Psychological Review
1977, Vol. 84, No. 1, 67–88

Contrasting Orientations to the Theory of Visual Information Processing

M. T. Turvey
University of Connecticut and Haskins Laboratories

In this paper the concepts of iconic memory and schematic memory are used to examine two fundamental and related features of the contemporary theory of visual information processing. One is the orientation of indirect realism which, to put it bluntly, emphasizes the equivocality and inadequacy of the light at the eyes and the necessity of epistemic mediation. The other is the analysis of visual processing into discrete temporal cross sections perpendicular to the flow of optical information. That the two features are closely cognate is revealed in the interpretation of event perception—the perception of change wrought over an object or object complex—as a deduction from or assimilation of (epistemic mediators) a sequence of static arrangements (discrete, temporal cross sections) represented iconically or schematically. On rational and empirical grounds, it is argued (a) that the discrete sampling of a continuous optical flow is not a tenable assumption, (b) that the informational support for event perception cannot be static iconic or schematic memories, and (c) that the perception of style of change cannot be epistemically mediated. Insofar as indirect realism receives little support from the analysis of event perception, *direct* realism is given due consideration as an alternative and radically different orientation to the theory of visual information processing.

It is a relatively commonplace understanding that visual processing can be characterized as a succession of stages, of both storage and transformation, which map the arrangement of light at the receptors onto progressively more abstract representations. There are, of course, variants on this theme, mostly along lines that liberate the relations among the representations. While the more familiar hierarchy tends to define the relation between representations as unidirectional and immutable, a heterarchy permits considerable cross talk and commutability of roles.

The writing of this paper was supported, in part, by the following grants to Haskins Laboratories: Grant HD-01994 from the National Institute of Child Health and Human Development and General Research Support Grant RR-5596 from the National Institutes of Health.

The author wishes to express his appreciation to Carol Fowler, Robert Shaw, Robert Remez, and William Mace for their conversations, questions, criticism, and suggestions during the development of this paper.

Requests for reprints should be sent to M. T. Turvey, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511.

It should be emphasized, however, that although a number of different schemes of visual information processing can be described, and although some are considerably more formal and detailed than others, it is fair to claim that they are all tokens of the same type. It is also fair to claim that, in bold outline, the current conceptions of the "flow of visual information" depart relatively little from their predecessors. Helmholtz's (1911/1925) sketch of the "flow of visual information" in the mapping from proximal stimulus to percept, and its distant cognate, that of Alhazen a millenium before (see Lombardo, 1973), have been filled in to a significant degree by contemporary investigators and theorists but have not been significantly altered. By way of illustrating this claim, here are the cornerstones of Helmholtz's theory: a finite set of primitive elements; procedures to make infinite use of this finite set; the dependence of current perception on stored memories about the world; the need for nonlocal processing to make local interpretations unambiguous; and the nonimmediacy of perception, that is, perception as a process over time (cf. Hochberg, 1974; Turvey, in press-a).

The traditional and contemporary departure point for theories of visual processing is the assumption that the light at the eye is equivocal and impoverished with respect to environmental facts and events. Insofar as perception tends to be veridical, the proximal stimulus is said to *underdetermine* perception, and epistemic mediators are said to restore the balance. The term *epistemic mediation* embraces a large number of processes and representations, of which inference, short-term memory, problem-solving, the retrieving of information from permanent memory, and the comparing of current and past inputs are among the more popular. It follows from these few remarks that visual processing theories are based on a version of realism that bears the epithet "indirect." The term *realism* identifies a belief in an objective world, detached from ourselves, that can be perceptually experienced. And the term *indirect* identifies the belief that our experience of that world is second-hand: Between the world and our experience of it, there intercedes a representation, or surrogate, of the world (see Shaw & Bransford, in press).

Let us consider two such surrogates, two links in the chain of epistemic mediation that characterizes a typical visual-information-processing account of the mapping from proximal stimulus to percept, namely, iconic and schematic memory. These memories or representations are adjudged to be similar in that they relate to the *visual* appearance of the stimulus (the point being that if the stimulus were a letter, neither representation would be a verbal recoding); but they are adjudged to be quite dissimilar in several, nontrivial respects. First, and perhaps foremost, the icon is *visible*, but the schematic representation is *not visible*. Second, the icon is described as "literal" or "nonsymbolic," while the schematic memory is described as "abstract." Third, the upper limits of iconic persistence are thought to be considerably less than those of schematic persistence. Fourth, the icon is a maskable representation —its persistence or quality is markedly affected by temporally proximate visual stimuli —while the schematic representation seems to be impervious to masking. Fifth, the persistence function of the icon does not appear to be affected by the complexity of the stimulus, although that of schematic memory does. Sixth, the iconic representation seems to be tied to the original coordinates of the stimulus, while the schematic representation does not. And seventh, iconic persistence is apparently indifferent to restrictions on processing capacity, while schematic memory, in its early stages, may be quite sensitive. For relevant references and discussion on most of these points, see Turvey (in press-a).

One may regard the present paper as simply an exploration of the concepts of iconic and schematic memory and the visual-information-processing philosophy of which they are a part. More precisely, however, the paper seeks to determine whether these concepts can play the roles nature requires of them. It will be shown both that the concepts are wanting and that there is reason to question the orientation of indirect realism that is the rationale for their existence.

Retinal Image Theory: The Simulative Assumption and the Image as an Anatomical or Ordinal Arrangement

Much of what is attributed to the icon—for example, that it is precategorical, passive, and impartial to the cognitive operations that follow in its wake—is due in very large part to the close (and unavoidable) relationship between the concept of the icon and the concept of the retinal image.

Perennially, the retinal image has been likened to a miniature replication or simulation of that part of the world at which the eye is directed. It has been understood, as a general tenet, that the simulation is imperfect, notably in that the retinal image has one less dimension than the world. The retinal image is only a picture.

The original impetus for the simulative assumption (Boring, 1942) and the source of its continued nourishment is the belief from indirect realism that perception is not actually of the world but of something happening in the perceiver (Lombardo, 1973). The homunculus of bygone days examined his personal copy of the world identified as the retinal image or a *simulation of the retinal image.* Thus, if the retinal image was a picture, then the simulation was a *phenomenal* picture. Echoing this precedent, the processor in current visual-information-processing schemes does not process the world, but neither does "he" process the retinal image—he processes the icon (Dick, 1974; Neisser, 1967). And if the icon is thought (implicitly or explicitly) to be a copy of the retinal image, then it makes little sense to think of the icon as something that is docile or malleable except in very uninteresting ways. After all, no amount of experience or cognitive effort could be expected to modify the retinal image qua retinal image. And further, if the icon simulates the retinal image, then it too must be a bidimensional picture frozen in time—an internalized snapshot that corresponds to the retinal snapshot.

The ageless alternative to the pictorial view is the interpretation of the retinal image as a set of points corresponding to a set of nonintersecting light rays that can differ only in intensity and wavelength. Analogously to the pictorial view, pointillism portrays the image as bidimensional; in addition, it denies the variable of pattern.

Pointillism, of course, sidesteps the simulative assumption with respect to the relation of the retinal image of the environment, but it need not necessarily do so with respect to the relation of the retinal image to the initial internal representation. Recall that Helmholtz, among many others, viewed the initial internal representation as a set of point sensations varying in hue and brightness. In contemporary pointillism, however, reasonably sophisticated processes are said to relate the retinal image and the initial internal representation, the icon. Thus, "preattentive processes" map the points into an array of segregated entities (Neisser, 1967), or feature-detection processes map the points into feature sets (Haber, 1971). Nevertheless, it is fair to claim that pointillistically motivated conceptions of the icon share the paraphernalia of the simulative assumption.

Of further and even greater significance to the theory of the icon are the following two conceptions: the retinal image as an anatomical arrangement and the retinal image as an ordinal arrangement (cf. Gibson, 1950). The question here, in essence, is whether the light at the eye is to be described in the coordinates of the retina or in the coordinates of the environment.

In the *anatomical* conception of the image, the position and movements of a point or of a pattern are with reference to the mosaic of receptor elements, that is, to the anatomy of the retina. The pattern of light at the eye is thought of as *an arrangement of receptors,* and in this view, the image is said to move relative to the retina.

In the *ordinal* conception, position and movement are defined relative to the arrangement of the stimulation itself. One way of construing the ordinal image is as *an arrangement of excitations indifferent to the actual set of receptors excited.* But a generalization of this construal, which takes us out of the eye as it were, is more useful. The ordinal "image" is an arrangement of differences in intensity in different directions from the eye (cf. Gibson, 1966). This ordered discontinuity of the light at the eye is a sample of the structure of ambient light, that is, radiant light as modulated by an environment. In this perspective, the ordinal "image" is actually a sample of a much larger "image," the ambient optic array (Gibson, 1966) in which the observer is immersed, which he scans, and through which he moves. From the anatomical-image view, we say that when the eye moves, the image moves relative to the retina; in the perspective of the ordinal view, we say rather that the retina moves relative to the image.

