rather similar in appearance, in spite of the fact that in one picture the eyes and mouth seem to be inverted. When the page is turned upside down, the two faces reveal a dramatic difference as a result of Gestalt organization. From Julesz (1984) after an idea of Thompson (1980).

the interested reader should see the brief and clear article of Rock and Palmer (1990). My own view on connectionist neural networks in early vision is given in Sec. III.D; however, I would not be surprised to see a neo-Gestalt revival explaining higher visual functions in the not-too-distant future.

II. ON THE CREATIVE PROCESS AND SCIENTIFIC BILINGUALISM

A. The manifold view of scientific interaction and conjugacy, an approach to discovery

From the Introduction it is quite clear that the study of the human cortex is so complicated that a gamut of disciplines—including sensory, perceptual and cognitive psychology, neurophysiology, neurology, neuroanatomy, embryology, neuropharmacology, mathematics, engineering, information theory, neural network theory, physics, and so on—is necessary to study its workings. The question arises of how any individual can cope with such a variety of different fields. A good answer is given by Michael Polanyi (1969), as follows:

> Even mature scientists know little more than the names of most branches of science. . . . The amplitude of our cultural heritage exceeds ten thousand times the carrying capacity of any human brain, and hence we must have ten thousand specialists to transmit it. To do away with the specialization of knowledge would be to produce a race of quiz winners and destroy our culture in favor of a universal dilettantism. . . . But how can anybody compare the scientific value of discoveries in, say, astronomy with those in medicine? Nobody can, but nobody needs to. All that is required is that we compare these values in closely neighboring fields of science. Judgments extending over neighborhoods will overlap and form a chain spanning the entire range of sciences.

The spread of knowledge between overlapping scientific disciplines according to this "manifold view," occurs naturally, similarly to the cooperation between members of a beehive. The only requirement is that specialists working, say, on the brain not have too narrow specializations, so they can indeed communicate with other specialists in overlapping areas of shared knowledge. Such cooperation can results in novel applications or technical breakthroughs. For instance, in the forties, electrical engineers started to develop special low-noise amplifiers that enabled neurophysiologists to record spike potentials in individual neurons. Another example is the development of surgical techniques to reduce epileptic seizures by splitting cortical hemispheres; these techniques enabled Sperry (1982) and his collaborators to study the mental competence of each separate hemisphere. For an up-to-date review of hemispheric localization, see Gazzaniga (1989). One may well ask whether there are ways to make this slow accumulation of knowledge more directed or conscious, thus accelerating progress. This question is intimately related to the essence of scientific creativity. While there must be many ways to create a new paradigm or get a novel insight, here I shall briefly discuss "conjugacy" and particularly "scientific bilingualism" as two approaches that can be used to advance science in general and brain research in particular.

The first approach, *conjugacy,* is the "trick" of establishing an equivalence relation between a difficult or unpected explored task (operation) and a familiar one, whose solution is already known. This is depicted in Fig. 4, where the difficult task **O** is to transport an object from point A to B through an impenetrable obstacle (wall). A possible way to complete this task is to drill a shaft **S** from point A, drill a tunnel **T** under the wall (assuming that drilling is a routine operation), and finally drill an inverse shaft $\mathbf{S}^{-1}$ to point B. One case in point is the facilitation of the operation **O** of multiplication (division) by introducing the logarithmic transformation **S** and its inverse $\mathbf{S}^{-1}$ that reduces the task to the much simpler operation **T** of addition (subtraction). Similarly, cross-correlation **O** of two functions can be reduced to the simple multiplication of the Fourier transforms **S** of the two functions, followed by taking the inverse Fourier transform. As a matter of fact, when neurophysiologists discovered Mexican-hat-shaped (Laplacian of a Gaussian) receptive field profiles of a concentric circular kind in retinal ganglion cells of the cat (Kuffler, 1953) and elongated field profiles in some orientation in the visual cortex of the cat and the monkey (Hubel and Wiesel, 1960, 1968), it was apparent that these spatial filter responses had to be cross-correlated with the visual image (brightness distribution cast on the retina). [The

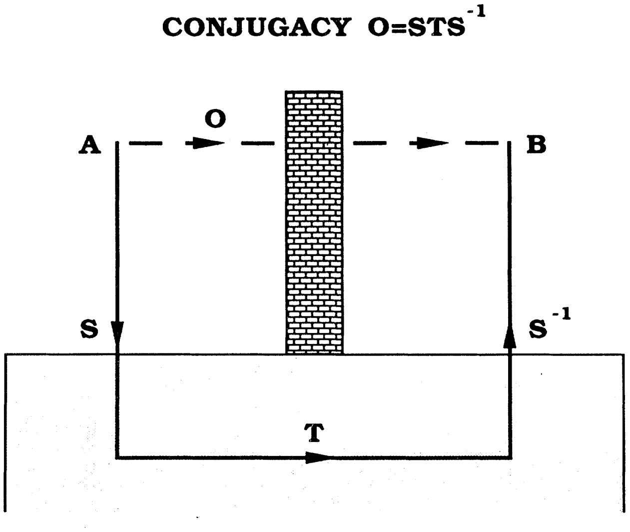


FIG. 4. The equivalence relation "conjugacy," or how to solve a difficult task by transforming it to an already familiar task.

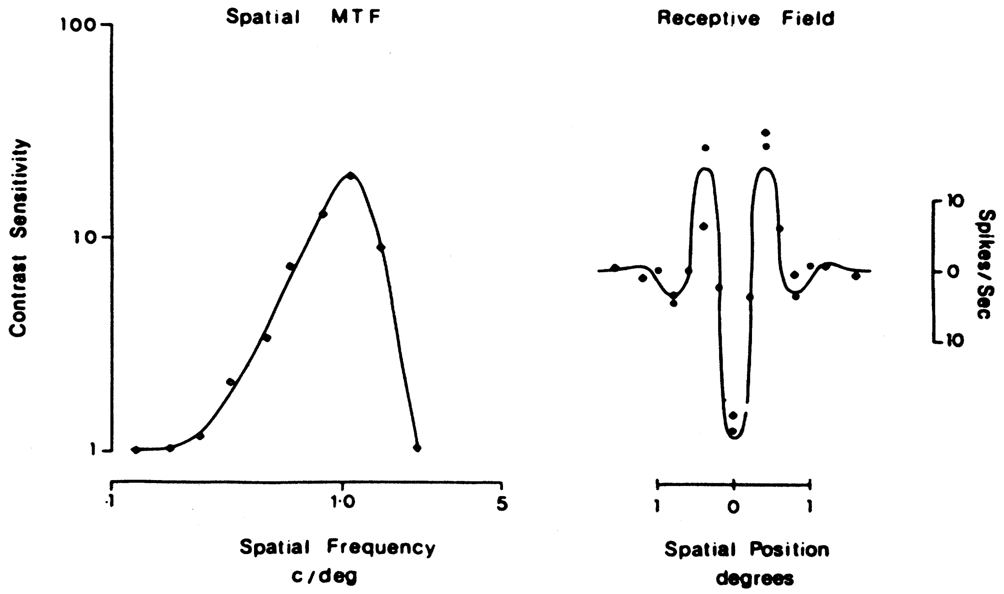


FIG. 5. Illustration of conjugacy in brain research. The neurologically determined double-Gaussian-shaped receptive field profiles and the psychophysically determined modulation transfer functions (by measuring the visibility thresholds to spatial luminance gratings with changing spatial frequencies) are related by the Fourier transform. From LeDoux and Hirst (1986).

reader unfamiliar with spatial filters might read the text and figure caption related to Fig. 15 below]. However, neurophysiologists were not at home in the cross-correlation domain, and it took the psychologists (Blakemore and Campbell, 1969) to show that sinusoidal luminance gratings presented at long duration would selectively fatigue the detection of test gratings when the test grating had spatial frequencies similar to the adapting grating. It is apparent from Fig. 5 that the visibility function of sinusoidal gratings, called the modulation transfer function (MTF), is a decomposition of perceptually weighted Dirac functions in the Fourier domain, which in turn is the Fourier transform of the neurophysiologically measured receptive field profiles. Had researchers been more familiar either with Fourier analysis or with cross-correlation techniques, they could have discovered the luminance grating adaptation paradigm years earlier, when the first neurophysiological findings were obtained.

I do not want to dwell on the use of conjugacy further, since the reader can make a long list of his or her own of novel unexplored tasks that have been transformed by some ingenious application of conjugacy to an already familiar one. (The interested reader should consult the inspiring mathematical monograph by Melzak, 1976.) However, an equally effective way to gain novel insight based on *scientific bilingualism* is less appreciated, and I turn now to this topic, after an analogy to linguistic bilingualism.

B. Scientific bilingualism, random-dot stereograms, and the false-target problem

The reader who is irritated by personal anecdotes should skip the first two paragraphs of this section.

The reader who speaks foreign languages in the same language family as his native tongue, say, Russian and French besides English, cannot appreciate the intellectual thrill one experiences when learning another language in a different language family. For a Hungarian child it was customary to learn an Indo-European language at an early age (since Hungarian belongs to the Finno-Ugric language family). I learned German and was quite surprised by the existence of a passive voice. In Hungarian one has to know the name of, say, a lecturer, in order to announce a lecture, since saying "someone will give a lecture on stochastic processes at noon" sounds comical even in Hungarian, while in all Indo-European languages one can easily announce that "a lecture will be given on stochastic processes at noon." This noncommittal passive voice is a great invention and is taken for granted by speakers of Indo-European languages, while Hungarians, Finns, Estonians, etc., have to struggle along with awkward situations that require them to know concretely who did what.

In 1959, after I arrived at Bell Laboratories as a young communications and radar engineer, I started to learn psychology, since I had joined a group dedicated to reducing the visual information in pictures without perceptual deterioration and I thought that some familiarity with human vision would be beneficial. Obviously, pictures are redundant, and the human visual system does not require half-tone images in their full detail; it can, for instance, recognize objects merely from their outlines, a fact exploited by cartoonists. But what is an outline (contour)? Could one design a machine (algorithm) that could emulate a skilled cartoonist by extracting meaningful contours from patches of various grays?

The idea of using the second spatial derivative (the Laplacian) for extracting contours was proposed and discussed by Ernst Mach (1886, reprinted 1959) and revived by David Marr (1982) and his followers, who averaged the image luminance with a Gaussian filter followed by the Laplacian. An image convoluted with the Laplacian of a Gaussian is almost identical to the double Gaussian-shaped (or Mexican hat function) receptive field profiles found by neurophysiologists in concentric form in retinal ganglion cells, and in elongated form in the striate cortex (V1), as portrayed on the right-hand side of Fig. 5. [These receptive field profiles are also quite similar to the cosine and sine functions weighted with a Gaussian envelope now called Gabor functions, in honor of Dennis Gabor (1946), who derived them as the optimal solution for the carrying of information simultaneously in both spatial and temporal domains.] Such simple linear operators applied to an image are surely inadequate to find contours and separate overlapping objects. For example, the image of a rabbit sitting before a fence has ears that could belong either to the fence or to the rabbit, and a great amount of familiarity with rabbits and fences is needed to separate the face from the background. Furthermore, the existence of "subjective contour" phenomena, discovered by Schumann (1904) and further elaborated by Kanizsa (1976), as demonstrated in Fig. 6, shows clearly that perceptually perceived contours exist in uniform areas where no luminance gradients occur, based on the linear continuation of quasicollinear line segments. In the cognitive psychology literature, subjective contour perception was traditionally regarded as a complex task of completing line segments by some higher-order reasoning. Actually, as neurophysiologists have recently shown (von der Heydt *et al.*, 1988), the completion of subjective contours occurs in V2 (an early cortical stage in the monkey cortex).

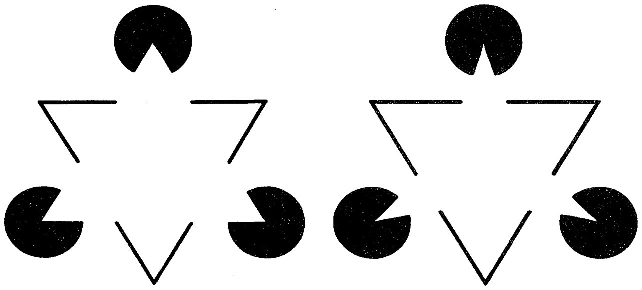


FIG. 6. Subjective contours as first described by Schumann (1904) and elaborated by Kanizsa (1976). One sees a triangle with a sharp boundary even though no luminance contours exist at most places. After Kanizsa (1976).

In essence, as I became acquainted with the psychological literature in 1959, it became apparent that the extraction of contours was intimately linked with the segmentation of objects, which in turn was based on the little-understood cues of semantic memory. So when I suggested using stereoscopic depth perception (stereopsis), based on horizontal disparity, to segment objects, it was pointed out that in order to find in a crowd the corresponding faces in the left and right retinal projections, one would first have to recognize the faces—a most complex feat, whose execution is shrouded in mystery. Already with four targets, not knowing which retinal projections correspond to each other, as depicted in Fig. 7, sixteen possible localizations can be made, out of which only four are correct and the rest are false (phantom) targets. With N targets, the number of phantom targets increases as (N^2-N). Thus, according to prevailing beliefs, monocular form cues and contours were essential for stereopsis, a belief shared by the leading expert of binocular depth perception of this period, Ken-

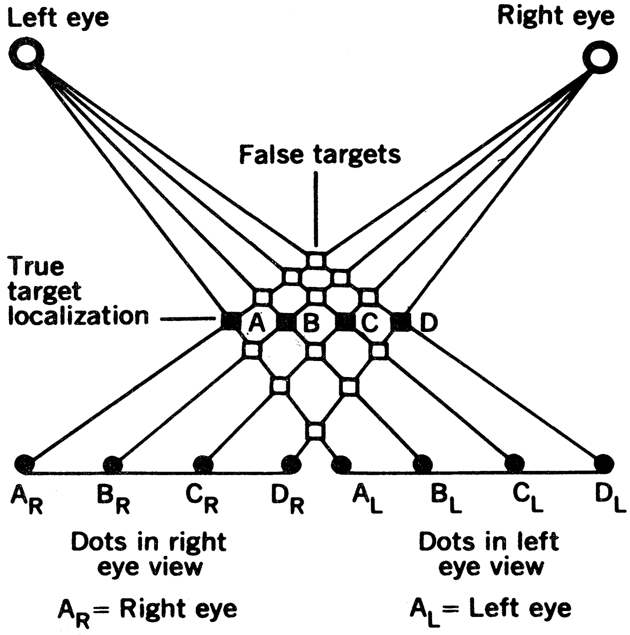


FIG. 7. Binocular matching, or the problem of false-target elimination. The four identical targets (dots) yield 16 possible localizations, out of which only four are correct. Without monocular labeling only global constraints (such as cross-correlation) can eliminate the false matches. From Julesz (1971).

neth Ogle (1964). As a former radar engineer, I knew that this view of the psychologists could not be valid. After all, in order to break camouflage in aerial reconnaissance, one would view aerial images taken from two different positions (the use of parallax) through a stereoscope, and the camouflaged target would jump out in vivid depth. Of course, in real life there is no ideal camouflage, and after stereoscopic viewing one can detect with a single eye a few faint cues that might discriminate a target from its surroundings. So I used one of the first big computers, an IBM704 that had just arrived at Bell Labs, to create *ideally camouflaged* stereoscopic images as shown in Fig. 8(a). Since the binocular disparity is an *integer multiple* of the randomly selected but correlated black and white dots (pixels) that make up the left and right arrays, these images, when viewed monocularly, are only aggregates of random dots with no break or gap between areas of different disparities. However, when binocularly fused, the correlated areas segregate in vivid depth according to these monocularly invisible disparities (Julesz, 1960, 1971). For a personal history with more detail, see Bernstein (1984), or my review in the 25th anniversary Silver Jubilee issue of *Vision Research* (Julesz, 1986a), and also Julesz (1990a, 1990b).

Figure 8(b) is similar to Fig. 8(a), except that the density of the random-dot stereogram (RDS) is reduced to a few percent. It is as easy to fuse as the dense RDS, and it is interesting to observe that the visual system interpolates a surface in depth where there are no dots present (as if it were a "convex hull"). [This "filling in" phenomenon inspired many model builders of early vision; the interested reader might consult Grossberg and Mingolla (1985).]

I have dwelt on this historical account merely to stress the point that the introduction of the RDS to psychology was just a lucky realization by a radar engineer that the common knowledge in aerial reconnaissance that *camouflage cannot exist in 3D* was unknown to psychologists. It was this *scientific bilingualism* that led to the discovery of RDS, although there must have been dozens of stereophotographs taken from balloons or airplanes of densely textured scenes that emulated random-dot stereograms quite well. (Babington-Smith, 1977, reprints a

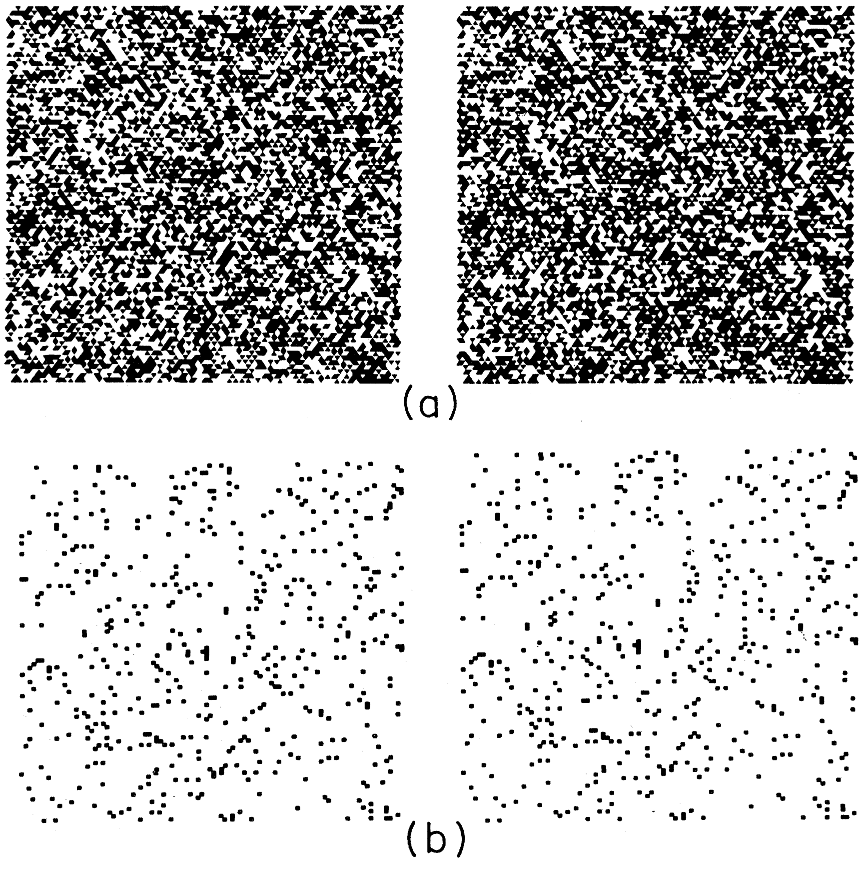


FIG. 8. A computer-generated random-dot stereogram (RDS) that, when monocularly viewed, appears as a random array of black and white dots (pixels), yet when binocularly fused (with the aid of a prism or by crossing the eyes) reveals a center target, which jumps out in vivid depth. (a) RDS with 50% density of black pixels; (b) RDS with 5% density of black pixels. From Julesz (1971).

British aerial reconnaissance photograph of Cologne taken from two different vantage points during WWII, in which one can see the floating ice on the River Rhine. When binocularly fused, the pieces of ice portray interesting surfaces in depth, particularly at the vicinity of the piers of a bridge, where the ice flow is altered.) Of course, the arrival of computers permitted *ideal* camouflage with absolute stimulus control, and nowadays one can generate *dynamic* RDS at a rate of 60 frames/sec or faster, portraying moving surfaces in depth; when these are viewed monocularly, only dynamic noise (snowstorm) is perceived.

[I referred above to the influential work of David Marr (1982) and his followers. In this historical recollection I want to point out that their ideas, particularly the close relationship between the early contour extraction in 2D and the 3D information obtained through binocular disparity, was strongly influenced by the findings with RDS a decade earlier and led to their "2.5-D sketch model." In the next section I shall discuss some of their contributions to the field of global stereopsis and give a critical evaluation of the role of the artificial intelligence approach in brain research.]