The distinction between these two conceptions, anatomical and ordinal, is not immaterial to visual information-processing theory. As I shall attempt to illustrate, the particular theory of processing one constructs is profoundly determined by one's choice of image. Let us begin with a simple and fairly prosaic example. Consider a stationary point of light in an object-cluttered environment and a simple translational movement of the eyes. The point of light will change its coordinates in the retinal mosaic, and from an anatomical-image point of view, this change could be interpreted as movement of the point. How do we conclude, veridically, that it was the eyes that moved and not the point? The frequently voiced answer takes this general form: The stationary status of the point is perceived by registering the direction and extent of eye movement (information that could be provided by muscle kinesthesis or by knowledge of the efferent commands to the eye muscles, depending on one's predilections) and subtracting the computed movement from, or comparing the computed movement with, the anatomically relative movement of the point. Generalization of this answer leads to the time-honored claim that a description of a point in the coordinates of the environment is computed, with the help of *extra* visual information, from a description of the point in the coordinates of the eye.

The same situation—stationary point, moving eyes—receives a quite different interpretation in the ordinal-image perspective. During movement, the point's relation to the arrangement as a whole remains unaltered; in short, *nonmovement* of the point is specified by an invariant arrangement in the ordinal pattern. Given an information-processing device that operates on the ordinal arrangement, recourse to nonvisual information (i.e., muscle-related information) and to unconscious calculations of the kind described would be superfluous. Additionally, there is the implication that registering the point in environmental coordinates is not a processing step that follows subsequent to the step of registering the point in anatomical coordinates. This is not to imply that only one kind of registration occurs; indeed, both may occur and probably do, but the implication is that they are not related.

The above example is a weak illustration of the distinction between the two conceptions. It constitutes a limiting case of what there is to be processed—a point—and a limiting case of observer motion—simple eye movement. As we step up the complexity and the naturalness of the situation, the distinction becomes more sharp.

Movements of the eyes are only one of a multitude of body activities that result in a change in the light at the retina. For example, walking, running, jumping, and rotating or flexing the head and trunk, singly or in combination, yield changes in the light at the eye that are inordinately more complex than those consequent to a simple eye movement: The latter results in a rigid translation; the former, in asymmetric, topological distortions of exceptional mathematical richness. To elaborate a little on the situation described above, consider again a stationary point, but now permit the observer to locomote and to rotate his head. How might we describe the processing, from an anatomical image view, that determines point stability during such a transform? Again, we may suppose that it would be necessary to determine the resultant motion of the eyeball—of the retina—from the particular body motions responsible for the point's journey within the retinal mosaic. Then, as before, this computed eye motion could be entered into some equation with the anatomically relative movement of the point to resolve the question of whether the point moved in addition to the observer. Now it goes without saying that this computation of eye motion would be considerably more difficult than that for simple eye movement in a stationary head on a stationary body.[1] But the difficulties augment quickly when we consider not a stationary point but a stationary object, which would be projected as a particular anatomical arrangement.

Envision an object located on the plane of

[1] Let me anticipate somewhat the developing argument: To determine the resultant motion of the eye, I must have *unequivocal* information about the movement of the eye relative to the head, the movement, of the head relative to the body, and the movement of the body relative to the environment. That is, for the formulation to work, there must be an unambiguous source of proprioceptive and exproprioceptive (see Lee, in press) information, one that does not need reinterpretation or verification by another source. Since "motor commands" and "motor effects" relate equivocally (see Bernstein, 1967; Greene, 1972; Turvey, in press-b), efference signals and their copies could hardly provide such a source other than in the most trivial of cases; and since articular (joint, muscle) proprioception and vestibular exproprioception are calibrated by vision (Lee, in press), they would comprise a second-rate candidate, at best, for the role of unambiguous source. But if, by implication, vision is the best candidate, then a serious paradox exists in the current formulation.

the ground and yourself as an observer walking in its direction. As you move toward and over the object, you will experience a succession of different anatomical arrangements as a result of the changing projection of that object onto the retina. Again, how might you determine whether this change is due simply to your movement and not to a movement of the object in addition to your own? Computation of the resultant eye motion will not be sufficient for this purpose. Not only will the object-produced pattern be displaced vertically from a lower to a higher region of the retinal mosaic—suggesting movement—but an increase or decrease in the number of involved receptor units is to be expected, owing to dilation and compression in the object's image, as is a change in the actual anatomical pattern owing to perspectival changes. To determine whether these variations are simply the result of your own motions, you will have to determine the shape of the object and then access knowledge about how your rectilinear motion alters the object's projected anatomical arrangement *and*, in addition, knowledge about how the possible motions of the object alter the projected anatomical arrangement. This solution is extremely cumbersome and appears to presuppose too much; obviously, it presupposes knowledge about how the projected anatomical arrangement of *any* object is altered by *any* of its possible motions and by *any* of yours.

A different and far less presumptuous solution to the foregoing puzzle derives from the ordinal conception of the light at the eye. A global transformation of the ordinal arrangement—an expanding optical flow—results from the observer's forward, rectilinear motion and is specific to it (Lishman & Lee, 1973; Warren, 1976); in contrast, a motion of the object results in a specific transformation that is restricted to an isolated part of the total arrangement (Gibson, 1966). Thus, if both you and the object are moving, an ordered discontinuity will exist between the ordinal arrangement as a whole and a region within it. If the expansion rate of the isolated region is greater than that of the entire optical flow, then the object is moving toward you as you move toward it; similarly, if the isolated region is a "contracting" pattern within the total optical-flow pattern, then you are moving toward the object but losing ground; and so on. In sum, for any particular relation between you and the object, there ought to be a specific mathematical discontinuity in the ordinal arrangement of the light at the eye. An information-processing device that is capable of registering the optical velocity field and its derivative properties will have at its disposal an *optical* basis for distinguishing, unequivocally, its own motions from the motions of objects in the environment that surrounds it. Formal steps toward verifying this thesis have already been taken (e.g., Lee, 1974; Nakayama & Loomis, 1974).

There are many examples that one could turn to to illuminate the consequences, for a theory of processing, of choosing one conception of the retinal image over another. It is hoped that for our present purposes, the example chosen will suffice. In the section that follows, I shall seek to demonstrate, in part, that visual experience is sensitive to both the anatomical pattern of cells that are excited (the anatomical image) and the relational pattern of places occupied by excited cells (the ordinal image). I shall move toward the claim that experiential correlates of the anatomical pattern are rare and insignificant to visual processing in natural circumstances.

The Icon in Relation to the Anatomical and Ordinal Arrangements

In terms of the anatomical image, position and movement are in reference to the coordinates of the eye, while in terms of the ordinal image, they are in reference to the coordinates of the environment. Afterimages are an obvious example of localization with reference to retinal coordinates: When the eyes move, an afterimage is not displaced across the retina but remains retinally fixed. A critical question is whether the icon, like an afterimage, is registered in the coordinates of the retina or whether, in contradistinction, it is registered in the coordinates of the environment. Let us adopt the oversimplification that masking is an index of the icon and proceed to respond to this question from the perspective of two ingenious experiments.

There is an often-made claim that masking is very much a matter of target-to-mask proximity in the anatomical image. However, with the commonplace tachistoscopic exposure, anatomical and ordinal image are confounded. The ordinal arrangement is instantiated by the anatomical arrangement. To separate the two kinds of "images," one needs to break the coincidence of retinal and environmental coordinates. This can be done through the introduction of an eye motion between presentation of target and presentation of mask. One of the experiments to be considered examines masking with saccadic movement (Davidson, Fox, & Dick, 1973), and the other examines masking with pursuit movement (White, 1976). Let us consider the saccadic case first.

A horizontal array of five letters is exposed briefly just prior to a saccadic movement that carries the eyes a distance of two letters to the left of fixation. Immediately upon termination of eye movement, a mask is presented in the environmental coordinates of the fifth letter but, owing to the eye movement, in the retinal coordinates of the third letter. The query of interest is which letter will be masked. The answer is provocative: The perception of the third letter is impeded although the mask is never seen spatially coincident with the letter it masks; the mask is seen to occur in its proper environmental position and, further, appears to cover one letter while occluding the perception of another. The implication is that there is a correlate of stimulation that moves when the eyes move and a correlate of stimulation that does not. The former correlate must be of the anatomical pattern, while the latter correlate, one may venture to claim, is of the ordinal pattern. May we take this experiment, therefore, as providing evidence of a sensitivity to both arrangements and, further, as providing evidence to suggest that these two sensitivities might be concurrent? In this respect, note that if one has an afterimage, casting the eyes about causes the afterimage to vary its location with reference to a perceptually stable environment.

The above observation entices one to argue as follows: The correlate of the anatomical pattern is the correlate most susceptible to masking, and since we have identified the icon as *the* maskable correlate of stimulation, then we can conclude that the icon is a record of the anatomical pattern of cells that are excited and, *ipso facto*, is localized with reference to the coordinates of the eye. By this argument, the source of iconic persistence emerges as a patch of excitation in the anatomical pattern, that is, excitation that is fixed in the mosaic of receptor cells and not displaced within the eye when the eye moves. This thesis is buttressed by an experiment with a rod monochromat which strongly suggested that the source of iconic persistence was localized primarily in the (rod) photoreceptors (Sakitt, 1975). White target letters were briefly superimposed on an intensely illuminated white background such that no matter how bright the target letters, they could not be seen. However, when the rod monochromat closed her eyes shortly after the target exposure, she experienced a visible and persisting record of the letters.

In Sakitt's interpretation, the background field saturates the rods so that any increments in intensity, such as would result from the target exposure, are indistinguishable. But since the letters are eventually distinguishable, the source of their persistence must be localized *prior* to the first stage of the visual system that saturates. The weight of the evidence suggests that the first stage that saturates is most likely the rod photocurrent, and the source of the visible icon is adjudged, therefore, to be in the photoreceptors. When the rod monochromat closes her eyes, the rods start to recover and the photocurrents dip below their saturation level. With the advent of this state, those photoreceptor locations that were more strongly stimulated, that is, those coincident with the letter display, induce a larger neural signal than the surrounding photoreceptor locations that were stimulated only by the background field. It is this "larger neural signal" that is said to give rise to the visible icon subsequent to closing of the eyes.

One might inquire whether this conclusion, that the source of iconic persistence is retinal, is contradicted by the fact that iconic persistence can be altered by dichoptically opposed stimulation (e.g., Meyer, Lawson, & Cohen 1975). We can answer confidently that it is not. The persistence in the firing of photoreceptors means that neuroanatomical channels allied to the receptor surface will be active for comparable periods. But if neuroanatomically more central components of these channels are perturbed in some fashion—such as by cross-adaptation—they may not be able to match the persistence of photoreceptor activity, and the visual experience, for which they are partly responsible, will be affected similarly.

Curiously, the conclusion that the icon is the correlate of the anatomical image is not supported by the second of the two experiments, which examines masking during *pursuit* movement. For when a target and an aftercoming mask are delivered to an observer who is visually tracking a moving dot, masking is maximal when the target and mask share the same environmental coordinates and minimal when they share the same retinal coordinates (White, 1976). Mimicking the line of reasoning used above, one might now be compelled to argue, contrarily, that the icon is a correlate of the ordinal arrangement.

To summarize briefly, if we choose to define as iconic that correlate of stimulation that is maskable, then there is not one style of iconic persistence but two—one that is in the coordinates of the retina and one that is in the coordinates of the environment. Official doctrine informs us that a description of a stimulus in the coordinates of the environment is mediated by a description in retinal coordinates. This is an assertion of the primacy of the anatomical arrangement in visual processing, and in part, the remarks in the previous section were meant to illuminate this sentiment. Now, given the two contrasting "icons," we should suppose, therefore, on the official doctrine, that the icon in environmental coordinates is computed from that in retinal coordinates. But the experiments of Davidson et al. (1973) and White (1976) offer little support for the orthodox view; indeed, they are more suggestive of a dissociation rather than a dependency. In the pursuit-movement experiment (White, 1976), masking in environmental coordinates occurred at target-to-mask intervals of 50 and 100 msec, intervals at which masking in retinal coordinates was absent. Unfortunately, mask delays of less than 50 msec were not assayed, but the masking functions reported suggest that an even finer time titration would not insure masking in retinal coordinates but would continue to insure masking in environmental coordinates. If true, this extrapolation is paradoxical given the traditional thesis that an object's environmental position is computed over time (with the aid of eye-movement information) from the object's position in the retinal mosaic.

Tentatively, we may hypothesize that the correlate of stimulation in environmental coordinates does not depend on that in retinal coordinates, although the two may occasionally interact. In the context of the previous section, the hypothesis takes this form: Processing of the anatomical arrangement and processing of the ordinal arrangement are largely separate activities.

But let us take a further, guarded step. Let us conjecture that in visual processing *au naturel*, the ordinal arrangement is what is processed and the anatomical arrangement is ignored. Under certain conditions, the experience of the anatomical arrangement may obtrude on the experience of the ordinal arrangement, but such obstrusion is rare. Generally, the anatomical image is *transparent* (in the literal sense of the word) to the information that is processed.

"Transparency" is the key to understanding masking in retinal coordinates. A saccadic movement per se does not erase the anatomical pattern of a previous fixation (Doerflein & Dick, Note 1). Erasure results from the new anatomical arrangement that is consequent to the saccade, superimposed on the old. With saccadic movements, the successive anatomical arrangements are discrete and separate from each other. On the other hand, the ordinal arrangement subsequent to a saccade is neither entirely new nor separate from its predecessor, since any two successive samples have a large overlap, the maximum displacement of an eye never exceeding its angular field. Indeed, almost all the fixations in the field of view of the head will have something in common with their topological neighbors. The degree of overlap is the degree of structure common to the successive sample, the ordinal arrangement that remains unchanged during movements of the eyes. Doubtless, what remains unchanged corresponds to the scene the observer is examining (Gibson, 1966). At all events, we may recognize that the masking in retinal coordinates, the annihilation of an anatomical arrangement that accompanies a saccade, enhances in simple but elegant fashion the transparency of the anatomical image. This would be to the benefit of a perceptual system that processes the mathematical order of stimulation, ideally unencumbered by its anatomical manifestation.

In some such way as the preceding argument, we may rationalize the retinally related masking that is incident to saccadic movement. But how is the environmentally related masking that is incident to visual tracking to be understood? White (1976) has remarked that since one's "goal" in pursuit movement is to maintain a fixation on an object, the perceptual suppression of nontarget, background objects is salutary. Thus in pursuit movement, masking is functionally more useful in environmental than in retinal coordinates. However, it is very difficult to conceive such masking ever occurring in more natural circumstances. Abrupt discontinuous changes at an environmental site are most improbable, and evolution would not have considered it an occurrence worthy of special mechanisms. It would seem that the rationale for environmentally related masking will have to be found elsewhere.

The Discrete, Static Sample Assumption

Iconic and Schematic Storage in Relation to Event Perception

I have referred to both the correlate of the anatomical arrangement and the correlate of the ordinal arrangement as iconic storage. At this time, it is both advisable and reasonable to identify the correlate of the ordinal arrangement as schematic memory and thus avoid an unnecessary proliferation of icons. Although it was suggested earlier that schematic memory is essentially nonmaskable and, therefore, unlike the icon, I choose now to qualify this assertion by remarking that schematic memory may be maskable under certain conditions—namely, pursuit movement. There is much to encourage this definitional alteration and consequent relabeling: most notably, the congruence of the overall distinction between iconic and schematic memory, commented on earlier, and the distinction developed over the last few pages between anatomical and ordinal sensitivity.

But focusing on the iconic/schematic distinction should not blind us to the shared characteristic of the two concepts: Both are conceived as static "snaphots." This commonality bears significantly on the question I shall pursue next: Given the concepts of iconic and schematic storage, how might we account for the perception of events, that is to say, the perception of change wrought over an object or complex of objects?

To set the stage for the critical issues that follow, consider a relatively mundane event such as the rotation of an object. Obviously, rotation occurs over time and, equally obviously, the optical information needed to determine that the object is rotating is temporally extensive. Let us refer to this flow of optical information as the transforming optic array (Gibson, 1966) and contrast it with the "information flow" that we have been concerned with primarily in visual-information-processing theory, which might be described as the flow of "neural information" or the transforming "neural array" (Neisser, Note 2).

Herein lies, as far as one can discern, the most significant feature of visual information processing as a philosophy and methodology: *The analysis of visual perception (or cognition) into discrete temporal cross sections perpendicular to the flow of optical information.* Essentially, the thrust of research and theory has been to describe the vicissitudes of retinal input, that is, the sequence of transformations that the retinally related neural array undergoes, *within* a discrete temporal cross section. The *in theory* argument is that a combination of these discrete temporal cross sections will yield the registration of an event, the optical information for which is realized over time. Considerable lip service has been paid to the eminent plausibility of this *in theory* argument, but curiously, the content of the argument has rarely, if ever, seen the light of day. Even worse, the enormously influential assumption of discrete sampling—and we may remark that it is, indeed, *the* overarching assumption of information-processing theory—has undergone very little serious scrutiny.

Wherein lies the motivation for the discrete-sampling assumption? Once again, we may identify the culprit of the piece as the retinal image construed as an anatomical pattern. On this point, a further, though brief, historical detour will serve us well.

At the onset of the 17th century, Kepler had been convinced by Della Porta of the appropriateness of the *camera obscura* as a metaphor for the eye, a metaphor that had been proposed originally by Alhazan and reinforced by Da Vinci (Lombardo, 1973). This conviction paved the way for Kepler's explanation of how the light rays are focused to produce a two-dimensional image on the surface of the retina. The camera metaphor had served a useful didactic purpose; however, it may be argued that the metaphor was to shackle thinking on visual processing in ways that, in retrospect, far outweighed its original usefulness. To pay heed to the camera metaphor is to liken the retinal photoceptors to photographic film. Consequently, it must be argued that in order to avoid blurrings, the retinal "film" must be exposed to incident light for but a brief moment, just enough to permit the capture of an image. Contemporary theorists, of course, decry the eye-as-a-camera theme; nevertheless, as far as the conventions of visual-processing theory are concerned, the die had been cast many centuries ago. The eye had been characterized as a device for capturing static images, and the visual system, as an instrument for analyzing them. The perception of form or static pattern was defined as the basic area of investigation, and the perception of change of events, was conceptualized as a deduction from sequences of static arrangements.