The essence of this somewhat personal note is that science usually progresses haphazardly through overlapping fields or disciplines, but more directed, conscious contributions can be made if a scientist is willing to learn two *remote* disciplines and apply this scientific bilingualism. This is nowhere more true than in brain research. For instance, physics and psychology, or physics and neurophysiology in my opinion are excellent backgrounds for attacking problems in psychobiology.

III. NEUROPHYSIOLOGY VERSUS ARTIFICIAL INTELLIGENCE

A. Global stereopsis and its linking to neurophysiology

I now return to my original purpose, which was to study stereopsis—and psychological phenomena in general—by emulating thermodynamics. The reader is able to see the "cyclopean" square in Fig. 8 after binocular fusion (aided by using an optical wedge to cross eyes, and wearing prescription glasses to correct for astigmatism—assuming, of course, that he or she does not belong to the 2% who are stereoblind); from this ability it becomes apparent that stereopsis is a bottom-up process that must occur before form perception, since the left and right images are devoid of all monocular cues, including shapes and their contours. Furthermore, it can be shown in the laboratory (Julesz, 1964; Julesz and Chang, 1976) by masking the stimulus 60 msec after the stimulus onset—and thus terminating the availability of the stereogram—that one can easily fuse and perceive complex surfaces in vivid depth during such a brief presentation, provided the maximum binocular disparity is within a limit, called Panum's fusional area. With dynamic RDS (which nowadays can be easily generated on personal computers and can be stored on video tape recorders), each subsequent frame contains uncorrelated random dots (but similar binocularly correlated areas), and these uncorrelated frames erase (mask) the previous frames. Therefore, in addition to the fact that the RDS are devoid of all familiarity cues, quick masking *prevents* the top-down processes from penetrating down in time and influencing the bottom-up processes of stereopsis.

There are many demonstrations that stereopsis of RDS (which I often refer to as either "global stereopsis" or "cyclopean perception") must be an early process, based on some correlationlike process prior to object (form) recognition. Here, I only take one example from my monograph (Julesz, 1971) showing that optical illusions must occur after binocular combination of information, that is, after several synaptic processing stages in the retina, lateral geniculate nucleus, and cortex. This technique of process localization without a scalpel, which I called "psychoanatomy," enables us to trace the information flow in the visual system by portraying visual information, not with the usual luminance gradients [Fig. 9(a)], but with binocular disparity gradients [Figs. 8(b) and 8(c)], using RDS (Papert, 1961; Julesz, 1971).

Figure 9(a) shows the classic Ebbinghaus illusion, in which identical center disks appear to have different sizes. When the reader inspects the cyclopean version of this illusion, by fusing the RDS of Fig. 9(b), it is apparent that the illusion is the same. Therefore the many processing stages prior in the "cyclopean retina" (an assumed stage at which binocular correlation first occurs) do not affect the illusion, and the optical illusion has to arise afterwards. With the technique of RDS one gains a new degree of freedom and can portray the center target at a different depth from the inducing figures, as shown in Fig. 9(c). When fusing Fig. 9(c), the reader can verify that the illusion is greatly reduced, if not gone. This suggests that optical illusions are the result of lateral interactions between neural pools tuned to the same binocular disparity, while pools tuned to different disparities do not interact with each other (Julesz, 1971).

Thus the RDS, since its inception in 1960, suggested that stereopsis must be an early process and much simpler than form recognition. It was the neurophysiologists who first realized these implications and switched from studying form to exploring binocular vision (Hubel and Wiesel, 1962, 1970; Barlow, Blakemore, and Pettigrew, 1967; Bishop, 1969; Blakemore, 1969). The perceptual finding that one can adapt to selected depth planes in a RDS by prolonged viewing and, as a result of fatiguing neural pools tuned to the corresponding binocular disparities, modify the amount of perceived depth (Blakemore and Julesz, 1971) suggested the existence of neurons tuned to specific binocular disparities. Nevertheless, only in 1984 did Gian Poggio find neurons as ear-

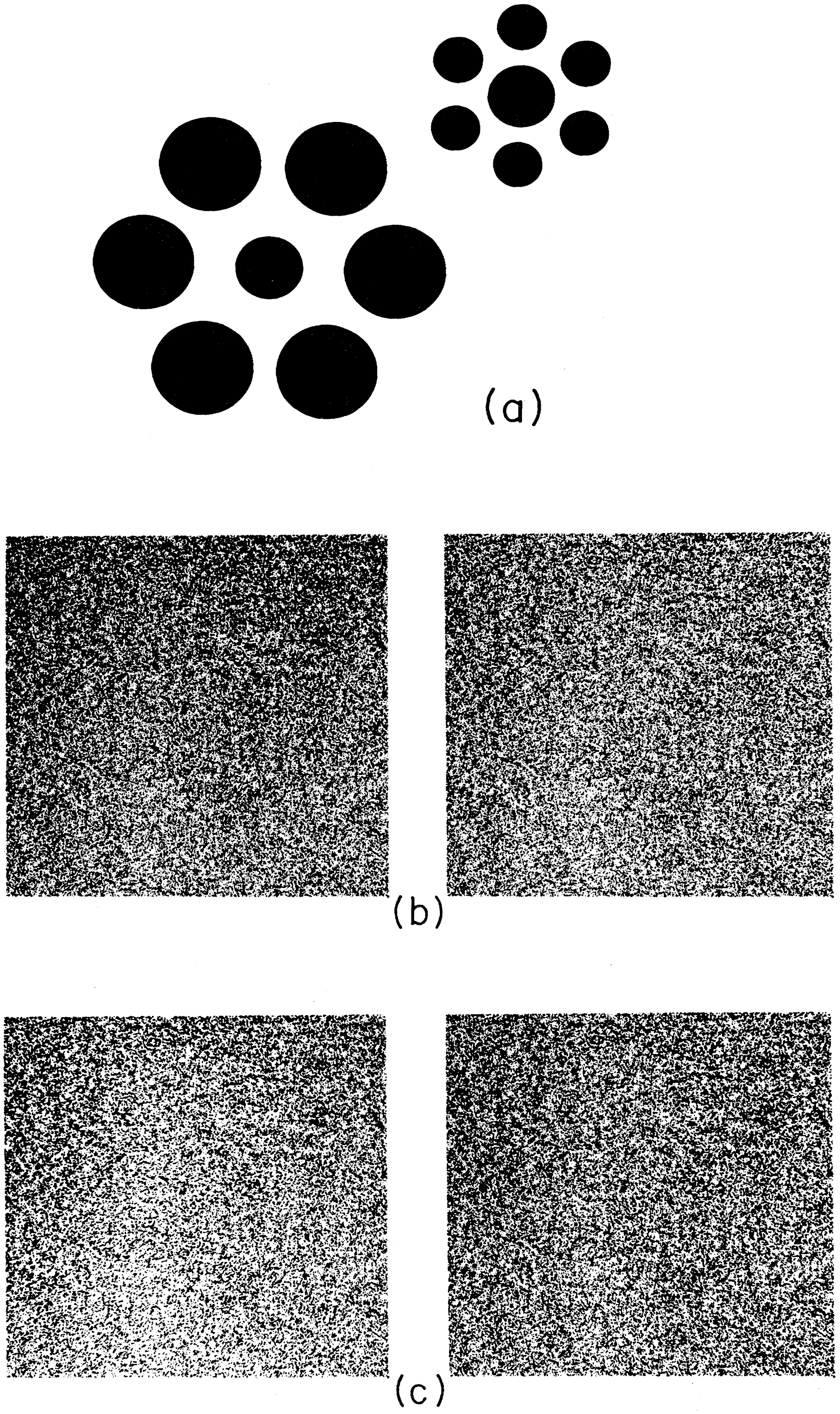


FIG. 9. Classically and cyclopeanly portrayed Ebbinghaus illusion. (a) The classical illusion portrayed by luminance gradients. The test figures in the center are actually identical, but due to differences in the inducing figures, they appear of different size. (b) The same illusion cyclopeanly portrayed by the method of RDS. Test and inducing figures are in the same depth plane when binocularly fused, and the perceived illusion is similar to the classical one. (c) Same as (b), except that the test and inducing figures are in different depth planes, and the perceived illusion is much reduced, if seen at all. From Julesz (1971).

ly as layer IVB in area V1 (the input stage to the visual cortex of macaque monkeys) that selectively responded to binocular disparities in dynamic RDS (Poggio, 1984). Since his stereograms were presented at rates of 100 frames/sec, there was no monocular contrast between the correlated areas, and therefore these early processing stages were able to solve the binocular matching problem (shown in Fig. 7) of performing correlationlike processing. While in area V1 only 20% of the probed units were cyclopean, as Poggio *et al.* (1988) probed higher areas, V2, V3, and V3A, the majority of neurons were found to be cyclopean (i.e., fired for dynamic RDS).

Poggio's discovery of cyclopean neurons in the input stages of the cortex confirms the psychological predictions that global stereopsis (i.e., stereoscopic depth perception of RDS) is an early process and neurophysiologically determines the first stages of the "cyclopean retina." That global stereopsis is such an early process has another interesting implication. Since in 3D there is no camouflage, stereopsis probably evolved in our insectivore primate predecessors (e.g., lemurs) rather late in the evolutionary time scale, in order to counteract the freeze response of insects, who would blend into the foliage at the sign of danger. [That in general there is no camouflage in three dimensions is the main insight gained by RDS; nevertheless, in a few rare cases, insects evolved such that their 3D shape mimics the 3D shape of leaves or reliefs of tree branches.] One would expect that such a late development would be relegated to an "attic" at some unused area of the cortex. Yet the emerging mechanisms of stereopsis were important enough to push aside existing machinery and grab the input stages of the visual cortex!

Perhaps the most important aspect of these developments is the linking of a rather complex mental event—global stereopsis—to some neurophysiological event—the firing of neural pools tuned to specific binocular disparities. As I mentioned previously, such a link was sought between neurophysiology and the sensory psychology of color vision years earlier, but this is the first time in the more complex field of perceptual psychology that such a link between mind and brain was found.

[When it comes to linking hypotheses between mind and brain, different schools of psychobiology have rather different criteria. I found a relevant essay by Davida Teller (1980) most entertaining. While writing this article, I came across a rather novel linking criterion by Salzman *et al.* (1990), who applied cortical microstimulation to an extrastriate area (MT or V5) of the monkey that plays a prominent role in extracting motion information. This *electrical* microstimulation biased the animals' *perceptual* judgments of motion. This finding implies that physiological events at the neuronal level can be causally linked to a specific aspect of perceptual performance. While the authors used motion in their study, they pointed out that they could also have applied the same experimental paradigm to color or depth perception.]

B. Global stereopsis is a bottom-up and cooperative process

One might wonder whether computer-generated RDS with rich disparity cues might not be a special case, and whether top-down processes might still interfere with global stereopsis in natural scenes when higher cognitive cues are present. An example might be depth from shading, noted by astronomers when viewing the craters of the Moon. When viewed upside-down, the concave craters appear to be convex mounds. Such a depth reversal is demonstrated in Fig. 10 when each of the stereo half-pairs is viewed monocularly, depending on whether the illumination comes from above or below. This "shape from shading" or "monocular depth from shading" might be construed as a top-down phenomenon interfering with the bottom-up processes of stereopsis. Indeed, Ramachandran (1988) exploited this phenomenon in his study of apparent motion perception and suggested that the "shape-from-shading" process operates prior to motion perception. The motion perception studied by Ramachandran is of the long-range kind that often disambiguates false matches by higher-order top-down processes, yet it is of interest to know whether the bottom-up process of short-range motion perception, which occurs for random-dot cinematograms and particularly for global stereopsis, is also influenced by shape from shading. After all, this phenomenon must be of a high-level kind, based on the fact that Earth has only one Sun that shines from above. If global stereopsis were to utilize such a complex top-down process, based on some learned or genetically inherited information, we would have a counterexample of global stereopsis' being based on early visual processing alone. In order to test this "counterexample," Jih Jie Chang and I (Chang and Julesz, 1990) constructed a randomly speckled egg-crate pair portraying the convex (or concave) depth-from-shading phenomenon when monocularly viewed (Fig. 10). However, 30% of the egg-crate pair is speckled with a RDS having a crossed disparity (if the reader views the stereogram with crossed eyes). Whether the reader views these figures right side up or upside-down, it becomes apparent that the convex shape determined by stereopsis will dominate depth from shading. Variations of this experiment, including *ambiguous RDS* (with two cyclopean shapes at different depth, but only one of them perceptible at a given instant), demonstrate that the monocular cue of depth from shading is rather weak (since observers perceive depth according to their natural bias) and therefore does not influence global stereopsis (Chang and Julesz, 1990). In summary, it is most likely that global stereopsis is mediated by early visual processes without top-down influences. It is, in the usage of Fodor (1983), a "cognitively impenetrable module."

In the previous paragraph, the problem of short- versus long-range processes in motion perception and stereopsis was briefly mentioned. My original use of "global" stereopsis and motion perception in RDS and

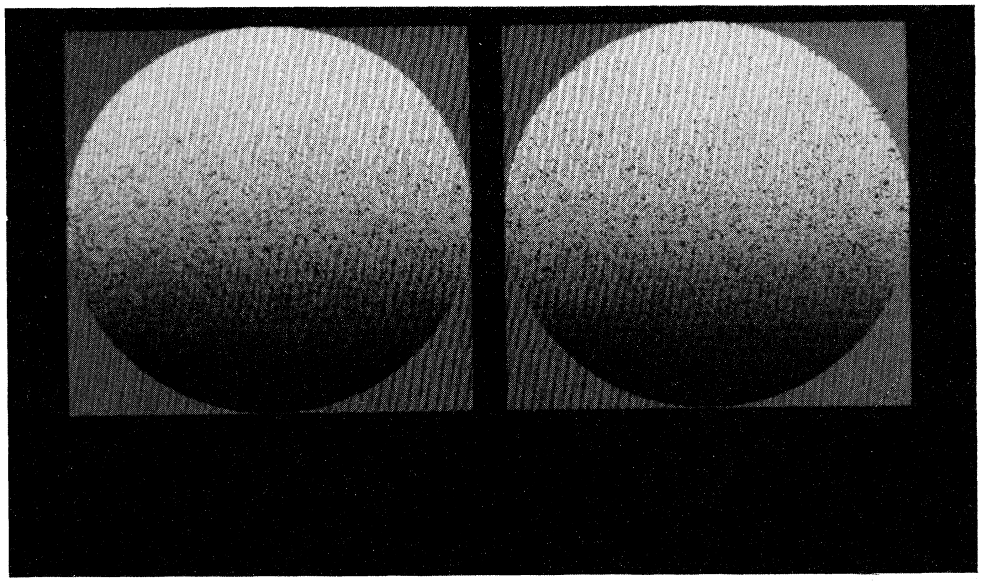


FIG. 10. Monocularly convex (concave) sphere due to depth (shape) shading, in which a random-dot stereogram (RDS) is mixed in. When binocularly fused, depth from stereopsis dominates perceived depth. When viewed monocularly, depth from shading yields a convex sphere, but appears concave when page is turned upside-down. Because the RDS has crossed binocular disparity, the binocularly fused image appears strongly convex. From Chang and Julesz (1990).

RDC, respectively, referred to the solving of the false-target problem *globally* (by cross-correlation), since the thousands of individual random dots could be paired in billions of incorrect ways. With increased disparities between the left and right half-pairs in a RDS, or between successive monocular arrays in a RDC, the number of possible false matches increases. Thus for global stereopsis and motion perception, the finding of matches within a short range is relatively easy, whereas over a long range it rapidly becomes difficult. However, for patterns that can be *locally* identified (by their color, shape, etc., as in a neon sign), the problem of correct matching is virtually independent of disparity, and stereoscopic depth or apparent motion can be obtained over a long range. Hence "local stereopsis" and "local motion perception" can also be called "long-range motion" or "long-range stereopsis." My use of local versus global is discussed in Julesz (1978) and Chang and Julesz (1983); the use of short- versus long-range motion is also discussed by Braddick (1974, 1980) and Snowden and Braddick (1990), as well as in a critical paper evaluating these dichotomies by Cavanagh and Mather (1989).

Let us turn now to the problem of artificial intelligence (AI) models. One could assume that, instead of neurophysiology, AI modeling could assume the role of an (I-1)-level description. While this is a theoretical possibility, it turns out that AI models are too robust—i.e., diametrically opposite models can explain perceptual phenomena equally well—and therefore one has to perform psychophysical and neurophysiological control experiments to check the biological plausibility of these models.

Besides the binocular matching problem (or false-target elimination problem) discussed in Fig. 7 there are three basic psychological findings of global stereopsis that model builders have to take into account. First is the finding the global stereopsis of RDS is *cooperative*—exhibiting multiple stable states, disorder-order transitions, and hysteresis. (For details, see the monograph and reviews by Julesz 1971, 1978, 1986a, 1990a). Here, I only mention an experiment with ambiguous RDS in which a bias of a few percent of unambiguous dots with given binocular disparities can perceptually affect the perceived depth (Julesz, 1964, 1978; Julesz and Chang, 1976)—a typical cooperative effect. The second finding is that global stereopsis utilizes spatial-frequency (SF)-filtered channels (Julesz and Miller, 1975). The spatial frequency spectrum of a left and right stereo half-image pair must overlap to some extent to obtain fusion. The third finding relates binocular disparity gradients to fusion and postulates "forbidden zones" where fusion cannot occur because the gradient exceeds a critical value (Burt and Julesz, 1980). That article demonstrates how nearby objects "warp" the disparity space, creating such forbidden zones.

C. Cooperative and noncooperative models of stereopsis

The brief (60 msec) processing time for global stereopsis and its cooperative behavior suggests a *parallel* mechanism. Furthermore, when one of the states of an ambiguous RDS is biased by 4% (in favor of the other state), a serial search mechanism would often stop at an almost perfect 96% binocular correlation, whereas, in fact, the stereopsis mechanism always chooses the perceptual state that yields 100% correlation, proving that not only the correlation mechanism but also the search mechanism for the best binocular match is parallel (Julesz and Chang, 1976). That perceptual phenomena are based on parallel processing is not a great surprise. Not only do 30 years of neurophysiological evidence point to this conclusion, but it is borne out by the following Gedanken-experiment: It takes about 500 msec or less to perform routine perceptual tasks (e.g., to recognize a face in a complex scene or to identify a cartoon), whereas a synaptic event (computation) requires 2–4 msec. Therefore the brain might utilize at most 100 stages of computational layers (instructions, or codes), which is in striking contrast to the thousands of codes necessary for AI algorithms running on the ubiquitous serial computers, which in addition perform perceptual tasks rather poorly.

My AUTOMAP-1 model was the first computer algorithm that could explain in successive iterations some of the cooperative phenomena of global stereopsis (Julesz, 1964). In order to give a better insight into the workings of such a model, I intentionally introduced a "spring-coupled-magnetic-dipole-model" (Julesz, 1971, 1978) that explained the pulling effect and plasticity, i.e., the fact that the left image of a RDS can be zoomed up or down in size by 10% with respect to the right without loss of fusion. (The reader interested in the history of stereopsis models might consult Julesz, 1990b.) Simultaneously to my cooperative spring-coupled-dipole model, Sperling (1970) proposed an "energy-well model" that had a heuristic appeal too, while Dev (1974) and Nelson (1975) offered cooperative models based on spreading inhibition between disparity neurons tuned to different disparities, and facilitation between neurons tuned to the same disparity. Marr and Poggio (1976) also published a cooperative model that resembles that of Dev and Nelson, in which inhibition occurs between disparity detectors tuned to different disparities, but these detectors fall on the same lines of sight. (This requirement seems important, since it has a heuristic appeal, but in fact is not, so the model is very similar to those mentioned above.) Perhaps the greatest contribution of the Marr and Poggio (1976) model is its easy implementation as a computer algorithm, even on PCs. Recently, Tomaso Poggio (1989) at MIT implemented a very fast stereo algorithm on the Connection Machine, and Mohowald and Delbrück (1989) at Caltech designed a one-dimensional version of the Marr and Poggio (1976) model on a chip that works very fast and solves the false-target problem in milliseconds. Cooperativity underlies even the computation of stereoscopic acuity (which is an order finer than visual acuity), as suggested by Westheimer (1979).