A commonplace current rationalization for the discrete-sampling assumption stems from the fact that perceiving an object-cluttered environment is achieved via scanning, where scanning is interpreted as a succession of discrete, static arrangements each separate from its neighbor. Respectfully, however, we should note that the static character of the "snapshot" provided by a fixation is guaranteed only if the observer is stationary and is scanning an unchanging, temporally frozen environment. But the absence of movement in observer and environment is a limiting case, and we will have to inquire, as before, what happens when one considers other, more natural instances of observer/environment interaction. First, however, let me preface these considerations with a perfunctory analysis of the limiting case.

Each fixation in a series of fixations results in a new anatomical image. Given that the dwell time—the duration of a fixation—in natural scanning is on the order of 200 msec, it has seemed most reasonable (to some students of perception) to contend that the perception of an unchanging scene is mediated by a succession of temporal cross-sections of iconic size (cf. Haber, 1971). Is perceiving a scene, then, just a matter of adding together persisting icons? Most obviously it is not, for as we have come to understand, each new anatomical arrangement annihilates the previous one; it follows that this must be the fate of successive icons as correlates of anatomical arrangements. Consequently, we are led to propose that whatever the function operating on the discrete temporal cross section, its variable is more likely to be the schematic rather than the iconic form of memory. In brief, we can think of a schematic memory as derived, during a fixation, from iconic memory or from the fixed ordinal arrangement; in either case, we can then speak of schematic "snapshots" and conjecture that a function on these determines the perception of a static scene.

But let us at this point unfreeze the environment and/or the observer. Consider situations in which one is scanning a dynamically transforming scene while locomoting or observing a rotating object from a fixed position of the eye. Here, again, traditional predilections might invite an analysis of these occurrences into successions of "snapshots" (cf. Neisser, 1967). Unfortunately, there is no obvious or simple mechanism for decomposing the optical flow into a series of discrete patterns such as that heretofore provided by successive fixations. Does not a dynamically transforming scene require such decomposition *within* fixations in order to differentiate the optical flow into static patterns? This appears to be a thorny problem for the theory of visual information processing. With a continously changing optical flow pattern, there is a continuously changing anatomical and ordinal arrangement. And in the absence of a source of discontinuity other than fixations, it would be nonsensical to speak of discrete anatomical or ordinal arrangements, and their respective iconic and schematic correlates, as the informational support for the perception of events. However, it does not stretch the imagination to concoct *other* sources of discontinuity. Here are two. First, discrete samples, corresponding to discrete anatomical arrangements, could be achieved by a mechanism that repeatedly turns the retinal mosaic on (for very brief periods) and off (for comparatively long periods). For didactic reasons, let us entertain such a mechanism, however improbable. Second, discrete static patterns could be achieved by a mechanism that sampled, discontinuously, the optical flow further upstream in the neural flow, precisely, at the level of schematic memory.

At all events, with some mechanism in mind by which the continuous optical flow could be packaged for processing into successive, temporally discrete samples, let us inquire, critically, about the style of a system that takes discrete inputs as intrinsic to its operation.

A time-honored distinction between perception and memory takes this form: The domain of perception is *adjacent* order as given in a simultaneous composite of elements; *successive* order, as given in a temporally distributed collection of elements, falls within the domain of memory, Consequently, to register an event involves, among other things, perceiving an adjacent arrangement and storing its percept, perceiving the next adjacent arrangement and storing *its* percept, and so on. From this perspective, to register an event requires both perception *and* memory. For purposes of illustration let us translate "percept" into "schematic memory" and let us identify the storage medium into which schematic memories are entered as short-term memory qua a medium which preserves short-term *memories*. Here, then, is a common formula for explaining the registration of events that occur over reasonably short periods of time: A succession of snapshots is entered into short-term memory and then analyzed or integrated in some fashion. An inquiry into the role adduced for short-term memory in this formula and also into the form of the proposed analytic or integrative operation will prove instructive.

Consider Figure 1. Decomposition of the optical flow into discrete patterns is assumed at the retinal level, but it could just as well be assumed at the schematic level. The successive snapshots (anatomical arrangements) of the events are ordered on the time line T. The internal representations (schematic memories) of these snapshots are ordered in short-term memory on the time line t, where $t < T$. On inspection of Figure 1, one is struck by the realization that the role attributed to short-term memory is that of compressing the time over which the internal representation of the event is distributed. Ideally, the mechanism of short-term memory makes the successive discrete samples contemporaneous. In its ideal form, this role of short-term memory manifests a most subtle alchemy: the conversion of successive arrangements to an adjacent arrangement, the translation of time into space. If short-term memory mimics such an alchemist, then presumably, perception—which, according to the time-honored distinction, is sensitivity to adjacent order—can take over to detect the event that transpired! Unfortunately, in this account, we have gone from perception to memory and

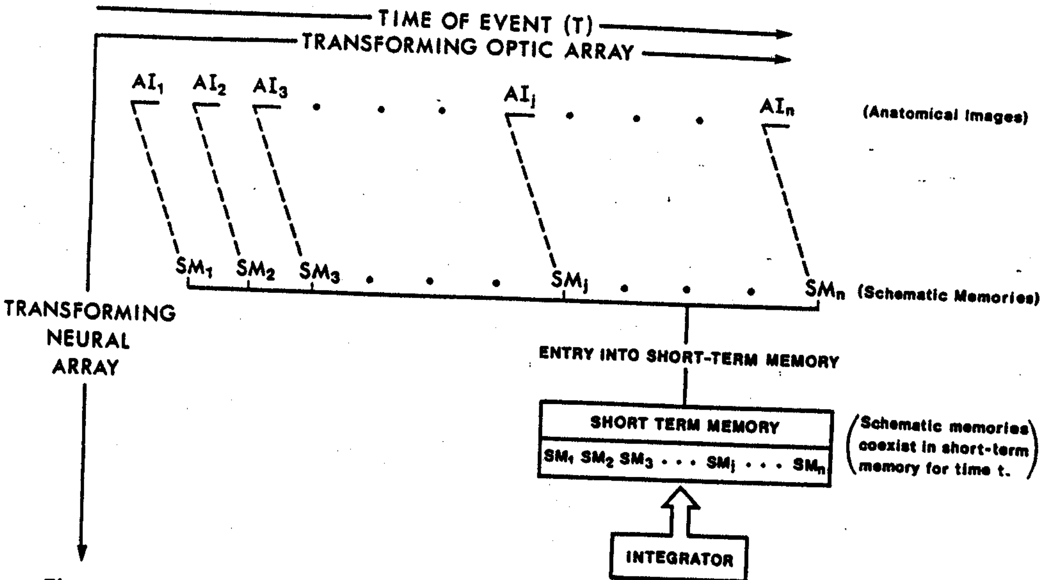

Figure 1. In one traditional view, the continuous optical flow concomitant to an event is decomposed into a sequence of static arrangements—anatomical images—which in the form of schematic memories are brought temporally closer together, even to the point of contemporaneity, by the function of short-term memory.

back again to perception, and the problem of explaining how an event is registered given a successive ordering of adjacent arrangements now becomes the problem of how an event is registered given only adjacent order. Of course, we could assume that the schematic memories are tagged for time-of-entry into short-term memory, but then we would have to suppose some internal process, akin to perception or to inference, that scans the memories according to their temporal order. Needless to say, that is much like the problem we began with, namely, explaining how an event—a change over time—is registered; and it suggests that we have gained little in the way of explanatory yardage from our appeal to the concept of short-term memory.[2]

Schematic Maps and Frame Systems

There is a perfectly good reason for rejecting the notion that the successive snapshots are preserved in short-term memory preparatory to registering an event: More often than not, the number of snapshots would exceed the useful capacity of the memory (Hochberg, 1968; 1970). Therefore, instead of a process that, say, draws inferences from a set of snapshots (schematic memories) in storage, we might wish to propose a set of operations that construct the event from the successive glimpses *as they occur,* and thereby ignore short-term memory altogether.

---

[2] Consider the problem of "slow events," changes that occur over long periods of time (e.g., aging). To explain our apprehension of such things, we might wish to make an appeal to the concept of long-term memory understood (similarly) as a collection of schematic memories. These memories, we may suppose, have been garnered erratically during the slow unfolding of the event. If cognizance of a slow event is interpreted as an inference from a circumscribed collection of schematic memories, then as a preliminary step, we must explain how any one schematic memory of the event comes to be related in long-term memory to any other schematic memory of the event. What relates two or more (indeed, all) of these schematic memories is that they are instances of the same object under transformation. But to establish the identity relation between the schematic memories necessitates knowledge of the transformation (see Shaw & Pittenger, in press). In sum, to relate the schematic memories in a way that makes possible the inference that this or that event transpired, *presupposes knowledge of the event that transpired.*

This has been the route championed by Neisser (1967), Hochberg (1968), and Minsky (1975), although only the latter two have provided anything like a formal account of such a procedure.

The essential idea is this: When one encounters a situation, one selects from one's long-term model of the world a substantial structure called a schematic map (Hochberg, 1968) or a frame (Minsky, 1975), which will be adapted to fit the the situation by changing details as necessary or which will be discarded if a fit cannot be made, in which case another map or frame is then selected. Given the right map or frame, the successive snapshots are fitted together or encoded into the remembered structure. It is this assimilation that determines the perception of a scene or situation. Hochberg defines a schematic map in this fashion: "The program of possible samplings of an extended scene and of contingent expectancies of what will be seen as a result of these samples" (Hochberg, 1968, p. 323). In brief, a schematic map is a matrix of space–time expectancies (Hochberg, 1968). A frame is defined similarly (Minsky, 1975). It is a data structure for representing a stereotyped situation; it is said to contain information about how to use the frame, what can be expected to happen next, and what to do if the expectations are not confirmed. Actually, when speaking of events we should be considering *frame systems* rather than frames, where the different frames of a system capture the object or scene from different perspectives and where change of scene or object is represented within the system by the transformation from one frame to another (Minsky, 1975). And, in like fashion, if we were to address events from Hochberg's stance, we should have to include what he calls schematic sequences, in addition to the schematic maps already described. Crudely speaking, schematic sequences are remembered temporal orders.

Enough has been said to suggest convergence between the terms proposed by Minsky and Hochberg, so further reference to schematic maps and sequences is not needed. We can take refuge in the more neutral and less-abused terms of *frames* and *frame systems*.

A very terse deliberation must suffice to illustrate what is entailed by a frame-systems approach to event perception. At the outset, we should suppose that context supplies a suitably colored backdrop of evidence and expectation for determining the selection of an initial frame system; otherwise, as with any other analysis-by-synthesis scheme, the process of registering the event would never get off the ground. Yet we must appreciate that events, like words, can be perceived with facility in the absence of, and out of, context and that the perception of context itself must rely on frame systems.

What if the initially selected frame system provides less than an adequate fit to the event? Presumably, there must be strategies which guarantee that the next system chosen provides a better fit. But when there is a misfit, there must be a procedure that distinguishes between the object and the form of change (the two aspects of the event instantiated by the frame system) as the source of incongruence. Furthermore, in the course of verifying a selected frame system, that is, while constructing the event, a record must be kept of those snapshots of the event that have been "absorbed" already; otherwise they may be reanalzyed and included as new information (Hochberg, 1970). This latter requirement sounds very much like an appeal to a short-term memory for successive snapshots, a mechanism I had hoped to avoid.

These preliminary remarks on a frame-system approach are testimony to the inelegance and, perhaps, implausibility of a device that perceives events by reference to *prior conceptual knowledge about events*. Let us pass, therefore, from frame systems to a further consideration of the discrete-sampling assumption, and move in a direction of a hypothesis that asserts, contrary to tradition, that the registration of events is not predicated on the registration of static patterns incident to discrete sampling.

Discrete Moment Versus Traveling Moment

Of immediate interest is the distinction between the *discrete moment hypothesis* and the *traveling moment hypothesis*. A simple metaphor, due to Allport (1968), contrasts the two hypotheses rather nicely. Consider a person standing on a railway platform looking into the windows of a passing passenger train for an expected friend. For our observer, the passengers on the train are revealed compartment by compartment as each window passes his or her point of view. In this sense, the glimpses our person has of the interior of the train are essentially discontinuous in time. They are analogous to a series of snapshots in which presence of a feature or object in one snapshot excludes its presence in the next.

But suppose now that we consider the scenario from the perspective of the expected friend *inside* the train. The field of view that he or she has of the platform is bounded by the compartment window; new aspects of the platform are revealed continuously at one side of the window, while other, older aspects are occluded continuously at the other side of the window. The moving window is the metaphor for "traveling moment."

There is a most elegant experiment, also due to Allport (1968), which puts the two hypotheses in competition and from which the traveling moment hypothesis emerges as victor. Suppose that one presents 10 lines in rapid succession on a cathode ray tube, beginning by displaying the line at the bottom of the tube, then turning it off and displaying the line just above it, then turning it off, and so on. The procedure continues until the 10th and top line has been displayed and turned off, at which point the procedure starts afresh with the bottom line. One can adjust the rate of presentation so that all lines are visible simultaneously; a slight reduction of this rate will result in nine lines visible simultaneously, with one missing. The missing line will appear as a shadow that moves gradually across the face of the tube. Now, interestingly, the discrete moment hypothesis predicts that the "shadow" will appear to travel in the direction *opposite* to that of the sequence of line positions, while the traveling moment hypothesis predicts motion of the shadow in the same direction as the sequence of line positions. The two predictions are depicted in Figure 2. It should be remarked, by way of summary, that observers unanimously see the shadow move in the direction predicted by the traveling moment hypothesis. It thus appears that processing is continuous, not intermittent.

But there is yet another and a very remarkable phenomenon that militates against

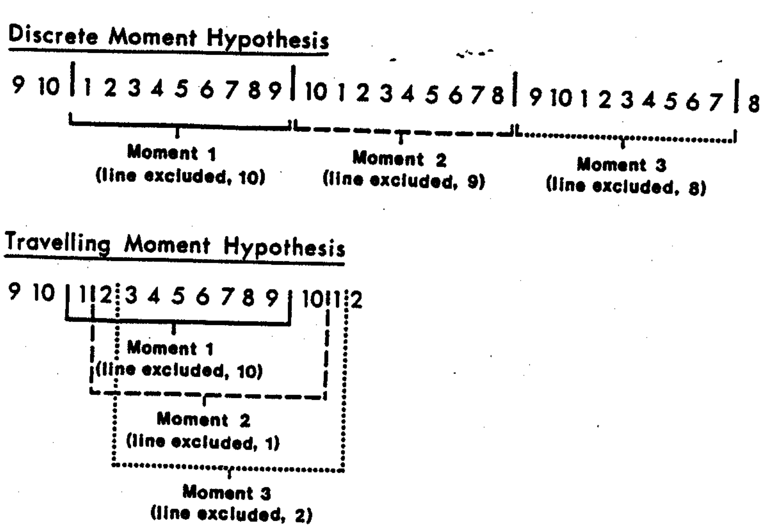

Figure 2. The contrasting predictions of the discrete moment and the traveling moment hypotheses (after Allport, 1968).

the discrete sampling notion.[8] When a cube is rotated on a face, its period of symmetry is four, since every 90° of rotation brings the cube into self-congruence; for rotation on a corner and rotation on an edge, the periods of symmetry are three and two, respectively. If a wire cube rotating at constant speed is strobed, a procedure that simulates discrete "snapshots," then whether or not an observer sees a cube and/or rotation depends on whether or not the strobe frequency is an integral multiple of the extant period of symmetry (Shaw, McIntyre, & Mace, 1974). By inference, if in everyday circumstances, the perceiver samples at an immutable rate, then the nonveridical perception of events should be commonplace; for surely, only on rare occasions would the perceiver's sampling rate be synchronized with the symmetry period of the event. Being able to control the rate at which the optical flow corresponding to an event is discretely sampled would not help either, because, for veridical perception, the perceiver would have to know in advance the nature of the object, the nature of the change, and the period of symmetry. To emphasize the absurdity of the discrete sampling notion, consider the problem of perceiving under natural circumstances two cubes rotating simultaneously, one on a face and one on an edge. In this case, in order to see the two events simultaneously, the observer would have to sample simultaneously at two different rates.

Let me complete this brief survey of phenomena disfavorable to the discrete-sampling assumption with the following observation: Patients with visual agnosia syndromes sometimes manifest a peculiar dissociation between (a) static bidimensional patterns and static three-dimensional objects and (b) these same patterns and objects when either they or the patterns and objects are moving (Botez, 1975; Zappia, Enoch, Stamper, Winkelman, & Gay, 1971). The patients are "blind" to the static objects, but they recognize them under dynamic transformation. If, again, we were to suppose that visual processing proceeds from static samples—procured by a stroboscopic-like operation—then, not unreasonably, we would have to expect these patients to be visually agnostic under any condition of observation, static or dynamic.

In these last two phenomena, the rotating cube and the agnosia, we can dimly see the outline of a potentially significant premise for the theory of visual information processing: Structural variation is not inferred from structural nonvariation, but rather, structural variation *reveals* structural nonvariation.

Limitations of Iconic and Schematic Memory

Our inquiry has brought us, in short, to the conclusion that discrete sampling of a continuous optical flow is not a tenable hypothesis. But the concepts of iconic and schematic representation, as they are most generally construed, are viable only in the context of discontinuous processing. Most obviously, a reappraisal of these concepts and of the visual-information-processing theory to which they are allied is in order.