While most models of stereopsis are cooperative, Marr and Poggio (1978) invented a second, noncooperative model of stereopsis. Their model is based on the use of the spatial frequency channels found by Julesz and Miller (1975). According to this model, coarse SF channels in the two retinal projections—which tolerate large disparity searches without false matches—are aligned first by vergence (convergence or divergence) movements of the eyes, and then by finer and finer matches of high-SF channels by consecutive vergence movements—under the assumption that increasingly fine image detail has a decreasing amount of disparity. Indeed, it is a reasonable assumption that only extended objects protrude in depth (except for thin wires, which we often trip over), and fine surface textures in relief have shallow depth.

Unfortunately, this ingenious scheme is not used by the human visual system. Among others, Mayhew and Frisby (1979) and Mowforth *et al.* (1981) have shown that high-SF RDS, with low and medium SF filtered out, can still elicit large vergence movements and yield fusion at large disparities during a brief flash (when vergence movements are prevented). The finding that the Marr and Poggio (1979) noncooperative stereo algorithm based on vergence eye movements is not used by the visual system does not mean that a modified version could not be used, based on neurological couplings such that low-SF channels would monotonically reduce the disparity deviations of the high-SF channels with increasing SF. Such a model, based on a neuronal shifter hierarchy, appears much more complicated than a model based on a connected neural net exhibiting cooperative phenomena, but it could, nevertheless, exist. For instance, Anderson and Van Essen (1987) have proposed models for such neural shifter networks in the early stages of the visual cortex, but until now no such shifter nets have been found. The fact that we see many objects at different velocities as equally sharp (without using a "shutter" as in a movie camera)—even though we cannot follow simultaneously all these moving objects by eye movement—also argues for a cortical shifter mechanism that can compensate for the displacements of these drifting objects within limits.

Of the many AI models of stereopsis, I mention a recent one by Lehky and Sejnowsky (1990) mainly because they discuss the interesting debate in brain research between those who favor local and distributed representations. Retinal position is locally represented by a very large number of neurons that fire when specific very narrow channels (positions) are stimulated. On the other hand, the many shades of color sensations are not conveyed by thousands of individually tuned neurons, but by the population of some broadly tuned channels for the three principal colors. Global stereopsis contains both systems of narrowly and of broadly tuned binocular disparity channels, as found by Gian Poggio (1984). It happens that Lehky and Sejnowsky postulated a model based on the distributed representation of the broad disparity channels. Here again one can see the robustness of AI models, since one could easily imagine that a dual model, based on local representation, might have worked equally well.

D. The role of artificial intelligence models, including spin-glass and connectionist models, in psychobiology

I have presented the noncooperative Marr and Poggio (1979) stereo algorithm to show that, after a psychological problem is clearly posed (i.e., the false-target problem in RDS), both cooperative and noncooperative models can offer a solution in a robust way. Furthermore, additional bottom-up processes (modules) based on occlusion, shape from shading, limits on disparity gradients, opacity, etc., help to disambiguate these already robust stereo models even more. Ultimately, only psychological and neurophysiological evidence can tell these models apart—that is, which kinds of models are certainly not used by the human visual system and which remain potential candidates. Since the human visual system evolved over millions of years, it is advisable that model builders, including practitioners of connected adaptive networks (e.g., Rumelhart and McClelland, 1986), be keenly aware of psychobiological insights before they invent some *ad hoc* schemes. By the same token, some insights from AI, like the influential work of David Marr (1982) or the unexpected analogies between human memory and spin glass drawn by Hopfield (1984), might inspire psychobiologists to search for some hidden structures or organizational principles in the brain. However, these insights by AI researchers are rather metaphors than deep analogies, and although they may be inspiring, the physicist should be warned that the purpose of vision is most likely not to look for generalized cylinders ("stick figures") as Marr claimed (after all, vision evolved prior to animals with axial symmetries), nor is the human brain—particularly the modules (transducers) of early vision—as similar to a spin glass or a parallel computer as devotees of parallel distributed processing (PDP) might like us to believe! However, it is intellectually exciting that the higher association areas (outside the scope of this review) might be modeled by neural computation that exhibits flow to attractors of the network dynamics.

Connectionism, particularly as represented by neural networks of the PDP kind (Rumelhart and McClelland, 1986), is popular among physicists who are not familiar with brain research but understand Hamiltonians, "strange attractors," thermal annealing, and so on. These physicists should be particularly on the alert. While the invention of "hidden layers" and convergence theorems of learning through "back-propagation" have extended the processing power of "perceptrons" considerably, it should be noted that the simplest network with a hidden layer is the "exclusive OR" (with three gates), and to learn this logical function from scratch requires over 500 trials. Now, let us assume that the number of trials for learning a task increases monotonically (linearly, polynomially, or exponentially) with the number of gates. A special mode of genetic learning (evolution) is to change the connectivity of neurons in the brain of a given species at each mutation. So, to develop the human brain with its 10^{15} gates (synapses) might require eons, and even at a high mutation rate, the time of life on Earth is too short to have accomplished this feat by PDP techniques. This criticism of mine is based on the monotonic increase of learning time with the number of synapses in a neural net; it would be groundless if one could prove that learning time stays invariant with the size of the network—an unlikely event. Indeed, Judd (1987) gives a proof that, in general, supervised training in connectionist networks is NP-complete; thus learning time goes up exponentially with the number of synapses (gates or nodes) in a network. Judd shows that, even for very restricted networks, teaching (loading) time to set the connectivity of the gates requires polynomial time. Based on the learning (loading) time's being a polynomial or exponential function of 10^{15} gates, and thus orders of magnitude larger than the age of our universe, where the unit of each learning step is "one tick of the clock of mutation" for a given species, the unlikelihood of the evolution of connectionist neural networks seems to be correct. Because humans are particularly bad at logical tasks of the exclusive OR kind, one could argue that this criticism of PDP is irrelevant, since neither people nor machines are good at this task. Of course, I could have selected hundreds of tasks of form recognition in which people excel and PDP algorithms fail miserably. For a more extensive critique, see Pylyshyn (1984) and Fodor and Pylyshyn (1988); for discussion of the "top-down versus bottom-up paradigm" in neurophysiology, cognition, and PDP, see Churchland (1986). For an interesting discussion of evolution, brain research, and AI models, see Edelman (1989).]

E. A strategic problem of 3D, mathematical models of perceptual constancies, and Lie germs

Perhaps the success of global stereopsis is owing in part to the likelihood that the problem *"how can the visual system reconstruct from two 2D retinal projections a best possible 3D representation of the environment?"* might prove to be a strategic question. Similar variants of this question, for example, *"how can the visual system reconstruct a 3D relief from a single 2D retinal projection containing the reflectance distributions of Lambertian surfaces from a specified light source?,"* have been successfully attacked by Horn and Brooks (1985), Pentland (1986), and Bülthoff and Mallot (1990), among others.

The reader may have noticed that, instead of superficially reviewing thousands of topics in vision research, I have concentrated on a few, with emphasis on, say, the false-target problem of stereopsis, and through such select problems I have introduced the current state of my field. Of course, even in stereopsis I have had to skip dozens of topics, from problems of stereo acuity to absolute depth. Because the latter is quite important, I end this section with a brief discussion of absolute depth perception. Stereopsis, that is, depth perception based on binocular disparity, utilizes only the *horizontal* component of the disparity, and slight misalignments in the vertical direction cause a rapid loss of perceived depth (Ogle, 1964; Julesz, 1971). Furthermore, stereopsis yields only *relative* depth. In machine vision one could use vertical disparities together with horizontal ones to compute absolute depth; human stereopsis, however, processes only horizontal disparities. Sensing the convergence angle of the eyes is a very coarse approximation and cannot be used to compute absolute depth by triangulation. However, with increasing convergence angles, the perceived depth also increases. [Fusing Fig. 8 with different prism angles will influence the perceived depth of the hovering square.]

Here I should like to mention the pioneering work of Hermann von Schelling (1956), who performed a very simple experiment (by looking at his outstretched palm and observing that, as he moved his head around the pivot of his neck, the perceived distance of the palm did not change) and made a sophisticated mathematical argument. He showed (by modifying Cayley's invariant metric under projective transformations) that a *unique metric* exists, that is, a distance can be defined that stays invariant under *any affine transformation,* provided the space has a constant negative curvature (i.e., the space is *hyperbolic).* If F is a point in space that is fixed (a finger, or our palm), and R and L denote the focal points of the right and left eyes, and O marks the pivot of the neck, the tetrahedron FRLO should be the subjectively fixed frame for binocular depth perception, independent of head and eye movements. Since an affine transformation changes any tetrahedron into another one, Schelling's assumption states that binocular vision should be invariant in a hyperbolic space, an insight whose truth has been verified by elaborate psychophysical experiments carried out by Luneburg (1950) and his followers. I have chosen this example to illustrate that a good mathematical background permits researchers in vision to design simple experiments, while with less sophistication the experiments in question appear so complex that the researcher avoids doing them or gets entangled in troubles. Similar perceptual invariance assumptions can yield important insights into the mathematical constraints with which the visual system has to cope when viewing solid objects both with eye movements (Hadani *et al.*, 1978, 1980) and with object movement (Ullman, 1977, 1979). Above I stated that stereopsis (alone) yields only relative depth. Absolute depth can, however, be perceived by humans as a result of internal feedback from the oculomotor system together with the visual system devoted to stereopsis and to monocular depth cues, an interesting problem beyond the scope of this review.

A much ignored although interesting theory that tries to link global psychophysics to local neural mechanisms is the Lie transformation group model of neurophysiology, first proposed by W. C. Hoffman (1970) and developed by the psychologist P.C. Dodwell (1983). This theory is based on the observation that during the act of seeing the visual cortex annuls, to a great extent, the effects of affine transformations, phenomena called size and shape constancies in visual perception. Indeed, as we bring our outstretched arm closer to our eyes, the perceived size will not appear to change much, in spite of large changes in retinal size. Solid objects in rotation will appear unchanged in shape (within limits, since faces are poorly recognized upside-down). Circular disks at oblique viewing angles do not appear elliptical, but continue to be perceived as circular, and so on. Hoffman's theory is based on the mathematical theory of continuous groups (using Lie algebras, created by Sophus Lie in the last century) to account for the perceptual constancies. He proposed that the visual cortex has built-in local Lie differential operators, called Lie germs. For details and criticisms the reader is referred to a volume of papers edited by Paillard (1977). Whether the human brain uses Lie germs is a controversial topic. We have tried successfully on the Connection Machine (a massive parallel computer) an algorithm incorporating Lie germs (in each of the 64k parallel subunits) of several spatiotemporal affine transformations (Papathomas and Julesz 1989). Here is a typical case in which machine vision might use tools that solve the problem of perceptual invariances in space and time, but the tools could very well differ from those utilized by brains.

F. On "maturational windows" and stereo blindness

The topic of stereopsis leads us to the concept of "maturational windows," one of the most important discoveries in developmental psychology. It seems that the genetic code is inadequate to carry all the information a highly developed organism needs, including the fine-tuning to a particular niche of the environment. The best example is language acquisition. While the "deep structure" for learning a language is provided at birth—whether one is English or Chinese—the actual learning has to take place within a rather narrow "maturational window" in time. The missing of this window has dramatic influences, as attested by the many stories of "wolf children" who could not be fully rehabilitated to learn, that is to understand and utter, speech after a "critical period."

One assumes that the reason why we are not born with the syntax and semantics of a given language is the huge amount of information such a hereditary transfer might necessitate. Therefore it is important that even such a low-level process as detecting the orientations of contours can be dramatically influenced by environmental manipulations within a specific critical period in the cat and monkey. For instance, Blakemore and Cooper (1970) and Hirsh and Spinelli (1970) restricted the visual input of young kittens to stripes of just one orientation. In contrast to normal kittens (for whom all orientations elicit neural responses), these kittens that underwent "environmental surgery" had functioning neurons mainly for the orientation they were exposed to during their critical period. For details on the effects of early visual experience, see Mitchell (1980) and the excellent textbook edited by Barlow and Mollon (1982). Humans with strong astigmatism that has not been corrected in infancy will have acuity defects at their original orientational defect even after their astigmatism is corrected by lenses (see Mitchell, 1980). Experiments of this kind cast new light on the centuries-long debate concerning heredity and environment or "nature and nurture."

Stereo blindness, which afflicts about 2% of the human population (Julesz, 1971), often has its origin in strabismus (cross-eyedness), which prevents binocular-disparity-tuned neurons from receiving correct stimulation. This incorrect stimulation leads to a condition of "*amblyopia ex anopsia*" ("lazy eye"), which means that one eye becomes dominant, while the other becomes practically nonoperative. After surgical intervention to correct strabism, the child usually remains stereo blind, so the surgery is merely a cosmetic intervention. The reason why surgery does not, in most cases, restore stereopsis is that it is done too late, after the critical period during which stereopsis becomes functional. In order to determine the critical period for stereopsis in human infants, two steps have been taken.

First, the technique of random-dot stereograms was extended to a new class of stimuli called random-dot correlograms (Julesz and Tyler, 1976), which are quite different from random-dot stereograms. A dynamic random-dot correlogram alternates between binocularly correlated and uncorrelated visual noise. A special case of uncorrelation is negative correlation (where the left and right images are complements of each other). The advantage of these stimuli is that, regardless of the position of the observer's head, the stimuli are either correlated (stimulating binocular neurons) or are not correlated. Furthermore, humans can detect changes from binocular correlation (order) into binocular uncorrelation (disorder) in as brief as 2 msec, while the reverse, detecting the change from disorder into order takes 20 msec. We named this entropylike behavior "neural entropy" or "neurontropy" (Julesz and Tyler, 1976). The second step was the use of dynamic random-dot stereograms and correlograms to evoke visual potentials on the human skull (Lehmann and Julesz, 1978; Julesz *et al.*, 1980). Dynamic random-dot correlograms are a robust method, which was used by Miezin *et al.* (1981) to evoke visual potentials in the monkey. More importantly, dynamic RDS and correlograms were used to determine the onset of functional stereopsis in human infants by Braddick *et al.* (1980) and Petrig *et al.* (1981). These studies reveal that evoked potentials appear first in normal human infants at 3.5 months after birth. Similar results were obtained using RDS but with behavioral techniques by Fox *et al.* (1980). In adults, the evoked potentials to a dynamic RDS portraying a cyclopean checkerboard that oscillates in depth at a rate of f Hz appear as a square wave with a period of f Hz, while a similar dynamic correlogram elicits a square wave at a frequency of $2f$ Hz (Julesz *et al.*, 1980). This doubling of frequency appears in human infants at about 6 months, as shown by Petrig (1980). It seems that functional binocularity appears at 3.5 months, while some more sophisticated binocular process (perhaps stereopsis) emerges at 6 months in human infants. For a discussion of electrical activities in human brains, including evoked potentials in infants, I recommend the thorough and informative monograph by Regan (1989). Gian Poggio (personal communication) attributes the frequency doubling for correlograms to the existence of two classes of binocular neurons, which fire either for correlation or uncorrelation, respectively.

Stereo deficiencies, including stereo blindness, are not serious handicaps, but there are several professions for which stereopsis is a must. For instance, fluoroscopists routinely perform the catheterization of the heart while looking at 3D screens of x-ray machines, they have for decades used the unfakeable RDS tests to weed out inappropriate applicants for this specialty. For astronauts, as for quality inspectors (who inspect VLSI chips under stereo microscopes for imperfections), good stereopsis is a job-related requirement. The combination of dynamic RDS with evoked potentials now permits a fast determination of functional binocularity in human infants. It is hoped that such an early diagnosis will lead to early surgery to correct for strabismus before the maturational window closes, which in turn will result, in many cases, in the restoration of stereopsis.

IV. TEXTURE DISCRIMINATION AND FOCAL ATTENTION

A. A brief note on motion perception

The advances in global stereopsis made possible by RDS were paralleled by advances in motion perception using random-dot cinematograms (RDC). Since motion phenomena require demonstrations for the nonexpert that do not lend themselves to the printed page, I shall skip them here and turn to human texture discrimination. The reader interested in motion perception can find excellent reviews by Anstis (1978), Van de Grind *et al.*, (1983), Hildreth (1984), Adelson and Bergen (1985), Nakayama (1985), Regan and Beverley (1985), Watson and Ahumada (1985), and Cavanagh and Mather (1989), among others. Since this review is intended for physicists, I draw special attention to the pioneering work on motion perception in the fly by the physicist Werner

Reichardt (1961) and co-workers (Reichardt and Egelhaaf, 1988). He found that correlationlike mechanisms are used in motion perception in the fly's visual system; it turns out that similar Reichardt detectors are used in the human visual system too (van Santen and Sperling, 1985). RDC exhibit the same false-target problem as RDS (see Fig. 7) except that, for the latter, one has to search only horizontally, whereas for the former, disparity is a vector and the search has to be carried out in two dimensions. Therefore RDC present a more complex problem. Reichardt and Egelhaaf (1988) pointed out mathematically that the ambiguity of motion in an aperture (called "the aperture problem"), contrary to common belief, can be locally solved by two correlation-type motion detectors. Similarly to stereopsis, motion perception exhibits cooperative phenomena (particularly hysteresis), as shown by Chang and Julesz (1984) and Williams *et al.* (1986).

With RDS several motion phenomena can be produced that differ from classically portrayed apparent (stroboscopic) motion, thus demonstrating that binocular-disparity-tuned neurons of motion have different time constants from monocular neurons of motion (Julesz and Payne, 1968).

B. Preattentive texture discrimination and focal attention

I turn now to preattentive texture discrimination (also called texture segmentation), a problem I posed almost simultaneously with the RDS paradigm (Julesz, 1962). We discussed how global stereopsis facilitates object segmentation based on binocular disparity differences. For objects in the same depth planes, segmentation can be aided by differences in surface textures. We restrict ourselves to textures in two dimensions (i.e., the monocular retinal projections of real-life textures in relief, such as bark on a tree, grass or plowed fields viewed from a distance, textiles, microscopic slices of biological tissues, etc., and artificial 2D arrays of repeated or randomly scattered elements such as wallpapers). Effortless or preattentive texture discrimination (segmentation) was illustrated by Fig. 2. There we discussed the difference between preattentive pop-out of certain textures from among others by a parallel neural mechanism (which is independent of the number of texture elements) versus serial element-by-element scrutiny by the "searchlight" of focal attention.

This dichotomy of preattentive-attentive processing in visual perception is very helpful in explaining texture discrimination. As we have shown (Sagi and Julesz, 1987), texture gradients pop out only if the element density is above a critical value. Even then, one can determine only the locations of the gradients in a brief flash (e.g., *where* are the horizontal or vertical line segments embedded in a diagonal array of line segments); to identify them (i.e., *what* they are: horizontal or vertical) requires scrutiny by focal attention, which depends on the number of texture gradients (Sagi and Julesz, 1985). We also determined the aperture of focal attention ("searchlight"), which increases with eccentricity (Sagi and Julesz, 1986). In the Introduction we discussed how the searchlight of attention scans rapidly, without eye movements, at 30–60 msec/item depending on the visibility of the texture gradients, and how some parallel mechanism seems to facilitate serial search (Kröse and Julesz, 1989; Wolfe and Cave, 1990). Particularly interesting are the findings of the cognitive psychologist Anne Treisman and co-workers (Treisman and Gelade, 1980; Treisman and Paterson, 1984), who postulated a "feature integration theory" of vision. According to this theory, "disjunctions" of features pop out, but "conjunctions" of features require serial search. For example, red horizontal line segments pop out from green ones, but in a mixture of red horizontal and green vertical line segments to find, say, red vertical line segments requires serial search (which depends on the number of elements). The psychological finding that color and line orientation are not coded simultaneously is quite unexpected from classical neurophysiological evidence [although recently Livingstone and Hubel (1987) found nonoriented blobs tuned to color surrounded by the usual orientation-tuned neurons that were not sensitive to color in area V1 of the monkey cortex]. Interestingly, if a third feature, such as depth or motion, is added to color and orientation, this "conjunction" of three features pops out preattentively (Nakayama and Silverman, 1986).