Most significantly, neither the icon nor schematic memory can be proposed as the object of perception in the sense of providing a data base for visual information processing. And generally, neither iconic nor schematic memory can be said to mediate perception or to be intrinsic to it. These representations have assumed significant roles in our theories because we have limited our experimental analysis to the perception of a frozen scene at a single glance from a stationary point of observation. That which we have labeled iconic storage is essentially the record of the anatomical pattern of excited cells, a record which is only isolable *and persistent* under very special constraints on viewing—for example, tachistoscopic exposure. There is no reason to believe that anything comparable to iconic persistence is manifest when one views a scene undergoing change or when one is in motion. And if iconic persistence is potentially present in the circumstance of

---

[8] I am indebted to Carol Fowler for pointing out the relevance of this phenomenon to the discrete-versus-continuous issue.

scanning a frozen environment from a stationary point of observation, there is no reason to believe that it contributes to the perception of the scene attendant to scanning. What then are we to make of the icon and its properties? The answer might be that the icon and its properties index some aspects of the neuroanatomic structures that support the act of perceiving; they are not, however, indices of processes constituent of that act.

Consider the following metaphor. I wish to understand the working of a telescope—how it magnifies—but my investigation draws me to chromatic aberrations manifested by one of the lenses. Using available techniques, I monitor this characteristic and probe the rules that govern its occurrence. The balance of the metaphor leaves little to the imagination: The object I have chosen to study is a property of the machinery that *supports* magnifying, but it is hardly intrinsic to the process.

But what of schematic memory? It has been defined as a correlate of the ordinal image, which is defined, in turn, as the relational pattern of places occupied by excited cells. The ordinal image incident to a fixation, however, is a limiting case of ordinal arrangement. When the eyes scan a frozen scene, successive fixations as purveyors of ordinal pattern are neither separate nor isolated; on the contrary, they are overlapping and mathematically related. In short, they comprise among themselves an ordinal arrangement. Similarly, and more prominently, in the case of a dynamically transforming environment and/or a moving observer, there is, over time, both change and nonchange in ordinal structure, more precisely, in the ambient optic array. Consequently, if schematic memory is defined as the correlate of the ordinal arrangement, then the static representation of ordinal structure incident to a fixation is simply the limiting case of schematic memory. And insofar as our acquaintance with schematic memory is restricted to the limiting case, we may concur that schematic memory is not something we know much about. We can claim, however, that schematic memory is not a constituent of but a consequence of visual processing, a consequence, that is, of the pickup of invariant and variant structure in the flux of light at the eye.

A Reappraisal of the Visual Processing Assumptions: The Perspective of Direct Realism

Let us conclude with a critical examination of the information-processing enterprise. As we have seen, visual information processing, as a methodology and as a philosophy, is devoted to the solution of the following problem: How is the information in the light entering the ocular system mapped onto perceptual experience?

On the topic of problem solving, Simon (1969) comments that it is essentially a matter of translating, back and forth, between two descriptions of the same phenomenon. On the one hand, there is the *state* description, which characterizes the phenomenon as we sense or intuit it, and on the other hand, there is the *process* description, which characterizes the phenomenon in terms of how we might produce it. As Simon remarks: "The general paradigm is: given a blueprint, to find the corresponding recipe" (p. 112).

A Consideration of the State Description for Haptic Information Processing

The state description of the visual-information-processing problem was identified at the outset, namely, that visual perceptual experience commonly corresponds veridically to the properties of the environment *even though* the proximal stimulus relates imperfectly to those properties. We have said, of the proximal stimulus, that it underdetermines perception. Consequently, the search is for the "recipe," or at least that class of recipes, that will effect this "blueprint," the blueprint of indirect realism. But suppose that we are mistaken in our description—that we have the wrong blueprint—then it matters little what recipes we discover.

The point of departure for these concluding remarks is the conjecture that the state description of the visual-information-processing problem has been inadequately specified, with nontrivial consequences for the understanding of the "how" of processing. Indeed, we will acknowledge the point of view that the most significant determinant of the manner in which we construe perceptual processing to occur is what we conceive the information in the proximal stimulation to be (cf. Garner, 1974; Gibson, 1966). In a phrase, our conception of the *what* of processing determines our investigation and interpretation of the *how* of processing. On this remark, let us return to our earlier discussion of Helmholtz.

Recall that in response to the hypothesized equivocality of proximal stimulation, Helmholtz proposed that a man or woman perceives that particular state of the environment that would normally fit the particular proximal stimulus received. For Helmholtz, knowledge of what is normal, or most probable, is given through experience. In the Helmholtzian position, the sensory mechanisms responding to the light at the eyes are not especially elaborate in that they yield only a patchwork of colors. The task of the perceiver is to recover from this colored, but relatively formless, tapestry the state of the environment. What the perceiver must come to learn, therefore, are procedures which permit him or her to do just that. But how can one acquire such procedures when one's visual relationship with the world is so limited? The visual sensations (or features), after all, are uninformative about the *facts* of the environment. Evidently, knowledge about the facts of the world must be culled from some source other than vision. The tradition of empiricism has looked to mechanical commerce with the world as the means by which the impressions of vision are rationalized: Proximal distributions of light come to be associated with particular tactile/muscle-kinesthetic states.

For example, I adjust my hands, that is, tailor the arrangement of the joints of the fingers, to the shapes of objects. But the immobile clasp of the hands provides a far from perfect mold even for those objects small enough to be grasped; and we do, after all, perceive objects in great detail, whether the objects be large or small. Supposedly, the tactile information about shape comes not from immobile clasps but from dynamic exploration. In like vein, there is the matter of locomoting toward objects and the fact that the amount of effort expended relates in some fashion to the distance traversed. This is perhaps, in part, how one comes to interpret the distance of objects from the size of their retinal images, as Berkeley (1709/1871) argued. Against this notion, however, are many counterpoints. To take but one, the retinal-image size that is contiguous and thus supposedly associated with the degree of kinesthesis and effort expended in locomoting is that retinal-image size which occurs at the *end* of locomotion and not that which was present at the outset. So how can kinesthetic correlates of the distances traversed to reach objects provide a basis for inferring distance from retinal-image size? This particular formula would work only if in our earliest years, we spent most of our time crawling and walking backwards! Undaunted, one could point out that the significance of mechanical commerce follows from the fact that it is intrinsically reinforcing. A hot stove or a lighted match is directly painful to the possessor of the hand that touches it. But in this role, tactile stimulation would tell us only that we liked or disliked some particular optical arrangement; it could not in any obvious fashion specify, say, the *shape* of the distal object to which the particular optical arrangement corresponded.

One might wish to argue that these last remarks do not necessarily allay the claim that tactually exploring and manipulating the environment is the source of knowledge about its properties. For the percept of an object involves more than an impression of its shape and its distance from us; it includes a detailed knowledge of the object's properties such as solidity, stickiness, sharpness, heaviness, coarseness, etc. And we have many intuitive reasons for describing these as nonoptical properties. As Gregory (1969) remarks, "To build a seeing machine we must provide more than an 'eye' and a computer. It must have limbs, or the equivalent, to discover nonoptical properties of objects for its eye's images to take on significance in terms of objects and not merely patterns" (p. 246).

The hypothesis, most obviously, is that mechanical commerce with the environment is privileged in that it permits the direct detection of object properties. In short, mechanical commerce with the world is directly meaningful. But this is a most curious point of view, for it eschews recognition of the fundamental sensationalism of Helmholtzian-type theory. Is it not the case that when an object deforms the skin, or induces a changing configuration of joint postures, all that is directly available is an array of tactile/kinesthetic sensations, or features? (Although as far as I know, features have never been identified for haptic processes.) And, further, is it not true that this tactile/kinesthetic "patchwork" demands interpretation as much as its visual counterpart? We may legitimately assert that empiricism's traditional solution to the problems raised by the nonspecificity and piecemeal nature of the proximal stimulus in vision is no solution at all, because faithful adherence to the traditional perspective results in assigning meaning to meaningless optical patterns by correlating them with meaningless tactile/kinesthetic patterns. Most obviously, the account from empiricism of tactile/kinesthetic perception ought to look just like the account from empiricism of visual perception; in which case neither perception, in theory, ought to occur.

Nevertheless, it is instructive to be delinquent (occasionally) in one's faith. If tactile/kinesthetic processes do register object properties in a fashion that does not have to resort to epistemic mediators, that is, to sources of knowledge other than the stimulation itself—a fashion which might be referred to as "direct"—then the ambiguity or inadequacy of the visual proximal stimulus is not a recalcitrant problem for an experience-based theory of the Helmholtzian kind, though it remains a difficult one. So let us, for pedagogical purposes, grant to tactile/kinesthetic processes this special virtue. Consequently, for mechanical commerce with the environment, the state description of the information-processing problem is as follows: Tactile/kinesthetic or (more appropriately) haptic perceptual experience corresponds to the properties of the environment because those properties determine the pattern of haptic stimulation, which in turn determines the perceptual experience. Let us seek a corresponding process description for this state description.