Treisman's theory, though superficially resembling my "texton theory" of preattentive vision, differs from it essentially. While Triesman does not specify here stimuli and often uses stimuli whose elements are distant from each other, my texton theory can only be applied for textures, which in turn requires that the elements be dense. (For critical distances within texture elements and critical distances between texture elements see Julesz, 1986b. For other work on texture density see Nothdurft, 1985, and Sagi, 1989.) I am skeptical that any theory can be postulated for distant elements at present, because they are influenced by Gestalt organizations in unknown ways, while the perception of textures is much simpler.

C. Human texture segregation: when the whole is less than the sum of its parts

Before I discuss the recent developments in the segmentation of textures by spatial filters, a brief survey on the evolution of the texton theory might be of interest to the physicist. In 1962 I realized that effortless texture discrimination might be a much less complex problem than the perception of form (Julesz, 1962), though it exhibits challenging properties. When the pair X and L is briefly presented, its discrimination is somewhat stronger than that of the pair L and T, but not by much, so it is not obvious why aggregates of X's and L's do pop out while aggregates of L's and T's do not. Texture elements that segregate in isolation and remain segregated when they are pooled to form textures comprise a far less interesting case than elements that become harder to discriminate when pooled. Indeed, it is not obvious how aggregates of texture elements that segregate in isolation become less and less discriminable as the number of elements in the aggregates increases.

Clearly, it is not interesting when a pair of elements in isolation is indistinguishable and remains so in aggregates. It is more interesting when a pair that is distinguishable in isolation becomes more and more discriminable as the number of elements increases in the two aggregates. For example, Caelli and Julesz (1979) took textures composed of dipoles (i.e., two nearby dots) as elements; dipoles in one texture had orientations in $\varnothing 1$ range, while in the other range of $\varnothing 2$. We showed both theoretically and experimentally that if $\Delta = \varnothing 2 - \varnothing 1$, and N is the number of dipoles, the psychophysical function at the visibility threshold of texture discrimination is

$$\log N = \log\Delta - \log(c-\Delta) + k$$

where c and k are experimental constants. Thus texture discrimination is a global function that increases monotonically with the number of texture elements and with the difference between the two dipoles' orientation range. It would be nice to develop a model that would predict why and how, say, discrimination between one T and L decreases with increasing numbers of elements, a problem to be discussed next.

There are several possible reasons for elements that are strongly discriminable in isolation to become less and less discriminable as the number of elements in the two aggregates increases. As of now I have identified three, based on the type of texture pair. One is shown in Fig. 11. Here an S- and a 10-shaped element can be easily discriminated from each other [Fig. 11(a)], but the texture pair formed by their aggregates is indistinguishable [Fig. 11(b)]. The reason for this phenomenon is probably lateral inhibition between adjacent elements. We have to shift our small disk of focal attention to remove this lateral inhibition, thus enabling scrutiny. Furthermore, subjective contours close gaps (Williams and Julesz, 1991), to be discussed later.

The second kind of indistinguishable texture pair is between, say, an R and its "mirror image dual," as shown in Figs. 12(a) and 12(b). In isolation the R and its mirrored dual can be discriminated in Fig. 12(a); however, in the laboratory this discrimination requires time-consuming "mental rotation" (Shepard and Metzler, 1971), which depends on the amount of angular difference between the pair. It appears that texture pairs like this [Fig. 12(b)] cannot be rotated all at once, only element by element. Indeed, one can discriminate the elements of the texture pair only by inspecting each element one at a time and trying to decide whether it is an R or its mirror image.

A third kind of indistinguishable texture pair is depicted by Fig. 13, where in isolation the T-shaped four-disk pair seems very different from the rectangle-shaped four-disk pair. Yet their aggregates yield indistinguishable texture pairs (Julesz, 1975). This texture pair is generated by the "four-disk method," which yields textures with identical second-order statistics [hence with identical autocorrelations and Fourier power spectra (Julesz *et al.*, 1978)]. The reason for our not being able to discriminate between these texture pairs appears to be the many irrelevant shapes that are formed by disks at the boundaries between adjacent texture elements, which in turn mask the dual elements.

These demonstrations already show that the "law of superposition" does not apply for preattentive texture segmentation (since aggregates of discriminable element

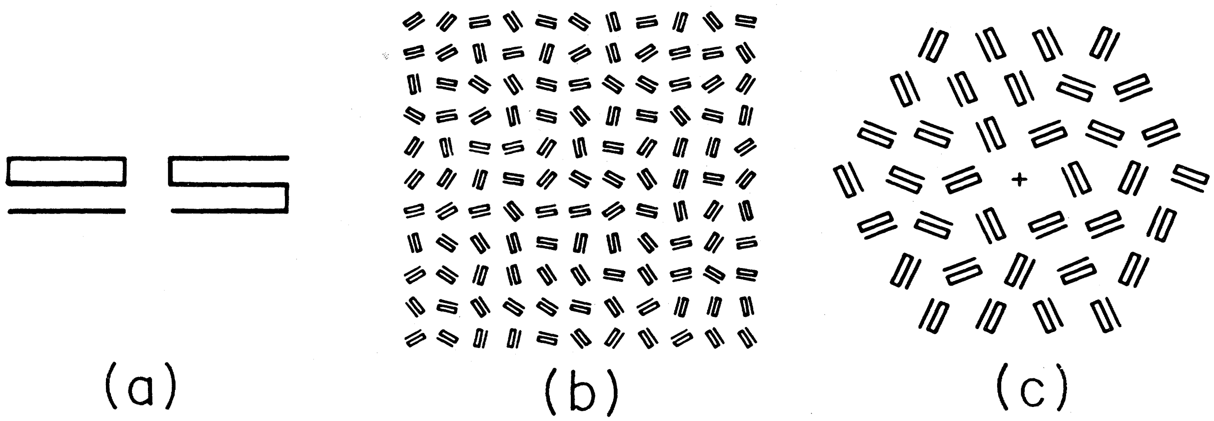


FIG. 11. Demonstration of a texture pair composed of S- and 10-shaped elements, which are discriminable in isolation, as shown in (a), but indistinguishable in an array, as shown in (b). (c) is similar to (b), except that only one target is presented, and to detect this target element-by-element scrutiny is required. From Julesz (1981).

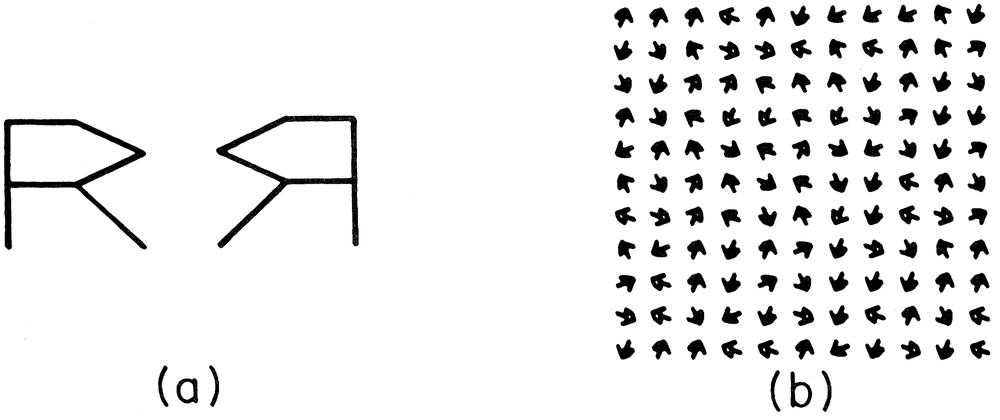


FIG. 12. Iso-second-order texture pair, composed of *R*'s and their mirror image duals; in isolation (a) the elements are discriminable, while in an array (b) they appear indistinguishable. From Julesz (1981).

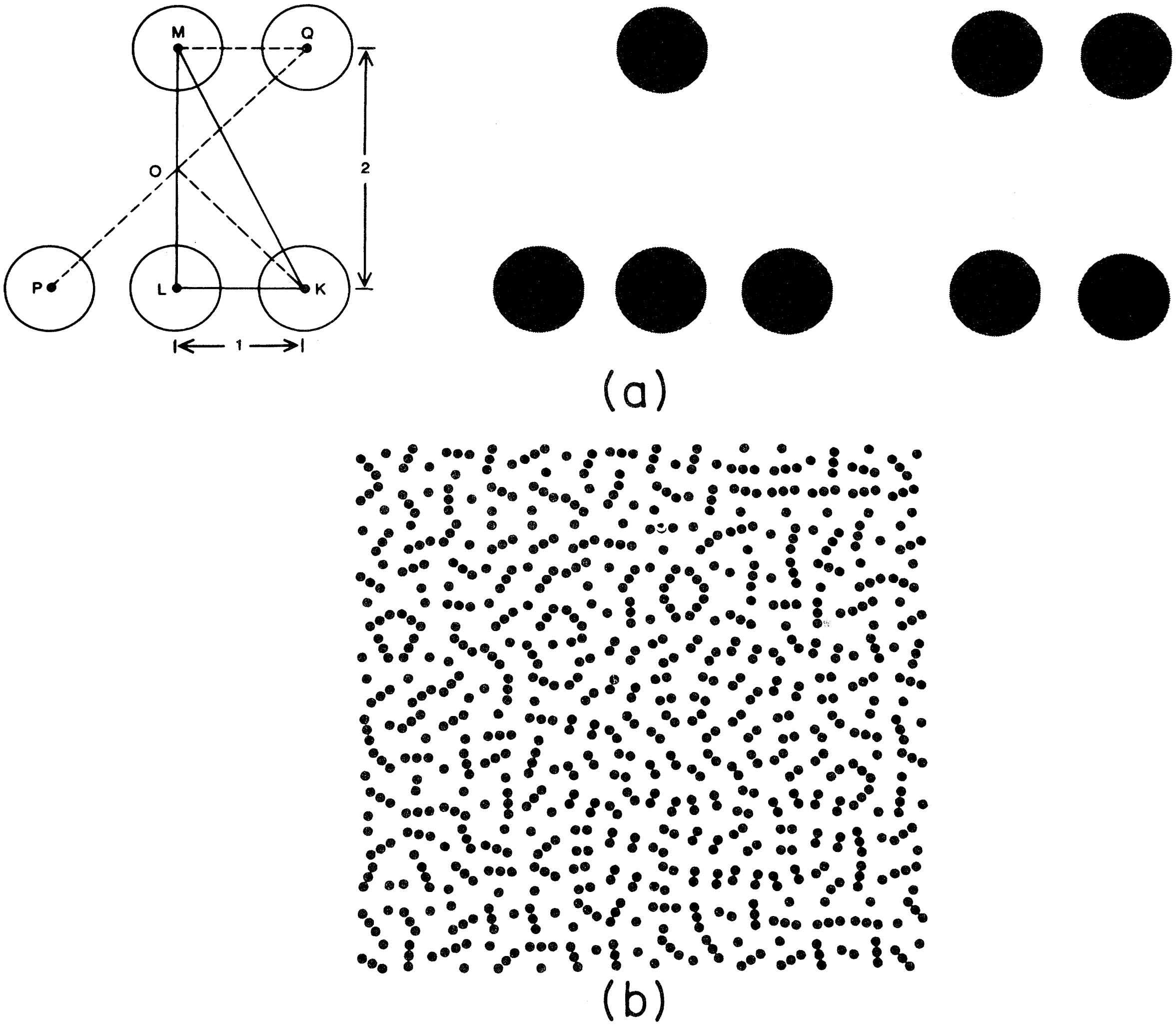


FIG. 13. Iso-second-order texture pair generated by the "four-disk method". From Julesz (1975) illustrated in (a). The pair becomes indistinguishable in (b).

pairs become indiscriminable), and therefore no linear spatial filters can model human texture discrimination. This essential nonlinearity of human texture discrimination will be demonstrated later in Fig. 17. Before I review recent work with nonlinear spatial filters, I briefly discuss theoretical studies aimed at generating stochastic textures with identical second- and third-order statistics, which led to the concept of *textons*—the basic perceptual units (perceptual quarks?) of preattentive texture discrimination.

V. FROM TEXTONS TO NONLINEAR SPATIAL FILTERS

A. A brief outline of the texton theory of texture discrimination

In 1962 I asked a combined mathematical and psychological question that has kept many mathematicians and psychologists busy ever since (Julesz, 1962). Because I knew that texture pairs that differed in their first-order statistics would be effortlessly segregated (based on differences in tonal quality) and assumed that differences in second-order statistics could be distinguished from each other (based on differences in granularity), I wanted to study textures with identical Nth-order statistics but different $(N+1)$th-order statistics. [Here I define Nth-order statistics as the probability that the vertices of an "N-gon" (e.g., a hexagon, pentagon, etc.) thrown randomly on a texture fall on certain N colors.] I wanted to determine the highest N that still yielded texture segmentation and wanted to know what perceptual quality would accompany such discrimination. For example, would texture pairs with identical second-order statistics (hence identical first-order statistics and identical Fourier power spectra) be discriminable, and what would the perceptual difference be? Surprisingly, at that time mathematicians did not know how to create such constrained stochastic textures, but from 1962 to 1975, Slepian, Rosenblatt, Gilbert, Shepp, and Frisch were instrumental in creating iso-second-order random texture pairs whose elements in isolation appear conspicuously different, yet as textures cannot be told apart. The indiscriminable texture pairs depicted in Figs. 12 and 13 were obtained by these efforts. It seemed that iso-second-order textures were so severely constrained globally that the visual system could not tell them apart. However, in 1977 and 1978, colleagues T. Caelli, E. Gilbert, and J. Victor helped me to invent stochastic texture pairs with *global* constraints of identical second-order (and even identical third-order) statistics that yielded preattentive texture discrimination based on some *local* conspicuous features, which I later called textons. Luckily, now that we know what textons are and their role in vision has been clarified, the reader need not take the tortuous mathematical path that led to their discovery (for details see Caelli, Julesz, and Gilbert, 1978; Julesz *et al.*, 1978; Julesz, 1981, 1984). [It can be mathematically proven that the four-disk method of Caelli, Julesz, and Gilbert (1978) is the only one using identical disks that can generate iso-second-order texture pairs in the Euclidian plane, thus permitting a thorough search for dual elements whose aggregates might pop out.] Figure 14(a) demonstrates the first iso-second-order discriminable texture pair we found, using the four-disk method. This figure, together with 14(b), and 14(c), which were generated with the help of the "generalized four-disk method" (in which the disks are replaced one by one by specific symmetric shapes), depicts iso-second-order texture pairs that are preattentively discriminable. Discrimination is based on local features which we named *"quasicollinearity," "corner,"* and *"closure"* (Caelli, Julesz, and Gilbert, 1978). Figure 14(d) shows iso-third-order textures (Julesz, Gilbert, and Victor, 1978) with the property that any triangle thrown on these textures has the same probability of its vertices falling on the same colors (however, the vertices of probing 4-gons will have different probabilities). As the reader can verify, discrimination is effortless and is obviously not due to computing differences in fourth-order statistics, but rather to elongated blobs of different aspect ratios and orientations.

What these textons really are is hard to define. For instance, in Fig. 14(a), besides quasicollinearity there are also more white gaps between these elements, giving rise to antitextons. As I pointed out (Julesz, 1986) it is not only the black (white) textons whose gradients yield texture discrimination but also the white (black) spaces between them, which act as textons too.

In essence, we found that texture segmentation is not governed by global (statistical) rules, but rather depends on local, nonlinear features (textons), such as color, orientation, flicker, motion, depth, elongated blobs, and collinearity, to name the most conspicuous ones that are both psychophysically and neurophysiologically accepted as being fundamental. Some less clearly defined textons are related to ends of lines or terminators, which occur in the concepts of "corner" and "closure" and are hard to define for halftone blobs. Particularly important is the realization that—contrary to common belief—texture segmentation cannot be explained by differences in power spectra. On the other hand, it became obvious that instead of searching for higher-order statistical descriptors, the visual system applies some local spatial filtering followed by some nonlinearity, and the results must be averaged again by the next spatial filter stage. This is depicted in Fig. 15, which illustrates how a Kuffler-type unit (instead of a Mexican-hat-function profile, a simpler spatial filter of 2×2 pixel center addition with a 2-pixel-wide surround annulus of subtraction, as shown in the inset) acts on the iso-third-order texture pair of Fig. 14(d), followed by a threshold-taking device (Julesz and Bergen, 1983). When viewing the output of this nonlinear spatial filter in Fig. 15, our visual system performs a second spatial filtering by separating the two areas of different lumi-

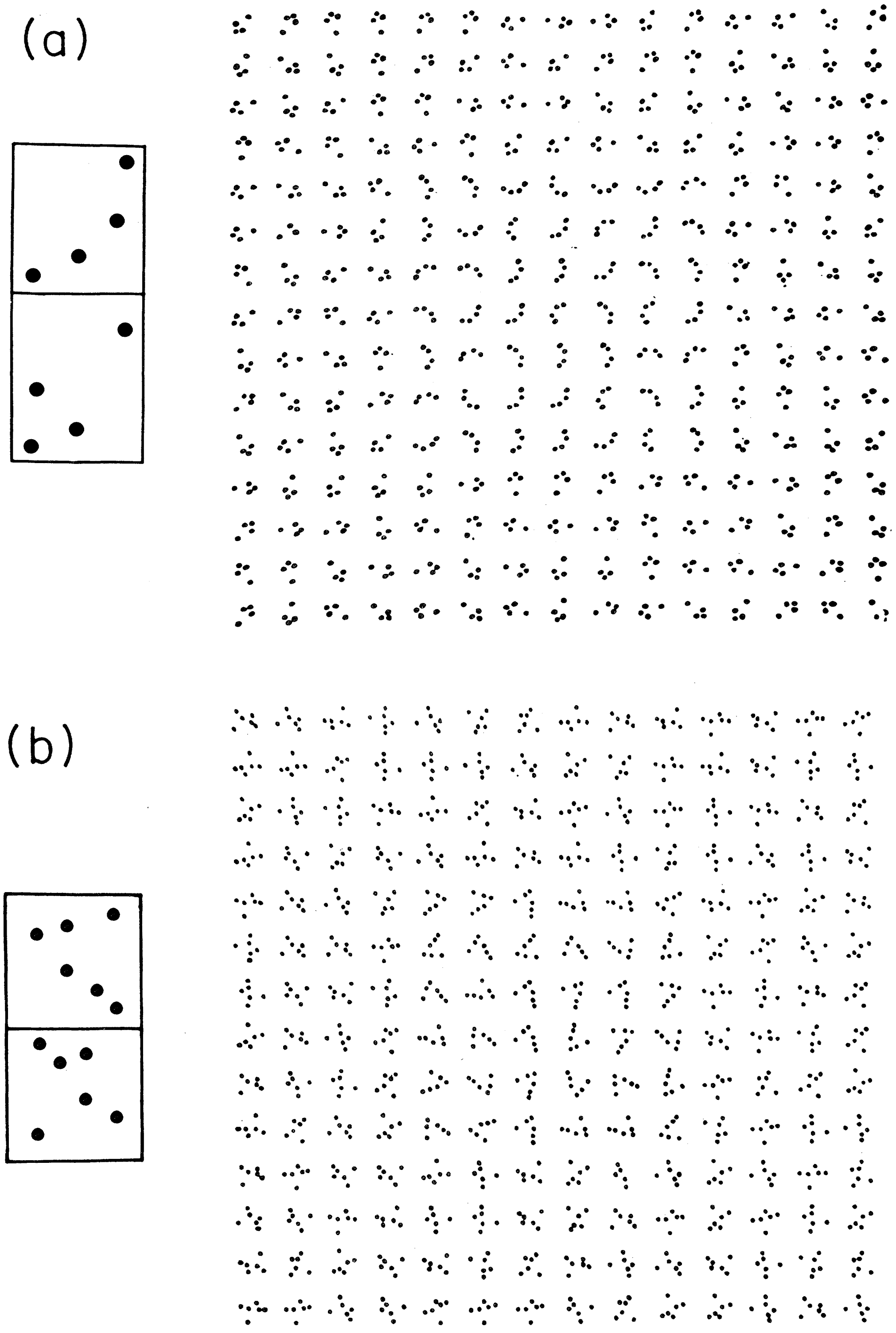


FIG. 14. Preattentively discriminable iso-second-order and iso-third-order texture pairs. (a) Iso-second-order texture pair that is indiscriminable due to the local conspicuous feature (texton) of "quasicollinearity." (b) Iso-second-order texture pair that is discriminable due to the local conspicuous feature (texton) of "corner." (c) Iso-second-order texture pair that is discriminable due to the local conspicuous feature (texton) of "closure." From Caelli, Julesz, and Gilbert (1978). (d) Iso-third-order texture pair that is discriminable due to the local conspicuous feature (texton) of "elongated blobs of specific orientation, width, and length." From Julesz, Gilbert, and Victor (1978).

nance distributions (that were obtained by the threshold taking).