Mechanical commerce with the environment permits the discovery of object properties, or so we hypothesize. Suppose that I wish to determine the consistency of an object; I will squeeze, prod, push, or stretch the object. If I wish to assess its texture, I will run my fingers—or some suitable extension of my fingers, such as a stick—over its surface; for, interestingly enough, I cannot obtain reliable discriminative information about surface texture by mere contact or pressure (Katz, 1925; Meenes & Zigler, 1923). To determine the mass of an object so that I might compare it with the mass of another, my best policy is to wield it in various ways. Throwing or tossing the object is far more informative about its mass than is passive grasping (Gibson, 1966). From such observations, we can infer the significance of change, that is, of transformations in tactile deformation and patterns of joint articulation, to the determination of object properties.

Obviously, since haptic perception is not epistemically mediated, *ex hypothesi,* there has to be, for any object property, some component of the changing stimulation which remains unchanged and which corresponds *unequivocally* to the property in question. How should we characterize this invariant? In principle, we must characterize it as an often complex mathematical arrangement that is revealed in the course of change. For each property of objects and surfaces detectable by the haptic system, we must suppose that there is a unique mathematical arrangement defined across the concomitant patterns of skin contacts and joint articulations covarying in time.

Roughness presents a limited "instance." The perception of roughness is only relatively slightly affected by a twenty-five-fold increase in the rate at which the fingers move across a surface, and relations among roughness perceptions of different surfaces are unaffected by the force with which the fingers are pressed against them (Lederman, 1974). A possible (quasi-static) candidate for the relevant invariant is the cross-sectional area of the amount by which the skin is depressed below its overall level (Taylor & Lederman, 1975).

At all events, by collecting the immediately preceding paragraphs, we may claim that haptic perception depends on the detection of *information about* something, in the sense of *specificity to* something, that is preserved over exploratory transformation. And thus, one part of the process description that we seek is clearly defined although difficult to implement—the delimiting and manipulating of invariants under relevant transformations. Another part follows.

Pursuing the gist of Helmholtzian-type theory, we might propose that there is a set of elementary properties, each property specified by a particular haptic arrangement, from which all other properties are derived. The distinguishing feature of these elementary properties is that they are the direct givens in the sense that the processes supporting their detection or determination, as already remarked, do not have to resort to sources of knowledge other than the stimulation itself. No epistemic operations (e.g., matching with memory or cross-correlating with the data of other modalities) are said to intercede between stimulation and the registering of these elements; the criteria for their detection are wholly in the stimulation (cf. Uttal, 1975).

Let the elementary properties be *a*, *b*, and *c*. What shall we say of property *d*? We can say that it is derived from *a*, *b*, and *c* and that its detection is by virtue of a *special* process, that is, one different in kind from that by which *a*, *b*, and *c* are detected. Since *d* is not an elementary property, then by definition, the criteria for its detection are not wholly in the stimulation but reside somewhere else, say, in memory. Obviously, the special process of which we speak is an instance of epistemic mediation. But epistemic mediation in haptic perception has been ruled out, *ex hypothesi*, so the detection of *d* cannot entail a special process that is *qualitatively* different from that supporting the detection of *a*, *b*, and *c*. With this line of reasoning, we reach a most curious conclusion: There can be no partitioning of haptic stimulation into primary and directly given properties on the one hand and secondary and derived properties on the other. Invariants specifying environmental properties may lie on a continuum from less to more complex, and they may be detectable with different degrees of facility, but the manner of their detection must be the same—that is, direct. In sum, the process description corresponding to the aforementioned state description of haptic perception is charged with specifying an account of perceptual processing that asserts the primacy of abstract, relational information and the direct pickup of such information. It cannot operate on the nominalistic principle that the perception of some set of elementary properties is a necessary prerequisite to the perception of other, higher order properties.

If we take stock of our analysis of haptic processing, it is evident that we have been describing a version of realism that is more aptly termed *direct*, in contrast to the *indirect* realism that customarily provides the backdrop for the inquiry into visual processing. By direct realism we understand, in part, that it is the world that is actually perceived and not some surrogate of it (see Shaw & Bransford, in press).

But this account of haptic information processing as nonmediated was necessitated by the desire to salvage the empiricist's formulation of visual information processing. That is, if haptic processing could reveal object properties unequivocally, then these products of haptic perception could be associated in experience with visual patterns and thus give the latter meaning. However, if we drop the straw-man stance of empiricism and accept phylogenetic experience as a viable source of knowledge about the world, then perhaps the above description of haptic perception is unwarranted. The significant question is this: Can we imagine a set of information-processing systems, *each operating in Helmholtzian style*, uncovering in the course of evolution those worldly facts and laws necessary to the interpretation of ambiguous and piecemeal sense data? That is to say, can the mediators of perception possibly have a perceptual origin? If they cannot, then we are confronted by an a priori statement of the worst kind: For any species, knowledge about the world was present from the very beginning. This kind of reasoning would seem to characterize the approach to scene analysis by machine (see Sutherland, Note 3). In reference to the question of how to program the perceptual process into a computer, Hunt (1975) remarks that we can do so only by finding a way to represent the world in the machine. In short, for the machine to come to perceive its world, it must already know its world.

One may hazard a guess that in order for a Helmholtzian-designed animal to evolve and relate adaptively to its ecological niche, it would be necessary for at least one perceptual system to acknowledge relevant environmental properties without epistemic mediation. In which case we could argue from a phylogenetic perspective, much as we argued from an ontogenetic perspective, for the direct realism of mechanical commerce with, or haptic processing of, the environment. But then we should ask: If evolution chose to go the direct-realism route with one processing system, why would she not choose to go that way with the others?

The main conjecture at which we have now arrived, in regard to the theory of visual information processing, is that we may look upon direct (but, of course, critical[4]) realism as the departure point for the theory. It is in the light of this conjecture that we take one further and final glimpse at the perception of events.

The Pursuit of Invariants Defined over Continuous Optical Transformations

In our previous consideration of event perception we learned of the nominalistic tenor of orthodox visual-processing theory. Precisely, with an event defined as the change wrought over an object or arrangement of objects, the "information" processed relevant to the resultant experience of the event is that of static bidimensional forms or static perspectives. Accordingly, the experience of the style of change is an abstraction from elementary aspects. What the orthodox view denies is the possibility of abstract relations, defined on optical structure over time, that are specific to the style of change per se. And, indubitably, it is the preclusion of such abstract entities, invariants, that entices—even demands—the interpretation of event perception as a synthesis, either from the memories of static perspectives and intellectual inference or from the assimilation of static perspectives by one of a large number of stored frame systems. But it is exactly such abstract entities and their apprehension without reference to memories, concepts, or any other form of epistemic mediation that is championed in the direct-realism approach to visual processing (see Gibson, 1966; Turvey, 1975).

There is an instructive illustration to be found in the simple change undergone by an object in its motion from one location to another. In an indirect-realism view, this event—translation—is perceived by virtue of the prior discrimination of spatial and temporal extent. The motion parameters of velocity and acceleration are inferred by noting the positional changes of the object and the time over which the changes occur; for acceleration, the spatial and temporal extents between two successive positions would need to be computed on several, separate occasions for there to be any degree of accuracy. The reader will agree that for this formulation to work, it is not sufficient to take discrete samples; rather, the discrete sampling must occur at a constant rate, and the rate must be known. Inasmuch as we have concluded previously that such discrete sampling is implausible, it would be strikingly odd if the evidence were favorable to this indirect-realism formulation. Fortunately, it is not. For it can be shown that the perceptions of velocity and acceleration are not determined by the prior discrimination of dis-

---

[4] In naive realism, it is assumed that all things can be perceived; in critical realism, it is assumed that not all things can be perceived. Direct, but critical, realism means that what can be perceived is perceived directly.

crete spatial and temporal positions but instead appear to be directly perceived attributes of the translatory event (Lappin, Bell, Harm, & Kottas, 1975; Rosenbaum, 1975). In keeping with the expectations inherent in a direct-realism view, the evidence is for abstract information defined on the relation between spatial and temporal changes that is specific to these parameters of translation (see Lappin et al., 1975). And the corollary is no less important, namely, that spatial and temporal positions are not percepual primitives from which are composed the perceptions of velocity and acceleration.

In our reflections on haptic processing as nonmediated, there was a conclusion reached to which we may now return and recognize as a fundamental precept of direct realism. The perception of each and every event or property is based on invariant information detected over time and not on the discrimination of elementary aspects (Gibson, 1966). Now as remarked previously, the two most significant components of events are the nature or style of the change (e.g., rotating, bouncing, running) and the object properties preserved over the change. And it follows, therefore, that for direct realism, there must be two kinds of invariant information in the ambient optic array corresponding to these two components—invariants which might be labeled, respectively, *transformational* and *structural* (Pittenger & Shaw, 1975). A transformational invariant is that information, specific to style of change, that is preserved over different structures "supporting" the change; a structural invariant is that information, specific to object structure, that is preserved over the styles of change in which the object participates.

Herein lies an important and far-reaching contrast with the story of event perception as told in the language of indirect realism. For in that language, there is no equivalent to a transformational invariant, and the concept closest, but not identical to, a structural invariant, is the static silhouette of an object as given by discrete sampling (cf. Pittenger & Shaw, 1975). In an indirect realist's account of event perception, knowledge about the style of change must be presupposed; it resides in memory by sleight of hand. In a direct realist's account, however, information about the kind of change is postulated to exist unequivocally, in principle, in the transforming optic array. We may say that transformational and structural invariants are "formless" (Gibson, 1973), for the crux of the matter is that no "form" remains in a continuous transformation—"form" or "perspective" or "silhouette" is annihilated. What does remain are those aspects of forms or of styles of change that are invariant over time. It is in this sense that invariants are said to be formless, for they are not themselves forms. The summary of these remarks has a curious ring to it: For the direct realist, the perception of events depends on the detection of formless invariants and not on the perception of forms.

Our simplified reconnaissance of visual processing as direct carries us but a little way toward an appreciation of the contrasting perspectives for a theory of visual processing. As significant as this subject is, to pursue it further would take us too far afield for the purposes of the present paper. Others have sought to sharpen and clarify the issues and distinctions I have barely touched upon, and to these authors the reader should turn for a fuller account (e.g., Gibson, 1966; Mace, 1974, in press; Shaw & Bransford, in press; Shaw & McIntyre, 1974; Turvey, 1975).

But there is one aspect of the contrast that we can, and should, pass comment on by way of conclusion. Presumably, the goal of visual-processing theory is to isolate and characterize that which is most eminently and directly responsible for our perceptual knowledge. In the view of indirect realism, the candidates for this honor are patently the postulated links in the internal chain of epistemic mediators from retinal image to perceptual experience. But the view of direct realism promotes a very different roster of candidates. They are, most obviously, the complex, nested relationships in the dynamically structured medium surrounding the observer that are specific to the properties of the environment in which he or she acts. What is prescribed by this latter view is an ecological attitude to research and theory on visual processing; an "ecologizing" of physics, as it were, so that we might better comprehend how the world relates to man and animal as knowing agents rather than as physical or biological objects (Shaw & Bransford, in press). An ecological attitude suggests a rather different tack in research from that which is currently being navigated and a different batch of skills from that at the disposal of contemporary explorers of visual information processing.

In sum, the issue of whether perception is direct or indirect cannot be treated lightly, for the consequences of its resolution will be far from trivial in the quest for an understanding of how we know our world by sight.

Reference Notes

1. Doerflein, R. S., & Dick, A. O. *Eye movements and perceived location in iconic memory.* Paper presented at the meeting of the Psychonomic Society, Boston, 1974.
2. Neisser, U. *On "Going beyond the information given": A reply to J. J. Gibson's memo.* Unpublished manuscript, Cornell University, 1972.
3. Sutherland, N. S. *Intelligent picture processing.* Paper presented at the Conference on the Evolution of the Nervous System and Behavior, Florida State University, Tallahassee, 1973.

References

Allport, D. A. Phenomenal simultaneity and the perceptual moment hypothesis. *British Journal of Psychology,* 1968, *59,* 395–406.

Berkeley, G. An essay toward a new theory of vision. 1709. In A. C. Fraser (Ed.), *The works of George Berkeley* (Vol. 1). Oxford: Clarendon Press, 1871.

Bernstein, N. *The co-ordination and regulation of movements.* New York: Pergamon, 1967.

Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century-Crofts, 1942.

Botez, M. I. Two visual systems in clinical neurology: Readaptive role of the primitive system in visual agnostic patients. *European Neurology,* 1975, *13,* 101–122.

Davidson, M. L., Fox, M. J., & Dick, A. O. Effects of eye movements on backward masking and perceived location. *Perception & Psychophysics,* 1973, *14,* 110–116.

Dick, A. O. Iconic memory and its relation to perceptual processing and other memory mechanisms. *Perception & Psychophysics,* 1974, *16,* 575–596.

Garner, W. R. *The processing of information and structure.* Potomac, Md.: Erlbaum, 1974.

Gibson, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

Gibson, J. J. *The senses considered as perceptual systems.* Boston: Houghton Mifflin, 1966.

Gibson, J. J. On the concept of "formless invariants" in visual perception. *Leonardo,* 1973, *6,* 43–45.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. M. Snell (Eds.), *Progress in theoretical biology* (Vol. 2). New York: Academic Press, 1972.

Gregory, R. L. On how little information controls so much behavior. In C. H. Waddington (Ed.), *Toward a theoretical biology* (Vol. 2). Chicago: Aldine, 1969.

Haber, R. N. Where are the visions in visual perception? In S. Segal (Ed.), *Imagery.* New York: Academic Press, 1971.

Helmholtz, H. von [*Treatise on psychological optics*] (J. P. Southall, Ed. and trans.). Rochester, N.Y.: Optical Society of America, 1925. (Originally published, 1911.)

Hochberg, J. In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception.* New York: Holt, Rinehart & Winston, 1968.

Hochberg, J. Attention, organization and consciousness. In D. I. Mostofsky (Ed.), *Attention: Contemporory theory and analysis.* New York: Appleton-Century-Crofts, 1970.

Hochberg, J. Higher-order stimuli and inter-response coupling in the perception of the visual world. In R. B. MacLeod & H. L. Pick, Jr. (Eds.), *Perception: Essays in honor of J. J. Gibson.* Ithaca, N.Y.: Cornell University Press, 1974.

Hunt, E. *Artificial intelligence.* New York: Academic Press, 1975.

Katz, D. *Der aufbau der tastewelt.* Leipzig: Barth, 1925.

Lappin, J. S., Bell, H. H., Harm, O J., & Kottas, B. On the relation between time and space in the visual discrimination of velocity. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *1,* 383–394.

Lederman, S. J. Tactile roughness of grooved surfaces: The touching process and effects of macro- and microsurface structure. *Perception & Psychophysics,* 1974, *16,* 385–395.

Lee, D. N. Visual information during locomotion. In R. B. MacLeod & H. L. Pick, Jr. (Eds.), *Perception: Essays in honor of J. J. Gibson.* Ithaca, N.Y.: Cornell University Press, 1974.

Lee, D. N. On the functions of vision. In H. Pick & E. Saltzman (Eds.), *Modes of perceiving and processing information.* Hillsdale, N.J.: Erlbaum, in press.

Lishman, J. R., & Lee, D. N. The autonomy of visual kinaesthesis, *Perception,* 1973, *2,* 287–294.

Lombardo, T. *J. J. Gibson's ecological approach to visual perception: Its historical context and development.* Unpublished doctoral dissertation, University of Minnesota, 1973.

Mace, W. M. Ecologically stimulating cognitive psychology: Gibsonian perspectives. In W. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes*. Hillsdale, N.J.: Erlbaum, 1974.

Mace, W. M., James Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, N.J.: Erlbaum, in press.

Meenes, M., & Zigler, M. J. An experimental study of the perceptions roughness and smoothness. *American Journal of Psychology*, 1923, *34*, 542–549.

Meyer, G. E., Lawson, R., & Cohen, W. The effects of orientation-specific adaptation on the duration of short-term visual storage. *Vision Research*, 1975, *15*, 569–572.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw Hill, 1975.

Nakayama, K., & Loomis, J. M. Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception*, 1974, *3*, 63–80.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Pittenger, J. B., & Shaw, R. E. Aging faces as viscal–elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 374–382.

Rosenbaum, D. A. Perception and extrapolation of velocity and acceleration. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 395–403.

Sakitt, B. Locus of short-term visual storage. *Science*, 1975, *190*, 1318–1319.

Shaw, R. E., & Bransford, J. Introduction: Psychological approaches to the problem of knowledge. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*, Hillsdale, N.J.: Erlbaum, in press.

Shaw, R. E., & McIntyre, M. Algoristic foundations to cognitive psychology. In W. Weimer and D. S. Palermo (Eds.), *Cognition and the symbolic processes*. Hillsdale, N.J.: Erlbaum, 1974.

Shaw, R. E., McIntyre, M., & Mace, W. The role of symmetry in event perception. In R. B. MacLeod & H. L. Pick, Jr. (Eds.), *Perception: Essays in honor of J. J. Gibson*. Ithaca, N.Y.: Cornell University Press, 1974.

Simon, H. *The sciences of the artificial*. Boston: MIT Press, 1969.

Taylor, M. M., & Lederman, S. J. Tactile roughness of grooved surfaces: A model and the effect of friction. *Perception & Psychophysics*, 1975, *17*, 23–36.

Turvey, M. T. Perspectives in vision: Conception or perception? In D. Duane & M. Rawson (Eds.), *Reading, perception and language*. Baltimore, Md.: York, 1975.

Turvey. M. T. Visual processing and short-term memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5). Hillsdale, N.J.: Erlbaum, in press. (a)

Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, N.J.: Erlbaum, in press. (b)

Uttal, W. R. *An autocorrelation theory of form detection*. Hillsdale, N.J.: Erlbaum, 1975.

Warren, R. The perception of egomotion. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 448–456.

White, E. W. Visual masking during pursuit eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 469–478.

Zappia, R. J., Enoch, M., Stamper, R., Winkelman, J. Z., & Gay, A. M. The Riddoch phenomenon revealed in non-occipital lobe lesions. *British Journal of Ophthalmology*, 1971, *55*, 416–421.

Received July 21, 1976 ■