One problem with such nonlinear simple filters is their inability to account for the asymmetry problem of human texture segmentation. It is known (Julesz, 1981; Gurnsey and Browse, 1987; Treisman and Gormican, 1988) that very often a given texture *A* pops out more strongly from a background of *B* than *B* does from a background of *A*,

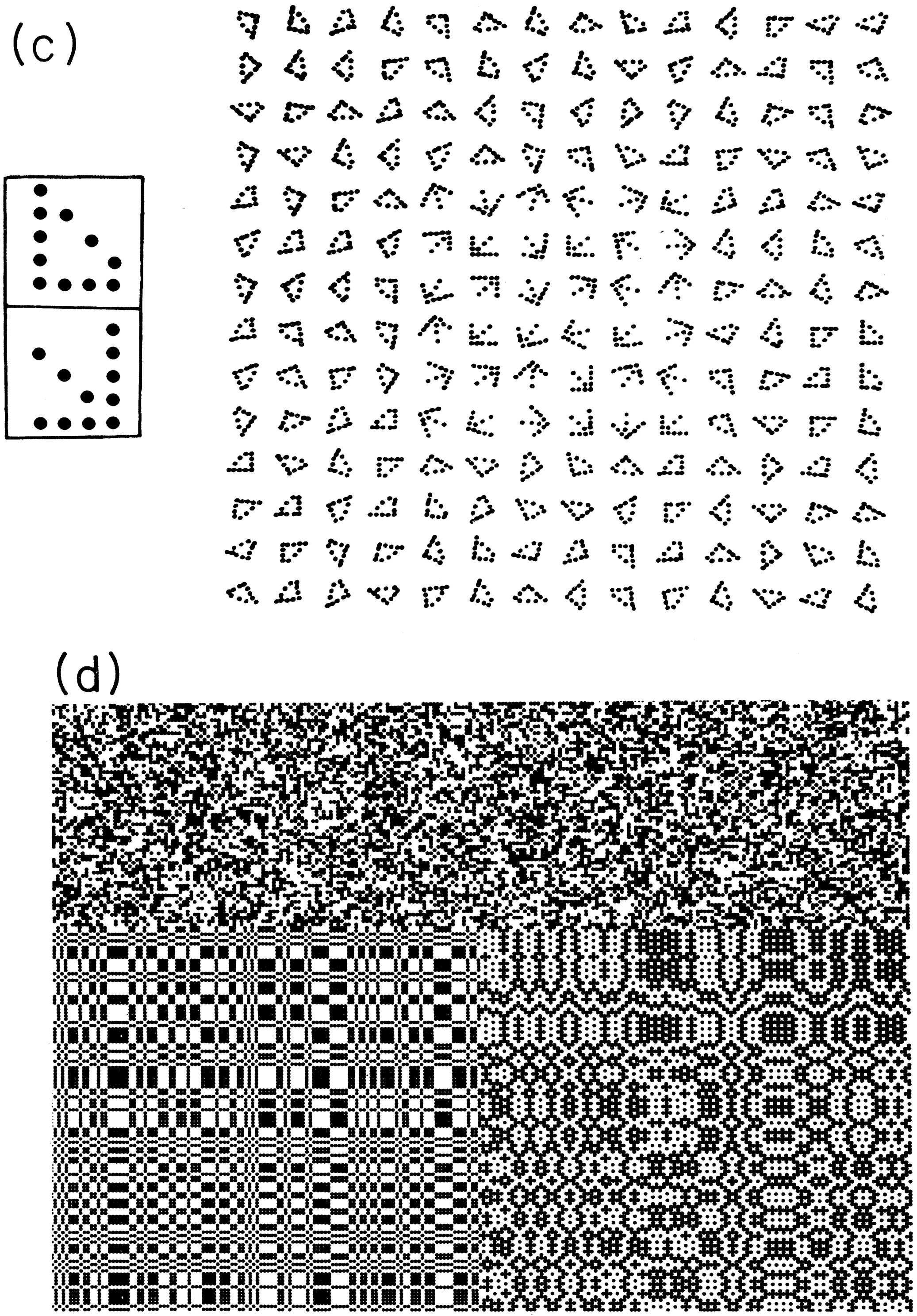


FIG. 14. (*Continued*).

as shown in Fig. 16(a). For many years I was worried that the asymmetry problem of preattentive texture discrimination might depend on top-down processes and complex figure-ground phenomena. Therefore I am glad to report that the asymmetry effect in Fig. 16 can be explained (Williams and Julesz, 1990, 1991) by assuming that the nonlinear operation is the subjective contour phenomenon that "closes the gaps." When we discussed Fig. 6, I pointed out that subjective (also called illusory) contours are extracted in V2 and therefore belong to early visual processes. Figure 16(b) yields a similar asymmetry of texture discrimination to that of Fig. 16(a), even though the orientation of the gaps is not jittered. We shall return to this demonstration in the next section.

B. The asymmetry problem of texture segmentation and nonlinear spatial filters

In 1988, several texture segmentation algorithms were developed based on the use of linear spatial filters followed by squaring or some other nonlinear operation (Bergen and Adelson, 1988; Voorhees and Poggio, 1988). (For a definition of spatial filters, e.g., the Laplacian of a Gaussian, or Gabor filters, see Sec. II.B.) For a critique showing that linear spatial filters cannot segment textures, see Julesz and Kröse (1988) and Julesz (1990c). Recently Williams and Julesz (1991) showed the nonlinear behavior of human texture discrimination as depicted in Fig. 17. Here the nondiscriminable iso-second-order texture pair, invented by Caelli, Julesz, and Gilbert (1978), is shown on the right side. (This texture pair is one of the rare cases that have iso-second-order statistics without having to rotate the texture elements randomly.) We were able to decompose this texture pair into the sum of a highly discriminable texture pair and a nondiscriminable texture pair, as shown on the left side. The fact that a discriminable texture pair becomes nondiscriminable when a nondiscriminable texture pair is linearly added shows convincingly the violation of the law of superposition for texture discrimination.

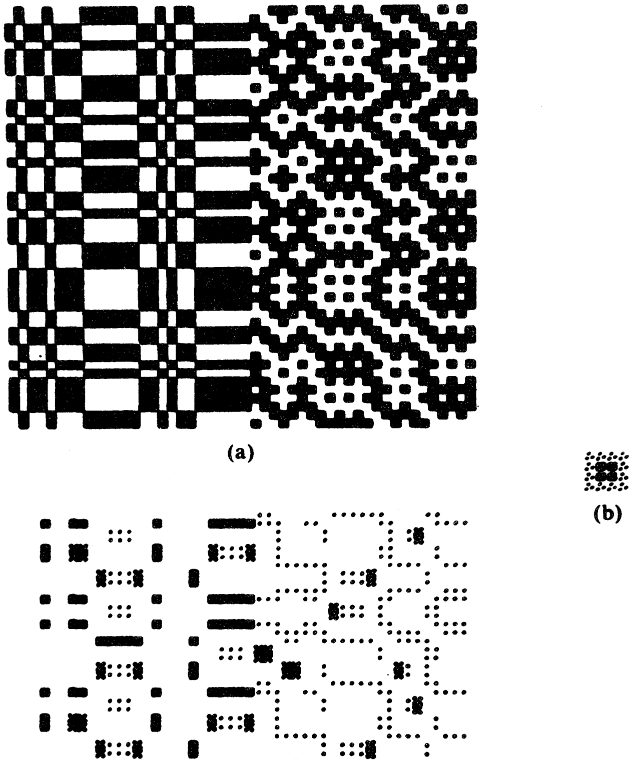


FIG. 15. Demonstration of how a simple local linear filter followed by a nonlinearity (threshold-taking) can segment the iso-third-order texture pair of Fig. 14(d). From Julesz and Bergen (1983).

More recently Fogel and Sagi (1989) and, independently, Malik and Perona (1990) developed texture segmentation algorithms based on the use of local spatial filters (oriented Gabor filters) followed by a quasilocal nonlinear operation (simple squaring in Fogel and Sagi's version and some inhibition between neighboring elements in the Malik and Perona algorithm), with a second spatial filter for final segmentation. It was most impressive that this approach emulated human texture discrimination performance as measured by Kröse (1987), but still could not account for the asymmetry effects. Therefore it is of great significance that Rubenstein and Sagi (1990) extended their model by determining the variances of the local texture elements' distributions after the nonlinear stage and found these variances asymmetric, particularly when the orientation of the elements was jittered, mimicking human performance. Their model could account for the textural asymmetries reported by Gurnsey and Browse (1987) and probably will be able to handle some other asymmetries of the kind shown in Figs. 18(a) and 18(b). In Fig. 18 a typical input-output pattern of the Rubenstein and Sagi (1990) algorithm is presented as it segments a texture pair (A in B) and its dual (B in A). It is most heartening that even the textural asymmetry effects that seemed to be based on figure-ground reversals—which in turn depended on unknown top-down processes—can be successfully explained by bottom-up processes modeled by relatively simple nonlinear spatial filters.

The Rubenstein and Sagi (1990) model can account for the asymmetry problem by assuming that jitter of line orientation accounts for increase in variance of their filter's output, hence increase in texture discrimination asymmetry. However, the demonstration of Figs. 16(a) and 16(b) clearly shows that in general the asymmetry problem of texture discrimination does not depend on orientational jitter. Indeed, recently, Williams and Julesz (1991) extended the texton theory to include illusory contours and "fill-in" phenomena between gaps and nearby elements, which could be regarded as *antitextons*. The filling in of the gaps by subjective contours can account for the asymmetry effect shown in Figs. 17(a) and 17(b), and the fill-in phenomenon between texture elements can explain many other asymmetries.

Antitextons together with the textons extend the theory of trichromacy, the only real scientific theory in psychology. The theory of trichromacy states that any color can be matched to a combination of three basic colors, red, green, and blue, such that the boundary between the selected color and the combination colors becomes minimum (or disappears without scrutiny). (As a matter of fact this theory, postulated by George Palmer in 1777, can be regarded as the *first scientific atom theory*, years before Dalton introduced atoms into chemistry.)

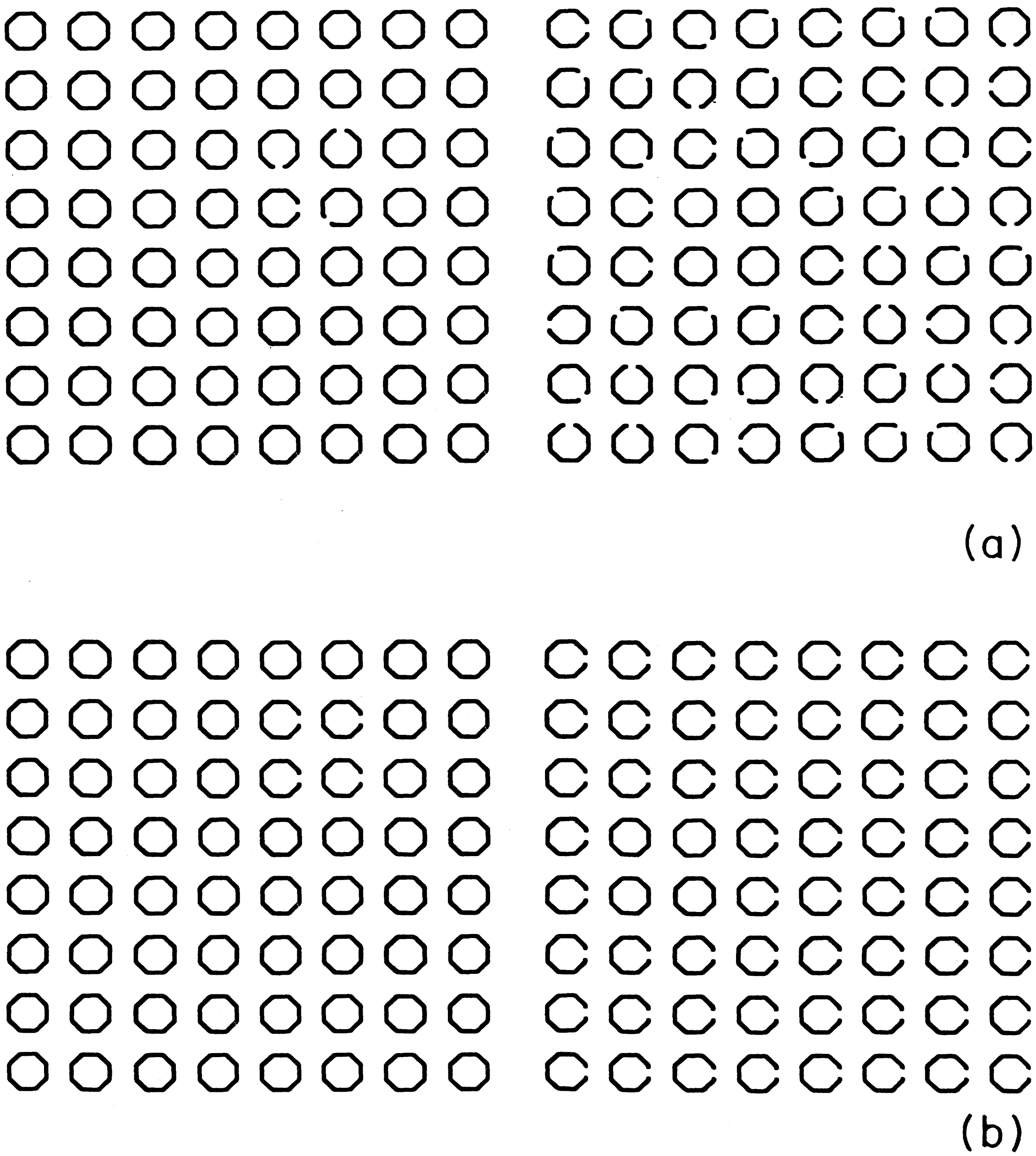


FIG. 16. Asymmetry of texture discrimination. (a) The perception of gapped octagons among closed octagons yields weaker discrimination than vice versa. (b) Similar to (a) but the position of the gaps is not jittered. This does not reduce the asymmetry effect. From Williams and Julesz (1991, in press).

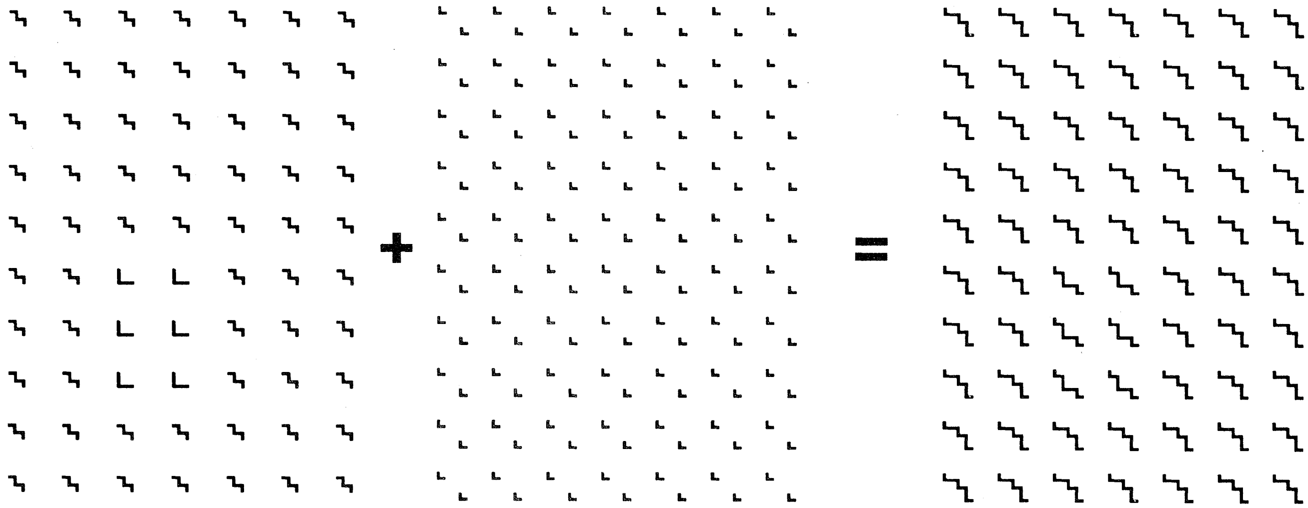


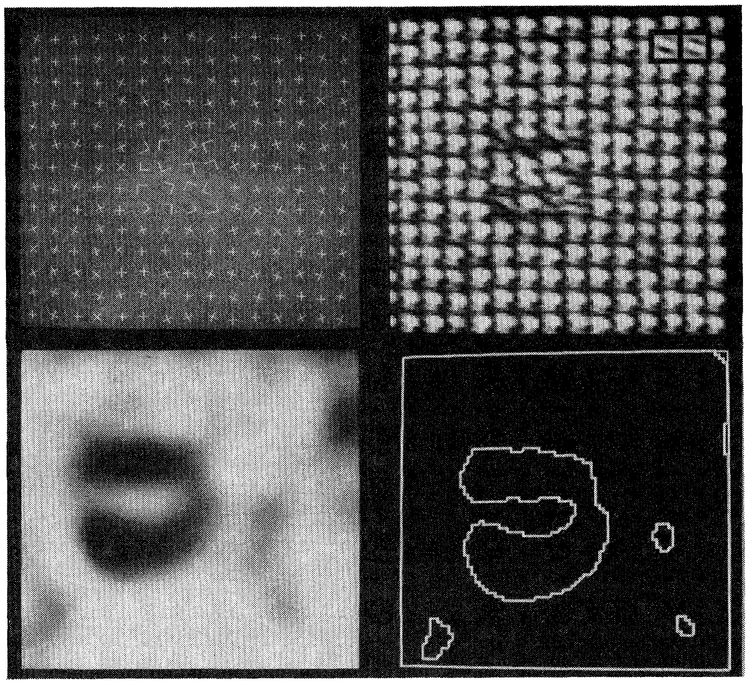
FIG. 17. Demonstration of the nonlinearity of human texture discrimination. Adding a nondiscriminable texture pair to a highly discriminable texture pair renders the latter nondiscriminable, thus violating the law of superposition. From Williams and Julesz (1991, in press).

When I introduced textons into psychology I wished to extend trichromacy to encompass textures as well as colors. I wanted to know whether any texture could be matched to a finite (and not too large) number of textons such that the boundary between any textural array and an array containing a mixture of textons would perceptually disappear without scrutiny. It seems now that this can be achieved. The fact that the gamut of colors can be matched by just three colors is in itself amazing. The finding that the infinitely richer variety of 2D textures could be matched to a mixture of a finite number of textons is even more unexpected!

In Fig. 17 it was shown that the law of superposition does not apply for texture elements. Whether the combination of textons themselves is a linear or nonlinear operation is not yet known. Howard L. Resnikoff (1987a) in his interesting monograph *The Illusion of Reality,* devotes an entire chapter to an early version of my texton theory and argues for the linear superposition of textons. Whether linear superposition still holds for the new texton theory, incorporating antitextons, remains to be seen.

For the author, who spent much of his scientific career in search of the elusive texton, it is anticlimactic, yet most satisfying, to find that quasilocal spatial filters can extract texton gradients without having to specify complex concatenation rules between adjacent textons. (I have no doubt that in the near future such filters will mimic human preattentive texture discrimination by incorporating several perceptual operations from the formation of subjective contours to the filling-in of gaps.) The reader familiar with speech research will recognize the similarity between "phonemes" and "textons." While phonemes were never well specified, and complex computer algorithms are now used to cope with the many *ad hoc* rules at their various concatenations in order to segment speech, nevertheless, the rudely defined phonemes permitted the development of phonetic writing, one of the great discoveries of human civilization. Had the development of phonetic speech coincided millennia ago with the invention of supercomputers that could automatically segment speech and talk, the skill of writing might never have developed. Of course, the fact that our voice organs limit the number of phonemes to a few dozen contributed to their universal acceptance. Similarly, the main insight from the texton theory was that, of the infinite variety of 2D textures, only a *limited number* of textons have perceptual significance and are evaluated *quasilocally* in effortless texture discrimination. (I use the term "quasilocal" instead of "local," because line segments, closed loops, corners, etc., have some finite dimensions.) Even though superfast computers will soon perform automatic texture segmentation, practitioners of visual skills—painters, designers of instrument panels or advertisements, directors of movies or TV shows—can benefit from the texton theory by the enhanced ability it gives them to manipulate the viewer's eye. Indeed, some of the great artists have instinctly known how to create a strong texton gradient to capture attention or create a texton equilibrium for which time-consuming scrutiny is needed to discover the hidden images.

I end this section with a mention of the work of Enns (1986), who made up arrays of little 2D perspective cubes with targets (cubes) biased to be perceived in one kind of 3D depth organization amidst cubes biased in the dual


(a)

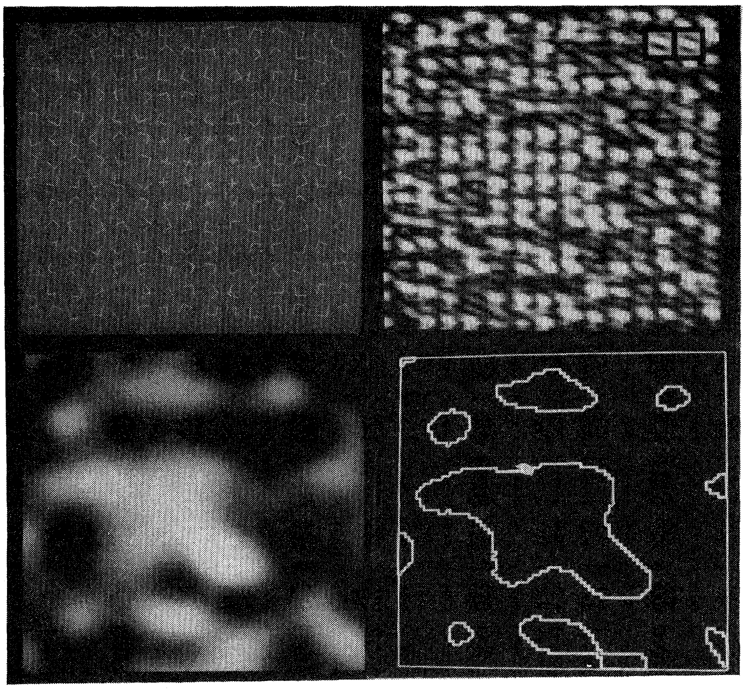

(b)

FIG. 18. Output of a nonlinear spatial filter that emulates the asymmetry phenomenon of human texture segmentation as illustrated by the dual texture pairs in (a) and (b). The first stage of Gabor filters is followed by a nonlinear operator, which in turn is followed by a second spatial filter. From Rubenstein and Sagi (1990).

3D organization. This depth from 3D perspective cues yields preattentive texture segmentation ("pop-out"), with the implication that, in addition to the textons of brightness, color, orientation, and aspect ratio of elongated blobs, flicker, motion, and stereopsis, even perceived depth in 2D perspective drawings might act as a texton. In my belief, some simple, quasilocal rules of 3D perspective, occlusion, transparency, etc., are probably utilized, without the need to invoke top-down processing.

C. Recent psychological and neurophysiological findings in texture discrimination

Perhaps the most important implication of the texton theory was its division of human vision into preattentive and attentive modes of action. Certain texture-gradient-like detections could be performed in parallel without scrutiny, while some other tasks that required identification needed serial search by attention. Recently, Braun and Sagi (1990) lent support to the "two-visual-system" concept by showing that while an observer's attention was loaded (by being asked to identify a letter) it was possible for the observer to carry out simultaneously the detection of texture gradients.

Other recent neurophysiological studies seem to support the texton-gradient notion of our perceptual studies. Van Essen *et al.* (1989) studied the responses of single units in visual areas V1, V2, and MT of the macaque monkey to stationary and moving patterns. In V1 and V2 the presence of a static texture surround (e.g., an array of parallel needles) somewhat decreases the response to a central texture element (single needle) within the classical receptive field if the orientation of the needles in the surround is perpendicular to that in the center. If they are parallel, the neural response is greatly reduced. Another neurophysiological finding is that of Robert Desimone and co-workers (Moran and Desimone, 1985; Desimone and Ungerleider, 1989), who located neurons in V4 whose firing for certain trigger features changes in accordance with the focal attention of the monkey.

D. Learning effects in early vision

One of the main themes throughout this review has been the phenomenological richness of early vision. Without cognitive and semantic cues, rather complex feats of false-target elimination can be performed in stereopsis and movement perception, asymmetry of texture discrimination takes place, subjective contours are formed, and so on. Even long-term memory effects occur in early vision. We have discussed some of the hysteresis effects that accompany the cooperative phenomena of global stereopsis and motion perception. Hysteresis is one of the simplest memory effects, in which some action modifies the outcome of a later response. For instance, Fender and Julesz (1967), using binocular retinal stabilization (by giving the subjects close-fitting contact lenses with mirrors attached), showed that a RDS had to be brought within Panum's fusional area (i.e., within 6 min.arc alignment) to obtain fusion. But after fusion, the left and right images could slowly be pulled apart by as much as 120 min. arc without breaking fusion. So, the fusional area depends strongly on the prior perceptual states. Another learning effect can be experienced when one first tries to fuse a RDS with large binocular disparities (Julesz, 1971 gives several demonstrations). At first, it might take minutes to achieve fusion, but even years later one can do it quite easily. This is not real perceptual (cortical) learning, but rather a procedural (cerebellar) learning. When fusing RDS with large disparities, the novice is trying large convergence movements to bring the corresponding areas of the RDS into Panum's fusional area, and it is this unconscious learning of proper vergence movements that is remembered years later (Julesz, 1986a). Some other real cortical learning phenomena of global stereopsis are also reviewed in the previous reference. Here I give only two examples of learning effects in preattentive texture discrimination:

First, when one presents some indistinguishable iso-second-order texture pairs that, however, are composed of element pairs with different convex hulls, after several hundred trials they can be effortlessly discriminated (Julesz, 1984).

Second, and even more interestingly, Karni and Sagi (1990) report a remarkable long-term learning effect in simple texture discrimination tasks were learning seems to be local in a retinotopic sense. What has been learned must be relearned for each different area of the visual field. Surprisingly, though learning is specific for target location, it is not specific for target orientation, but rather for background element orientation. These authors briefly presented, in an array of horizontal line segments, a few targets of adjacent line segments that were tilted from the horizontal orientation. With small tilts, it took several sessions to detect these targets correctly. This improvement was retained for the next sessions in the same retinal quadrant but was not transferable to other retinal quadrants. Changing the orientation of the targets (from left oblique to right oblique) had no effect on the learned performance. However, chaning the orientation of the background array (from horizontal to vertical) obliterated learning. These plasticity effects are of great interest, and it seems that perceptual learning in early vision might be a useful tool in understanding the mysteries of human memory.

I conclude this review with some perceptual phenomena that are not just bottom-up, but require semantic memory and other top-down processes for their explanation. I have emphasized throughout this review that most of the perceptual processes in global stereopsis, motion perception, and texture discrimination are essentially bottom-up; their linking to present neurophysiological results obtained in the early cortical areas of V1, V2, V3, V4, or MT is now possible. This is also in agreement with David Marr's view of computational vision's being basically bottom-up. However, in a recent short monograph *Visual Processing: Computational, Psychophysical, and Cognitive Research,* Roger Watt (1988) argues for algorithms in early vision that are under the control of high-level processes and memory. Indeed, in cognition there are many perceptual phenomena that depend on high-level processes, including semantic memory. A well-known example in cognition is the "word-superiority effect," in which the recognition of certain letters is superior when they are contained in an English word to when they are in a nonsense word or presented in isolation. This makes a good deal of sense, since recognition of letters and words is surely a high-level semantic process. However, as Naomi Weisstein and Charles Harris (1974) have shown, the same phenomenon exists in visual perception, where they call it the object-superiority effect. The detection of a line segment of certain orientation was greatly improved if the segment belonged to a line drawing that portrayed a 3D object; it deteriorated if the segment belonged to a random line drawing and was the worst if the line segment was shown in isolation.

Of course, object and form recognition are more complex, high-level processes in which semantics and Gestalt organization play a prominent role. Therefore the experiments of Gorea and Julesz (1990) are of special interest. We converted the object-superiority effect from an identification paradigm into a detection paradigm, as follows: We presented an array of oblique line segments into which three horizontal and one vertical line segments were inserted. These four nonoblique line segments were clumped either in random fashion or representing a primitive human face (two horizontal lines representing the eyes, the vertical line segment between the eyes representing the nose, and the bottom horizontal line segment portraying the mouth). Observers were not aware that occasionally a face was presented, and were only asked to detect any line segment that was not oblique. Surprisingly, observers detected the horizontal and vertical line segments significantly better when they belonged to the face than when they belonged to a random clump (or to a four-line-segment symmetric pattern) that was not a face. I have always assumed that the detection of a line segment in a texture (based on a texture (texton) gradient between adjacent orientation differences) was a simple parallel bottom-up process. And here is a case in which even such a simple perceptual task might depend on top-down processing! I say "might" because the effect is very small (though statistically significant) and only four observers were tried. Because of the importance of this experiment, I would welcome the attempt of others to repeat this study of ours with more observers and perhaps some other experimental design!

VI. CONCLUSION

A physicist reader who only glanced through this article might be surprised by the lack of explicit mathematical equations. Obviously I did not want to bore the reader with the difficult proofs of iso-*N*th-order texture generation, including their ergodicities. Furthermore, the internal structure of the cooperative computer algorithms modeling stereopsis or texture segmentation are much more complex in detailed mathematical notation than the usual differential equations of the Maxwell or Schrödinger kind. Furthermore, physicists have a knack of ignoring "dirty" problems, such as computing the shape of a puddle of spilt milk on a kitchen floor (a favorite example of George Sperling, 1978). [When they are forced to do so, to compute, say, the shape of a plasma in a magnetic bottle, they are confronted with the same difficulties as their colleagues in psychobiology.] Indeed, a "thought" might correspond to the "shape of a puddle" of cooperating neural pools of a certain activity.

I hope I have made it clear that progress in psychobiology is not hampered so much by the lack of mathematical tools, as by our inability to find the proper levels of complexity for describing mental phenomena. It is my opinion that whenever a canonical problem is raised in brain research, mathematical problems become simplified or vanish. Strategic insights might well be gained, not so much as the result of a new mathematical tool that will enable us to handle many-body nonlinear systems, but as the result of a novel way of looking at some emerging property at a proper level of complexity that we have missed so far! Indeed, in modern molecular biology much of the action can be understood at a phenomenological level as proteins fold and unfold, without having to compute the many-body problems of van der Waals forces. Even the genetic code has turned out to be a redundant three-letter code, instead of some complex error-correcting code envisaged by information theorists. This does not mean that some insights from theoretical physicists, particularly from experts in complex adaptive systems, might not be crucial for psychobiology.

Luckily in early vision we do not have to wait for such insights. The modular nature of early vision enables us to attack problems with adequate psychophysical, neurophysiological, and mathematical tools at our disposal. Obviously there are many interesting problems—from perceptual learning to the kind of information one can collect without focal attention—that will keep us busy for many years to come. The only unanswered question is whether such a reduced psychobiology of early vision is an adequately interesting subfield of human vision for intensive study. In my opinion, it is a promising beginning until novel techniques—such as behavioral studies following microablations of targeted brain tissues using monoclonal antibodies, or direct observations of optical changes in firing neurons [as pioneered by Grinvald and co-workers (Ts'o *et al.*, 1990)]—give us insight into the workings of the higher brain centers.

The physicist who wants to contribute to vision research can either learn psychobiology from scratch and apply his or her knowledge of mathematics and physics where it is applicable or try to suggest some novel principles to brain researchers regardless of whether these are neurophysiologically plausible or not. After all, what is not plausible in psychobiology now might become probable decades later! The latter kind of contribution can be useful as an AI model or can be applied in robotics or machine vision. For both approaches there are many famous examples.

The best known physicists and engineers who have directly contributed to psychobiology are Hermann von Helmholtz, Ernst Mach, Georg von Békésy, and Werner Reichardt, among others. The physicists and engineers who indirectly contributed to brain research by introducing fundamental concepts, such as the programmable digital computer, information theory, cybernetics, analogies to memories, AI models, and so on are Jonn von Neumann, Leo Szilárd, Allan Turing, Claude Shannon, Norbert Wiener, Dennis Gabor, David Marr, and John Hopfield, to name a few of the best known.

As an example of the latter approach, I should like to draw attention to the delightful monograph by Valentino Braitenberg (1984). The author's main message, that "analysis is uphill, synthesis is downhill," is illustrated by building a succession of simple miniature vehicles of increasing complexity. The vehicles have two motors and a few sensors. Already such a simple operation as crossing the left and right photodetectors to be connected to the right and left rear motors (that drive the wheels) results in behavior that an outside observer might interpret as "fear" and "aggression." The adding of inhibitions between sensors and motors produces behavior that an anthropomorphic observer might construe as "love." Some nonmonotonic connections result in behavior that resembles "decision making." The inclusion of wires that exhibit hysteresis yields behavior resembling "concepts." Introducing some random connections (mutations) and selection-of-the-fittest conditions (lethal mutations cause some vehicles to fall off the tabletop), which the author calls "the impersonal engineer," results in vehicles of ever-increasing sophistication, with concepts of "forms," "ideas," "rules and regularities," "trains of thought," "foresight," "egotism," and "optimism." This wide range of complex behavior is merely the outcome of manipulating the interconnections between a few sensors and motors and observing the outcome. If instead of this synthetic approach (the essence of the AI paradigm), a designer were asked to construct a machine with some given complex behavior (the analytic approach of science), usually this could not be done at present. Of course, to regard such crude behavior as, for instance, "the circling of an object and periodically returning to it" as "love" borders on the comical. Nevertheless, thirty years ago theorists of behaviorist psychology still had considerable influence on psychologists, and their restriction of studying mental phenomena by behavioral manifestations alone still has followers who might take the philosophy of vehicles quite seriously.

Similarly, one can treat the transducers of brightness and color as an "ideal observer" and ask how well the visual system emulates such a hypothetical observer. For instance, Buchsbaum and Gottschalk (1983) utilized the ideal-observer approach to extract maximum information from the red, green, and blue color mechanisms of given spectral characteristics. After solving the eigenfunctions of the connectivity matrices, they found a system of connections that in essence are the Hering (1878) color system of red-green and yellow-blue opposing mechanisms. By the same token, I am curious as to why nobody has found, so far, Hadamard transforms that would optimize signal-noise ratios, which might be a desirable requirement for the visual system that can detect a few quanta. [On quantum efficiency of vision, I recommend the beautiful textbook by Cornsweet (1970).] I selected this example, out of many, to illustrate the point that interesting mathematical and physical ideas might have been stumbled upon by "the impersonal engineer" during hundreds of millions of years of evolution; nevertheless, it is almost as interesting when human ingenuity suggests solutions that are better than those found by trial and error in Nature (such as the Hadamard filters nowadays routinely used in measuring devices of all kinds). Another approach to applying mathematical ideas to brain research is provided by Resnikoff (1987b). He assumes some similarity between graph theory and brains by regarding the human brain as being randomly connected (a rather unlikely hypothesis) and applying the stochastic theorem of Erdös and Rényi (1959) [that estimates the minimum number of connections (synapses) between neurons that would permit a random graph (neural network) to be completely connected]. He argues that the human brain consumes about 15 watts (that is, the equivalent of a dim lightbulb, so a "bright idea" is not really very bright) and assumes that it has 10^{12} neurons. Assuming further that each neuron has about 10^4 synapses and using the Erdös-Rényi stochastic theorem, he finds that the brain must have at least 14 connections per neuron to be totally connected with a probability of 0.988, while with fewer connections the brain would fall into disconnected parts.

For the physicist, who is accustomed to dealing only with local interactions, early vision must be a familiar discipline. Indeed, much of the interaction occurs between adjacent neurons often exhibiting cooperative perceptual phenomena. When the stimulus moves fast, reaching neural propagation velocities (which are quite slow for nonmyelinated axons), one can actually observe some of the relativistic foreshortening phenomena of special relativity theory (Caelli *et al.* 1978b), where instead of the speed of light one has to postulate the maximum neural propagation velocity as the ultimate limit in the Lorentz transformation. For instance, disks moving increasingly faster might appear as shrinking ellipses in the direction of travel. However, there are processes that travel much faster through myelinated axons, which appear as "tachyons" representing "action at a distance" among the slowly propagating non-myelinated neural networks. Here is a fascinating world of relativistic and nonrelativistic phenomena mixed together in sophisticated ways.

The theoretical physicist fascinated with problems of "complexity," "self-similarity," "scale," "fractals," and "chaos" will find the human brain the proper structure for studying such problems. As an example, the reader can easily draw some connected and disconnected complex line drawings, that cannot be perceived as such by the human visual system. On the other hand, by reducing the complexity of these line drawings [e.g., as in Fig. 11(a)], we can effortlessly perceive connectivity. However, these same simple line drawings in aggregates do not yield texture discrimination without scrutiny [as illustrated by Fig. 11(b)]. Our limitations in perceiving connectivity in complex line drawings and in aggregates of simple line drawings is related to the self-similarity of perceiving connectivity at different scales. From simple and complex "mazes" to "impossible figures," one can juxtapose local and global rules of connections such that, at some scale, the solution of the maze or the impossibility of an object pops out, while at a finer scale these properties become hidden.

Finally, the concept of "chaos" in deterministic systems is not very new to me. Over decades I used to get the largest (unclassified) pseudo-random-number generators with the longest cycles of repetition. I used the same pseudo-random generator to create the left and right images of a dynamic RDS, respectively, and even after millions of such stereograms composed of 1000×1000 pixels, one could stereoscopically fuse them provided the computer did not make some error. This should drive home my point that computers are deterministic systems. They can exhibit complex behavior that resembles chaos, but cannot in fact generate randomness! Nevertheless, I am impressed by the chaotic behavior, exhibited as fractals, at the boundaries between stable domains in the phase space of complex adaptive systems. It is suggested by Stewart Kauffman, among others, that learning (e.g., the evolution of life) does take place at "the edge of chaos."

In such a condensed review, I have had to omit hundreds of important contributions and concentrate on authors who share my outlook. I have no doubt that if other psychologists had the opportunity to write a review for this journal, a very different story might be told, probably with a completely different list of references. Indeed, workers in visual perception belong to many camps. The majority believe in the "direct vision" paradigm of J. J. Gibson (1966). Gibson—whose ideas had a great effect on me in my youth—assumed that mental processes "resonate" to certain hidden structures in the environment, as a tuning fork might resonate to a specific tone. [However, metaphors of this kind, according to Peter Medwar, as quoted by Ramachandran (1990), are "mere analgesics; they dull the ache of incomprehension without removing the cause."] Others adhere to the view that perception is based on "unconscious inferences" as originally posited by Helmholtz (1867). [Interestingly, Helmholtz introduced the concept of the "unconscious" many years before Sigmund Freud.] Researchers in AI believe in the paradigm of "natural computation." Several leading psychologists cling to old paradigms, from behaviorism to Gestalt theories, while others embrace radically different paradigms, such as mental imagery. [That paradigms die hard has been convincingly discussed by Kuhn (1972), and indeed Newton's theory of gravitation is still taught, although Einstein's general relativity theory has "falsified" it.] Some others, particularly Ramachandran (1990), assume that visual perception is based on a "bag of tricks." This last-mentioned paper by Ramachandran contains much useful information and several perceptual demonstrations. I end this review with a quotation from Ramachandran (1990, p. 22): "[The early visual system] . . . *takes advantage of constraints that incorporate general properties of the world rather than top-down influences that depend on high-level semantic knowledge of specific objects. Thus the visual system may have built-in knowledge about surfaces, depth, movement, etc., but not about umbrellas, chairs, and dalmatian dogs.*"

ACKNOWLEDGMENTS

I thank Dr. Gérard Toulouse for inviting me to write this review on early vision and advising me to assume that the readers have no previous knowledge of psychobiology. He also advised me to give a set of introductory references before the Introduction.

I thank Dr. Douglas Williams, Dr. Stanley Klein, Dr. Jochen Braun, and Dr. Francis Crick for their helpful comments while reading this manuscript, and Ms. Carol Esso for helping me with the word processing, particularly with the references.

REFERENCES

Adelson, E. H., and J. K. Bergen, 1985, "Spatio-temporal energy models for the perception of motion," J. Opt. Soc. Am. **2**, 284–299.

Amit, D. J., 1989, *Modelling Brain Function: The World of Attractor Neural Networks* (Cambridge University, Cambridge/New York).

Anderson, C. H., and D. C. Van Essen, 1987, "Shifter circuits: A computational strategy for dynamic aspects of visual processing," Proc. Nat. Acad. Sci. USA **84**, 6297–6301.

Anstis, S. M., 1978, "Apparent movement," in *Handbook of Physiology. Vol. VIII. Perception,* edited by R. Held, H. W. Leibowitz, and H. L. Teuber (Springer, Berlin), pp. 655–673.

Babington-Smith, B., 1977, "A wartime anticipation of random-dot stereograms," Perception **6**, 233–234.

Barlow, H. B., C. Blakemore, and J. D. Pettigrew, 1967, "The neural mechanism of binocular depth discrimination," J. Physiol. **193**, 327–342.

Barlow, H. B., and J. D. Mollon, 1982, Eds., *The Senses* (Cambridge University, Cambridge, England).

Bergen, J. R., and E. H. Adelson, 1988, "Early vision and texture perception," Nature **333**, 363–364.

Bergen, J. R., and B. Julesz, 1983, "Parallel versus serial processing in rapid pattern discrimination," Nature **303**, 696–698.

Bernstein, Jeremy, 1984, *Three Degrees above Zero: Bell Labs in the Information Age* (Scribner, New York), Chap. 3, "Bela Julesz," pp. 39–48.

Bishop, P. O., 1963, "Neurophysiology of binocular single vision and stereopsis," in *Handbook of Sensory Physiology,* edited by R. Jung (Springer, Berlin), Vol. 7, pp. 255–303.

Blakemore, C., 1969, "Binocular depth discrimination and the nasotemporal division," J. Physiol. **205**, 471–497.

Blakemore, C., and F. W. Campbell, 1969, "Adaptation to spatial stimuli," J. Physiol. **200**, 11–13.

Blakemore, C., and G. F. Cooper, 1970, "Development of the brain depends on the visual environment," Nature **228**, 477–478.

Blakemore, C., and B. Julesz, 1971, "Stereopsis depth aftereffect produced without monocular cues," Science **171**, 386–388.

Boff, K. R., L. Kaufman, and J. P. Thomas, 1986, *Handbook of Perception and Performance. Vol. 1. Sensory processes and perception* (Wiley, New York).

Boynton, R., 1986, Ed., Reprinted version of the 25th Jubilee Issue of Vision Res. **26**(9), 1601–1629.

Braddick, O., 1974, "A short-range process in apparent motion," Vision Res. **14**, 519–527.

Braddick, O. J., 1980, "Low-level and high-level processes in apparent motion," Philos. Trans. R. Soc. London, Ser. B **290**, 137–151.

Braddick. O., J. Atkinson, B. Julesz, W. Kropfl, I. Bodis-Wollner, and E. Rabb, 1970, "Cortical binocularity in infants," Nature **288**, 363–365.

Braitenberg, V., 1984, *Vehicles: Experiments in Synthetic Psychology* (MIT, Cambridge, MA).

Braun, J., and D. Sagi, 1990, "Vision outside the focus of attention," Percept. Psychophys. **48**, 45–58.

Broadbent, D. C., 1958, *Perception and Communication* (Pergamon, Oxford).

Bruce, V., and P. Green, 1985, *Visual Perception: Physiology, psychology, and ecology* (Erlbaum, Hillsdale, NJ).

Buchsbaum, G., and A. Gottschalk, 1983, "Trichromacy, opponent colours coding and optimum colour information transmission in the retina," Proc. R. Soc. London, Ser. B **220**, 89–113.

Bülthoff, H. H., and H. A. Mallot, 1990, "Integration of stereo, shading, and texture," in *AI and the Eye,* edited by A. Blake and T. Troscianko (Wiley, New York), pp. 119–146.

Burt, P., and B. Julesz, 1980, "A disparity gradient limit for binocular fusion," Science **208**, 615–617.

Caelli, T. M., and B. Julesz, 1979, "Psychophysical evidence for global feature processing in visual texture discrimination," J. Opt. Soc. Am. **69**(5), 675–678.

Caelli, T. M., B. Julesz, and E. Gilbert, 1978, "On perceptual analyzers underlying visual texture discrimination: Part II," Biol. Cybernetics **29**, 201–214.

Caelli, T. M., W. C. Hoffman, and H. Lindman, 1978, "Subjective Lorentz transformations and the perception of motion," J. Opt. Soc. Am. **68**, 402–417.

Cavanagh, P., and G. Mather, 1989, "Motion: The long and short of it," Spatial Vision **4**, 103–129.

Chang, J. J., and B. Julesz, 1983, "Displacement limits, directional anisotropy, and direction versus form discrimination in random-dot cinematograms," Vision Res. **23**, 639–646.

Chang, J. J., and B. Julesz, 1984, "Cooperative phenomena in apparent movement perception," Vision Res. **24**, 1781–1788.

Chang, J. J., and B. Julesz, 1990, "Low-level processing of disparity-tuned binocular neurons takes precedence of shape from shading," Invest. Ophthalm. Visual Sci. **31**(4), 525 (Proceedings of ARVO Meeting, May, 1990).

Churchland, P. S., 1986, *Neurophilosophy. Toward a unified science of the mind-brain* (MIT, Cambridge, MA).

Cornsweet, T. N., 1970, *Visual Perception* (Academic, New York).

Desimone, R., and L. G. Ungerleider, 1989, "Neural mechanisms of visual processing in monkeys," in *Handbook of Neurophysiology. Vol. 2,* edited by F. Boller and J. Grafman (Elsevier, New York/Amsterdam), pp. 267–299.

Dev, P., 1974, "Segmentation processes in visual perception: a cooperative neural model," COINS Tech. Rep. 74C-5, Computer Information Science, University of Massachusetts.

Dev, P., 1975, "Perception of depth surfaces in random-dot stereograms: A neural model," Int. J. Man-Mach. Stud. **7**, 511–528.

DeValois, R. L., D. G. Albrecht, and L. G. Thorell, 1978, "Cortical cells: Bar detectors or spatial filters?" in *Frontiers of Visual Science,* edited by S. J. Cool and E. L. Smith (Springer, Berlin).

Dodwell, P. C., 1983, "The Lie transformation model of visual perception," Percept. Psychophys. **34**, 1–16.

Edelman, G. M. 1989, *The remembered Present. A biological theory of consciousness* (Basic, New York).

Edelman, G. M., W. E. Gall, and W. M. Cowan, 1984, Eds., *Dynamic Aspects of Neocortical Function* (Wiley, New York).

Enns, J., 1986, "Seeing textons in context," Percept. Psychophys. **39**, 143–147.

Erdös, P., and A. Rényi, 1959, "On random graphs. I." Publ. Math. Debrecen **6**, 290–297.

Fender, D., and B. Julesz, 1967, "Extension of Panum's fusional area in binocularly stabilized vision," J. Opt. Soc. Am. **57**(6), 819–830.

Fodor, J., 1983, *Modularity of Mind* (MIT, Cambridge, MA).

Fodor, J. A., and Z. W. Pylyshyn, 1988, "Connectionism and cognitive architecture: A critical analysis," in *Connections and Symbols,* edited by S. Pinker and J. Mehler (MIT, Cambridge, MA), pp. 3–71.

Fogel, I., and D. Sagi, 1989, "Gabor filters as texture discriminators," Biol. Cybernetics **61**, 103–113.

Fox, R., R. N. Aslin, S. L. Shea, and S. T. Dumais, 1980, "Stereopsis in human infants," Science **207**, 323–324.

Gabor, D., 1946, "Theory of communication," J. Inst. Electr. Eng. (London) **93**, 429–457.

Gazzaniga, M. S., 1989, "Organization of the human brain," Science **245**, 947–952.

Gorea, A., and B. Julesz, 1990, "Context superiority in a detection task with a line-element stimuli: A low-level effect," Perception **19**, 5–16.

Gregory, R. L., 1987, Ed., *The Oxford Companion to the Mind* (Oxford University, Oxford/New York).

Gregory, R. L., 1990, *Eye and Brain: The Psychology of Seeing* (Princeton University, Princeton, NJ).

Grossberg, S., 1988, Ed., *Neural Networks and Natural Intelligence* (MIT, Cambridge, MA).

Grossberg, S., and E. Mingolla, 1985, "Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations," Percept. Psychophys. **38**, 141–171.

Gurnsey, R., and R. Browse, 1987, "Micropattern properties and presentation conditions influencing visual texture discrimination," Percept. Psychophys. **41**, 239–252.

Hadani, I., G. Ishai, and M. Gur, 1978, "Visual stability and space perception in monocular vision: A mathematical model," Biomedical Engineering Technical Report: TSI (Israel Institute of Technology) May 1978.

Hadani, I., G. Ishai, and M. Gur, 1980, "Visual stability and space perception in monocular vision: mathematical model," J. Opt. Soc. Am. **70**, 60–65.

Harris, C. S., 1980, Ed., *Visual Coding and Adaptability* (Erlbaum, Hillsdale, NJ).

Heilman, K. M., R. T. Watson, E. Valenstein, and M. E. Goldberg, 1987, "Attention: Behavior and neural mechanisms," Chapter 11 in *Handbook of Physiology, Section 1: The Nervous System,* Volume V, edited by V. B. Mountcastle, F. Plum, and S. R. Geiger (American Physiological Society, Bethesda, MD), pp. 461–481.

Helmholtz, H. von, 1896, *Handbuch der Physiologischen Optik. Dritter Abschnitt, Zweite Auflage* (Voss, Hamburg). [Published in English as *Helmholtz's Treatise on Physiological Optics,* by J. P. C. Southall, 1924 (The Optical Society of America); republished by Dover, New York, 1962.]

Hering, E., 1878, *Zur Lehre vom Lichtsinn* (Gerold's Sohn, Vienna).

Hildreth, E., 1984, *The Measurement of Visual Motion* (MIT, Cambridge, MA).

Hillis, D. W., 1985, *The Connection Machine* (MIT, Cambridge, MA).

Hirsch, H. V. B., and D. N. Spinelli, 1970, "Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats," Science **168**, 868–871.

Hoffman, W. C., 1970, "Higher visual perception as prolongations of the basic Lie transformation group," Math. Biosci. **6**, 437–471.

Hopfield, J. J., 1984, "Neural networks and physical systems with emergent collective computational abilities," Proc. Nat. Acad. Sci. USA **79**, 2554–2558.

Horn, B. K. P., and M. J. Brooks, 1985, "The variational approach to shape from shading," MIT Artificial Intelligence Lab. Memo No. **813**, 1–32.

Hubel, D. H., 1988, *Eye, Brain, and Vision* (W. H. Freeman, New York).

Hubel, D. H., and T. N. Wiesel, 1960, "Receptive fields of optic nerve fibres in the spider monkey," J. Physiol. **154**, 572–580.

Hubel, D. H., and T. N. Wiesel, 1962, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," J. Physiol. **160**, 106–154.

Hubel, D. H., and T. N. Wiesel, 1968, "Receptive fields and functional architecture of the monkey striate cortex," J. Physiol. **195**, 215–243.

Hubel, D. H., and T. N. Wiesel, 1970, "Cells sensitive to binocular depth in area 18 of the macaque monkey cortex: Stereoscopic vision in the macaque monkey," Nature **225**, 41–42.

Hubel, D. H., and T. N. Wiesel, 1974, "Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor," J. Comp. Neurol. **158**, 295–306.

Humphreys, G. W., and V. Bruce, 1989, *Visual Cognition: Computational, Experimental, and Neuropsychological Perspectives* (Erlbaum, Hillsdale, NJ).

James, W., 1890, *The Principles of Psychology* (republished in 1925 by Dover, New York).

Judd, S., 1987, "Learning in networks is hard," in *IEEE First International Conference on Neural Networks,* Vol. II, pp. 685–692.

Julesz, B., 1960, "Binocular depth perception of computer-generated patterns," Bell Syst. Tech. J. **39**, 1125–1162.

Julesz, B., 1962, "Visual pattern discrimination," IRE Trans. Inf. Theory **IT-8**, 84–92.

Julesz, B., 1964, "Binocular depth perception without familiarity cues, Science **145**, 356–362.

Julesz, B., 1971, *Foundations of Cyclopean Perception* (University of Chicago, Chicago).

Julesz, B., 1975, "Experiments in the visual perception of texture," Sci. Am. **232**, (April), 34–43.

Julesz, B., 1978, "Global stereopsis: Cooperative phenomena in stereoscopic depth perception," in *Handbook of Physiology, Vol. VIII, Perception,* edited by R. Held, H. W. Leibowitz, and H. Teuber (Springer, Berlin), pp. 215–256.

Julesz, B., 1980, "Spatial frequency channels in one-, two-, and three-dimensional vision: Variations on an auditory theme by Bekesy," in *Visual Coding and Adaptability,* edited by C. S. Harris (Erlbaum, Hillsdale, NJ), pp. 263–316.

Julesz, B., 1981, "Textons, the elements of texture perception and their interactions," Nature **290**, 91–97.

Julesz, B., 1984, "Toward an axiomatic theory of preattentive vision," in *Dynamic Aspects of Neocortical Function,* edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 585–612.

Julesz, B., 1986a, "Stereoscopic vision," Vision Res. **26**(9), 1601–1612.

Julesz, B., 1986b, "Texton gradients: The texton theory revisited," Biol. Cybernetics **54**, 245–251.

Julesz, B., 1990a, "AI and early vision," in *AI and the Eye,* edited by A. Blake and T. Troscianko (Wiley, New York), pp. 9–20.

Julesz, B., 1990b, "Concepts in early vision," in *Synergetics in Cognition,* edited by H. Haken and M. Stadler (Springer, Berlin), pp. 218–232.

Julesz, B., 1990c, "Early vision is bottom-up, except for focal attention," in *Symposium #55: The Brain* (Cold Spring Harbor Laboratory Press, Long Island, NY).

Julesz, B., 1991a, "Some strategic questions in visual perception," in *Representations in Vision: Trends and Tacit Assumptions in Vision Research,* edited by A. Gorea (Cambridge University, Cambridge, England), pp. 331–349.

Julesz, B., 1991b, "Consciousness and focal attention," Behav. Brain Sci. (in press).

Julesz, B., and J. R. Bergen, 1983, "Textons, the fundamental elements in preattentive vision and perception of textures," Bell Syst. Tech. J. **62**(6), 1619–1645.

Julesz, B., and J. J. Chang, 1976, "Interaction between pools of binocular disparity detectors tuned to different disparities," Biol. Cybernetics **22**, 107–119.

Julesz, B., E. N. Gilbert, and J. D. Victor, 1978, "Visual discrimination of textures with identical third-order statistics," Biol. Cybernetics **31**(3), 137–140.

Julesz, B., W. Kropfl, and B. Petrig, 1980, "Large evoked potentials to dynamic random-dot correlograms and stereograms permit quick determination of stereopsis," Proc. Nat. Acad. Sci. USA **77**(4), 2348–2351.

Julesz, B., and B. Kröse, 1988, "Visual texture perception: Features and spatial filters," Nature **333**, 302–303.

Julesz, B., and J. E. Miller, 1975, "Independent spatial frequency tuned channels in binocular fusion and rivalry," Perception **4**, 125–143.

Julesz, B., and R. A. Payne, 1968, "Differences between monocular and binocular stroboscopic movement perception," Vision Res. **8**, 433–444.

Julesz, B., and C. W. Tyler, 1976, "Neurontropy, an entropy-like measure of neural correlation in binocular fusion and rivalry," Biol. Cybernetics **23**, 25–32.

Kandel, E. R., and J. H. Schwartz, 1985, *Principles of Neural Science,* 2nd edition (Elsevier/North-Holland, New York/Amsterdam).

Kanizsa, G., 1976, "Subjective contours," Sci. Am. **234**, 48–52.

Karni, A., and D. Sagi, 1990, "Where practice makes perfect in texture discrimination," The Weizmann Institute of Science Technical Report No. CS90-02, Jan. 1990. Also in Proc. Nat. Acad. Sci. USA (1991, in press).

Kröse, B. J. A., 1987, "Local structure analyzers as determinants of preattentive pattern discrimination," Biol. Cybernetics **55**, 289–298.

Kröse, B. J. A., and B. Julesz, 1989, "The control and speed of shifts of attention," Vision Res. **29**(11), 1607–1619.

Kuffler, S. W., 1953, "Discharge patterns and functional organization of mammalian retina," J. Neurophysiol. **16**, 37–68.

Kuhn, T., 1972, *The Structure of Scientific Revolutions.* New edition (University of Chicago, Chicago).

LeDoux, J. E., and W. Hirst, 1986, Eds., *Mind and Brain: Dialogues in Cognitive Neuroscience* (Cambridge University, Cambridge, England).

Lehky, S. R., and T. J. Sejnowsky, 1990, "Neural model of stereoacuity and depth interpretation based on a distributed representation of stereo disparity," J. Neurosci. **10**, 2281–2299.

Lehmann, D., and B. Julesz, 1978, "Lateralized cortical potentials evoked in humans by dynamic random-dot stereograms," Vision Res. **18**, 1265–1271.

Lettvin, J. Y., H. R. Maturana, W. S. McCulloch, and W. H. Pitts, 1959, "What the frog's eye tells to the frog's brain," Proc. Inst. Radio Engrs. **47**, 1940–1951.

Livingstone, M. S., and D. H. Hubel, 1984, "Anatomy and physiology of a color system in the primate visual cortex," J. Neurosci. **4**, 309–356.

Luneburg, R. K., 1950, "The metric of binocular visual space," J. Opt. Soc. Am. **40**, 627–642.

MacAdam, D. L., 1970, *Sources of Color Science* (MIT, Cambridge, MA).

Mach, E., 1886, *The analysis of sensations and the relation of the physical to the psychical.* First German edition in 1886; republished by Dover (New York, 1959), using the 5th German edition revised and supplemented by S. Waterlow.

Mahowald, M. A., and T. Delbrück, 1989, "Cooperative stereo matching using static and dynamic image features," in *Analog ELSI Implementation of Neuro Systems,* edited by C. Mead and M. Ismail (Kluwer Academic, Norwell, MA), pp. 213–238.

Malik, J. M. A., and P. Perona, 1990, "Preattentive texutre discrimination with early vision mechanisms," J. Opt. Soc. Am. **7**(5), 923–932.

Marr, D., 1982, *Vision* (Freeman, San Francisco).

Marr, D., and T. Poggio, 1976, "Cooperative computation of stereo disparity," Science **194**, 283–287.

Marr, D., and T. Poggio, 1979, "A theory of human stereopsis," Proc. R. Soc. London, Ser. B **204**, 301–328.

Mayhew, J. E. W., and J. P. Frisby, 1979, "Convergent disparity discriminations in narrow-band-filtered random-dot stereograms," Vision Res. **19**(1), 63–71.

Melzak, Z. A., 1976, *Mathematical ideas, modeling & applications.* Vol. II of Companion to *Concrete Mathematics* (Wiley, New York/London).

Miezin, F. W., J. Myerson, B. Julesz, and J. M. Allman, 1981, "Evoked potentials to dynamic random-dot correlograms in monkey and man: A test for cyclopean perception," Vision Res. **21**, 177–179.

Mitchell, D. E., 1980, "The influence of early visual experience on visual perception," in *Visual Coding and Adaptability,* edited by C. S. Harris (Erlbaum, Hillsdale, NJ), pp. 1–50.

Moran, J., and R. Desimone, 1985, "Selective attention gates visual processing in the extrastriate cortex," Science **229**, 782–784.

Mountcastle, V. B., 1979, "An organizing principle for cerebral function: The unit module and the distributed system," in *The Neurosciences: Fourth Study Program,* edited by F. O. Schmidt and F. G. Worden (MIT, Cambridge, MA), pp. 21–42.

Mowforth, P., J. E. W. Mayhew, and J. P. Frisby, 1981, "Vergence eye movements made in response to spatial-frequency-filtered random-dot stereograms," Perception **10**(3), 299–304.

Müller, J., 1844, *Handbuch der Physiologie des Menschen,* 4th Edition (J. Hölscher, Coblenz).

Nakayama, K., 1985, "Biological image motion processing: A review," Vision Res. **25**, 625–660.

Nakayama, K., and M. Mackeben, 1989, "Sustained and transient components of focal visual attention," Vision Res. **29**, 1631–1647.

Nakayama, K., and G. H. Silverman, 1986, "Serial and parallel processing of visual feature conjunctions," Nature **320**, 264–265.

Nelson, J. I., 1975, "Globality and stereoscopic fusion in binocular vision," J. Theor. Biol. **49**, 1–88.

Nothdurft, H. C., 1985, "Sensitivity for structure gradient in texture discrimination tasks," Vision Res. **25**, 1957–1968.

Ogle, K. N., 1964, *Researchers in Binocular Vision* (Hafner, New York).

Paillard, J., 1977, Ed., "The Lie transformation group model for perceptual and cognitive psychology," Cah. Psychol. **2** (entire issue).

Papathomas, T. V., and B. Julesz, 1989, "Stereoscopic illusion based on the proximity principle," Perception **18**(5), 589–594.

Papert, S., 1961, "Centrally produced geometrical illusions," Nature **191**, 733.

Penrose, R., 1989, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford University, New York/Oxford).

Pentland, A. P., 1986, "Shading into texture," Artif. Intell. **29**, 147–170.

Petrig, B., 1980, "Nachweis von Stereopsis bei Kindern mittels stochasischer Punktstereogramme und der zuhörigen evozierte Potenziale," Ph.D. thesis, ETH, Zürich.

Petrig, B., B. Julesz, W. Kropfl, G. Baumgartner, and M. Anliker, 1981, "Development of stereopsis and cortical binocularity in human infants: Electrophysiological evidence," Science **213**, 1402–1405.

Poggio, G. F., 1984, "Processing of stereoscopic information in primate visual cortex," in *Dynamic Aspects of Neocortical Function,* edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 613–635.

Poggio, G. F., F. Gonzalez, and F. Krause, 1988, "Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity sensitivity," J. Neurosci. **8**, 4531–4550.

Poggio, T., V. Torre, and C. Koch, 1989, "Computational vision and regularization theory," Nature **317**, 314–319.

Polanyi, M., 1969, *Knowing and Being* (University of Chicago, Chicago).

Posner, M. I., 1978, "Orienting of attention," Q. J. Exp. Psychol. **32**, 3–25.

Pylyshyn, Z. W., 1984, *Computation and Cognition. Toward a Foundation of Cognitive Science* (MIT, Cambridge, MA).

Ramachandran, V. S., 1988, "Perception of shape from shading," Nature 163–166.

Ramachandran, V. S., 1990, "Visual perception in people and machines," in *AI and the Eye,* edited by A. Blake and T. Troscianko (Wiley, New York), pp. 21–77.

Regan, D., 1989, *Human Brain Electrophysiology* (Elsevier, New York/Amsterdam).

Regan, D., and K. D. Beverley, 1985, "Visual responses to verticity and the neural analysis of optic flow," J. Opt. Soc. Am. **2**, 280–283.

Reichardt, W., 1961, "Autocorrelation, a principle of evaluation of sensory information by the central nervous system," in *Sensory Coding,* edited by W. A. Rosenbluth (Wiley, New York), pp. 303–317.

Reichardt, W., and M. Egelhaaf, 1988, Naturwissenschaften **75**, 313–315.

Resnikoff, H. L., 1987a, *The Illusion of Reality* (Springer, New York).

Resinkoff, H. L., 1987b, "Concurrent computation and models of biological information processing," in *Advances in Cognitive Science,* edited by M. Kochen, and H. M. Hastings (Westview, Boulder, CO), pp. 78–98.

Richards, W. A., 1988, Ed., *Natural Computation: Selected Readings* (MIT, Cambridge, MA).

Rock, I., 1984, *Perception* (Sci. Am. Library, Freeman, San Francisco).

Rock, I., and S. Palmer, 1990, "The legacy of Gestalt psychology," Sci. Am. **12**(Dec.), 84–90.

Rubenstein, B., and D. Sagi, 1990, "Spatial variability as a limiting factor in texture-discrimination tasks: Implications for performance asymmetries," J. Opt. Soc. Am. A **9**, 1632–1643.

Rumelhart, D. E., and J. L. McClelland, 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (MIT, Cambridge, MA).

Saarinen, J., and B. Julesz, 1991, "The speed of attentional shifts in the visual field," Proc. Nat. Acad. Sci. USA **88**, 1812–1814.

Sagi, D., 1991, "Detection of an orientation singularity in Gabor textures: Effect of signal density and spatial-frequency," Vision Res. (in press).

Sagi, D., and B. Julesz, 1985, "'Where' and 'what' in vision," Science **228**, 1217–1219.

Sagi, D., and B. Julesz, 1986, "Enhanced detection in the aperture of focal attention during simple discrimination tasks," Nature **321**, 693–695.

Sagi, D., and B. Julesz, 1987, "Short-range limitation of detection of feature differences," Spatial Vision **2**, 39–49.

Saltzman, Z. D., K. H. Britten, and W. T. Newsome, 1990, "Cortical microstimulation influences perceptual judgments of motion direction," Nature **346**, 589.

Schumann, F., 1904, "Einige Beobachtungen über die Zusammenfassung von Gesichtseindrucken zu Einheiten," Psychol. Stud. **1**, 1–32.

Sekuler, R., and R. Blake, 1985, *Perception* (Knopf, New York).

Shepard, R. N., and J. Metzler, 1971, "Mental rotation of three-dimensional objects," Science **171**, 701–703.

Snowden, R. J., and O. J. Braddick, 1990, "Differences in the processing of short-range apparent motion at small and large displacements, "Vision Res. **30**, 1211–1222.

Sokoloff, L., 1984, *Metabolic Probes of Central Nervous System Activity in Experimental Animals and Man* (Sinauer Associates, Sunderland, MA).

Sperling, G., 1970, "Binocular vision: A physical and neural theory," Am. J. Psychol. **83**, 461–534.

Sperling, G., 1978, "The Goal of Theory in Experimental Psychology," Bell Teleph. Lab. Tech. Memo #78-1221-12.

Sperry, R. W., 1982, "Some effects disconnecting the central hemispheres, "Science **217**, 1223–1226.

Sternberg, S., 1966, "High-speed scanning in human memory," Science **153**, 652–654.

Teller, D. Y., 1980, "Locus questions in visual science," in *Visual Coding and Adaptability,* edited by C. S. Harris (Erlbaum, Hillsdale, NJ), pp. 151–176.

Thompson, P., 1980, "Margaret Thatcher: A new illusion," Perception **9**, 483–484.

Treisman, A., and G. Gelade, 1980, " A feature-integration theory of attention," Cognitive Psychol. **12**, 97–136.

Treisman, A., and S. Gormican, 1988, "Feature analysis in early vision: Evidence from search asymmetries," Psychol. Rev. **95**, 15–48.

Treisman, A., and R. Paterson, 1984, "Emergent features, attention and object perception," J. Exp. Psychol. **10**, 12–31.

Ts'o, D. Y., R. D. Frostig, E. E. Lieke, and A. Grinvald, 1990, "Functional organization of primate visual cortex revealed by high-resolution optical imaging," Science **249**, 417–420.

Ullman, S., 1977, "The Interpretation of Visual Motion," Ph.D. thesis, Department of Electrical Engineering and Computer Science, MIT.

Ullman, S., 1979, *The Interpretation of Visual Motion* (MIT, Cambridge, MA).

Ullman, S., 1984, "Visual routines," Cognition **18**, 97–159.

Uttal, W. R., 1990, "On some two-way barriers between models and mechanisms," Percept. Psychophys. **48**, 188–203.

Van de Grind, W. A., A. J. Van Doorn, and J. J. Koenderink, 1983, "Detection of coherent movement in peripherally viewed random-dot patterns," J. Opt. Soc. Am. **73**, 1674–1683.

Van Essen, D. C., E. A. DeYoe, J. Olavarria, J. Knierim, J. Fox, D. Sagi, and B. Julesz, 1989, "Neural responses to static and moving texture patterns in visual cortex of the macaque monkey," in *Neural Mechanisms of Visual Perception,* edited by D. M. K. Lam and C. D. Gilbert (Portfolio, Woodlands, TX), pp. 137–154.

van Santen, J. P. H., and G. Sperling, 1985, "Elaborated Reichardt-detectors," J. Opt. Soc. Am. **A2**, 451–473.

von der Heydt, R., E. Peterhans, and G. Baumgartner, 1988, "Illusory contours and cortical neuron responses," Science **224**, 1260–1262.

von Schelling, H., 1956, "Concepts of distance in affine geometry and its applications in theory of vision," J. Opt. Soc. Am. **46**, 309–315.

Voorhees, H., and T. Poggio, 1988, "Computing texture boundaries from images," Nature **333**, 364–367.

Watson, A. B., and A. J. Ahumada, 1985, "Model of human visual-motion sensing," J. Opt. Soc. Am. **A2**, 322–342.

Watt, R., 1988, *Visual Processing: Computational, Psychophysical, and Cognitive Research* (Erlbaum, Hillside, NJ).

Weichselgartner, E., and G. Sperling, 1987, "Dynamics of automatic and controlled visual attention," Science **238**, 778–779.

Weisstein, N., and C. S. Harris, 1974, "Visual detection of line segments: An object-superiority effect," Science **186**, 752–755.

Westheimer, G., 1979, "Cooperative neural processes involved in stereoscopic acuity," Exp. Brain Res. **36**, 585–597.

Williams, D. W., and B. Julesz, 1990, "Perceptual asymmetries in texture discrimination can be explained by bottom-up processing alone," in OSA Annual Meeting Technical Digest 1990, Vol. 15 of the OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1990), p. 51.

Williams, D. W., and B. Julesz, 1991, "Filters versus textons in human and machine texture discrimination," in *Neural Networks for Human and Machine Perception,* edited by H. Wechsler (Academic, Orlando, FL), in press.

Williams, D. W., G. C. Phillips, and R. W. Sekuler, 1986, "Hysteresis in the perception of motion direction: Evidence for neural cooperativity," Nature **324**, 253–255.

Wolfe, J. M., and K. R. Cave, 1990, "Deploying visual attention: The guided search model," in *AI and the Eye,* edited by A. Blake and T. Troscianko (Wiley, New York), pp. 79–103.

Zeki, S. M., 1980, "The representation of color in the cerebral cortex," Nature **284**, 412–418.

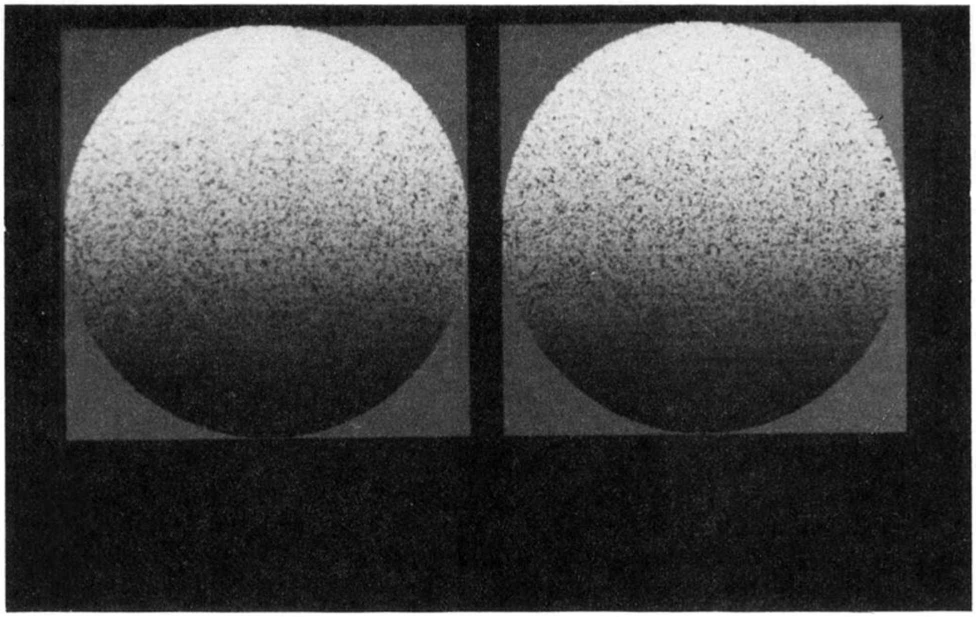
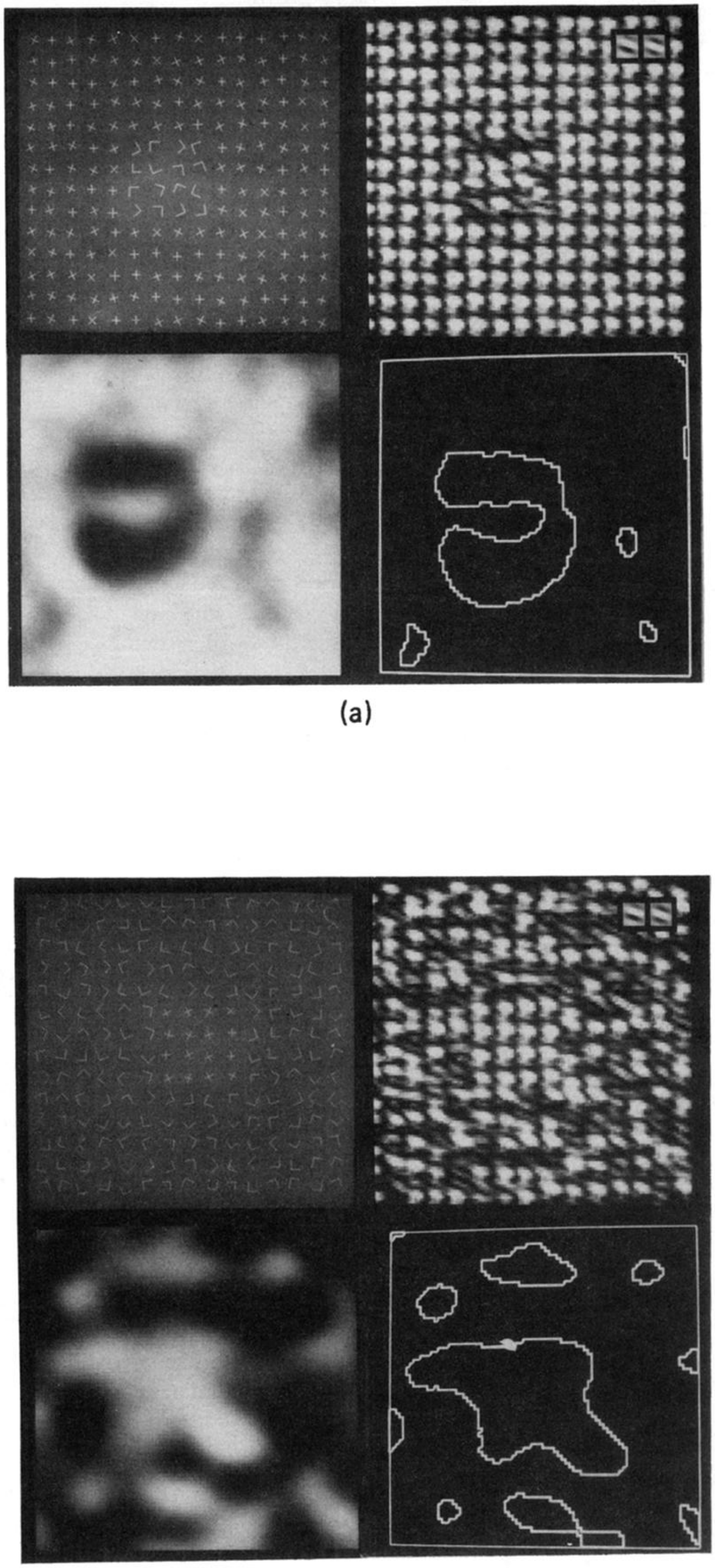


FIG. 10. Monocularly convex (concave) sphere due to depth (shape) shading, in which a random-dot stereogram (RDS) is mixed in. When binocularly fused, depth from stereopsis dominates perceived depth. When viewed monocularly, depth from shading yields a convex sphere, but appears concave when page is turned upside-down. Because the RDS has crossed binocular disparity, the binocularly fused image appears strongly convex. From Chang and Julesz (1990).



(a)

(b)

FIG. 18. Output of a nonlinear spatial filter that emulates the asymmetry phenomenon of human texture segmentation as illustrated by the dual texture pairs in (a) and (b). The first stage of Gabor filters is followed by a nonlinear operator, which in turn is followed by a second spatial filter. From Rubenstein and Sagi (1990).

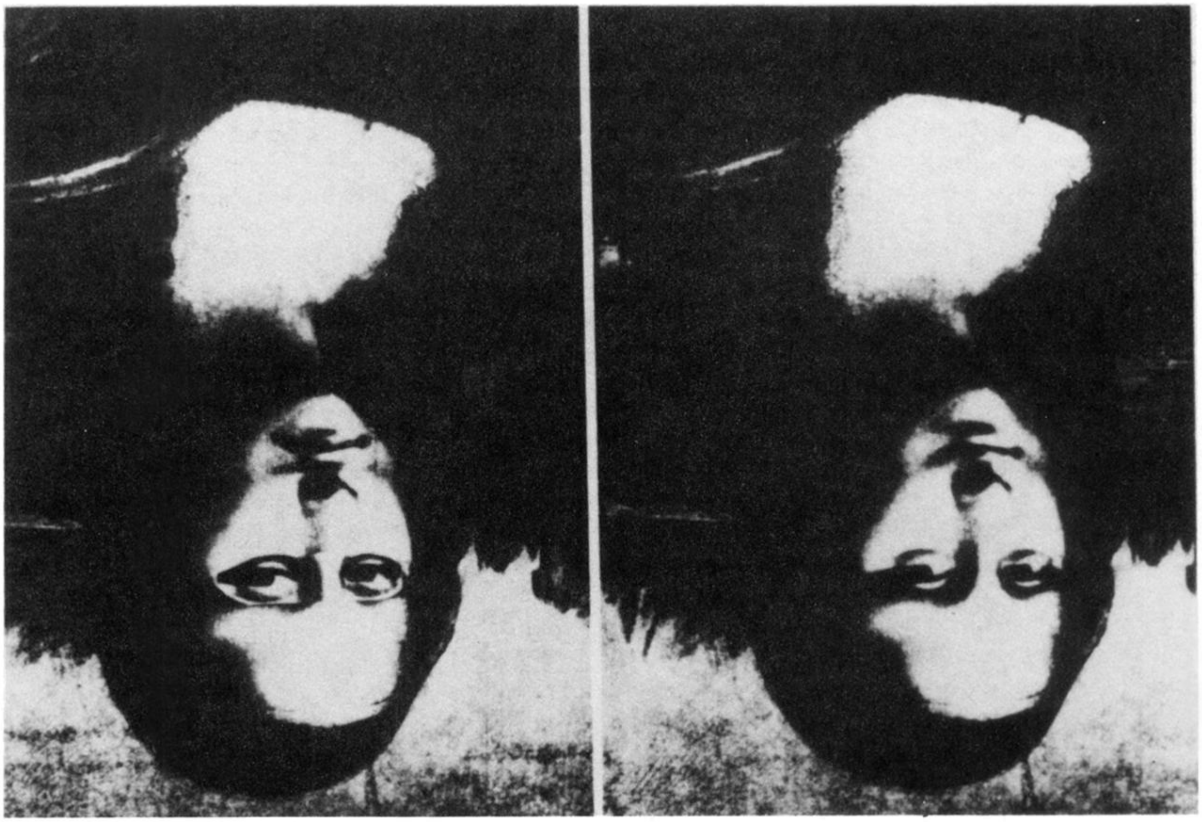


FIG. 3. Demonstration of Gestalt. The upside-down pictures appear rather similar in appearance, in spite of the fact that in one picture the eyes and mouth seem to be inverted. When the page is turned upside down, the two faces reveal a dramatic difference as a result of Gestalt organization. From Julesz (1984) after an idea of Thompson (1980).