5

A Theory of Direct Visual Perception

James J. Gibson

The theory to be outlined is partly developed in *The Senses Considered as Perceptual Systems* (Gibson, 1966), especially in chapters 9–12 on vision. It is related to, although a considerable departure from, the theory presented in *The Perception of the Visual World* (Gibson, 1950). Some of its postulates go back 20 years to that book, but many are new.

What is "direct" visual perception? I argue that the seeing of an environment by an observer existing in that environment is direct in that it is not mediated by visual sensations or sense data. The phenomenal visual world of surfaces, objects, and the ground under one's feet is quite different from the phenomenal visual field of color–patches (Gibson, 1950, Ch. 3). I assert that the latter experience, the array of visual sensations, is not entailed in the former. Direct perception is not based on the having of sensations. The suggestion will be that it is based on the pickup of information.

So far, all theories have assumed that the visual perception of a stable, unbounded, and permanent *world* can only be explained by a process of correcting or compensating for the unstable, bounded, and fleeting sensations coming to the brain from the retinal images. That is to say, all extant theories are sensation-based. But the theory here advanced assumes the existence of stable, unbounded, and permanent stimulus-information in the ambient optic array. And it supposes that the visual system can explore and detect this information. The theory is information-based, not sensation-based.

Perception and Proprioception

Simplifying a distinction made by Sherrington, the term *perception* will be used to refer to any experience of the environment surrounding the body of an animal, and the term *proprioception* for any experience of the body itself (including what Sherrington called *interoception*). Far from being one of the senses, then, proprioception is a kind of experience cognate with perception. Proprioception *accompanies* perception but it is not the same thing as perception.

An awareness of the body, however dim, does in fact seem to go along with an awareness of the world. Conversely, an awareness of the body, however intense, even an experience of pain, is never wholly without some awareness of the environment. And this reciprocity is only to be expected since the very term "environment" implies something that is surrounded, and the term "observer" implies a surrounding world.

The difference between perception and proprioception, then, is one of function, not a difference between the receptors stimulated as Sherrington assumed, that is, the exteroceptors and the proprioceptors. Perception and proprioception both depend on stimulation, but the visual system, for example, can isolate from the flux of stimulation that which is extero-specific (specifies the world) from that which is propriospecific (specifies the body). Vision, in other words, serves not only awareness of the environment but also awareness of self.

For example, the motion of an object relative to the stationary environment can be detected by vision, and this is a case of *perception*. Likewise the motion of one's body relative to the stationary environment, whether active or passive, can be detected by vision, and this is a case of *proprioception*. Locomotion, as distinguished from object motion, is specified by transformation of the ambient optic array as a whole. An observer can ordinarily distinguish the two cases with no difficulty, and so can animals, even species with very simple eyes.

Note that proprioception, as here defined, it not to be confuse with *feedback* in the modern usage of the word, that is, a return input to the nervous system from a motor action. The movements and postures of the body are detected (in several independent ways) whether they are imposed by outside forces or are obtained by an action of the observer himself. Proprioception can be passive or active, just as perception can be passive or active. The above hypothesis is elaborated in Chapter 2 of *The Senses Considered as Perceptual Systems*. The classical doctrine that proprioception is one of the sense modalities is familiar, and is still taught, but it simply will not work. The evidence is against it.

It should already be evident that this theory of perception does not accept the usual analogy between the brain and a computer, and rejects the idea that perception is a matter of processing the information fed into a computer. No one has suggested that a computer has the experience of being "here."

Optical Stimulation and Optical Information

The theory distinguishes between stimulation by light and the information in light. The difference is between light that is seen and the light by which *things* are seen. Light as energy is treated by physical optics. Light as information is treated by an unfamiliar discipline called ecological optics (Gibson, 1961; 1966, Ch. 10). The facts of physical optics are sufficient for a psychophysics of the light sense, and of the elementary visual sensations. But the facts of ecological optics are required for an understanding of direct visual perception.

The relation between optical stimulation and optical information seems to be as follows. The stimulation of photoreceptors by light is a necessary condition for visual perception. The activity of the visual system depends on ambient light; there is no vision in darkness. But *another* necessary condition for visual perception is an *array* of ambient light. It must be structured or differentiated, not homogeneous. With homogeneous ambient light, perception fails although the sensation of light remains. Such is the case in dense fog, empty sky, or in the experiment of wearing plastic diffusing eye-caps, an experiment that we repeat every year at Cornell. In homogeneous darkness, perception fails because stimulation is absent. In homogeneous light, perception fails because stimulus *information* is absent although stimulation is present. We conclude that stimulus energy is a necessary but by no means sufficient condition of stimulus information.

The meaning of the term "information." There are currently two radically different usages of the word "information" in psychology. One I will call *afferent-input information* and the other *optic-array information*. The former is familiar; it is information conceived as impulses in the fibers of the optic nerve. Information is assumed to consist of *signals*, and to be *transmitted* from receptors to the brain. Perception is a process that is supposed to occur *in* the brain, and the only information for perception must therefore consist of neural inputs *to* the brain.

Optic-array information is something entirely different. It is information in light, not in nervous impulses. It involves geometrical projection to a point of observation, not transmission between a sender and a receiver. It is outside the observer and available to him, not inside his head. In my theory, perception is *not* supposed to occur in the brain but to arise in the retino-neuro-muscular system as an activity of the whole system. The information does not consist of signals to be interpreted but of structural invariants which need only be attended to.

It has long been assumed by empiricists that the only information for perception was "sensory" information. But this assumption can mean different things. If it means that the information for perception must come through the senses and not through extrasensory intuition, this is the doctrine of John Locke, and I agree with it, as most of us would agree with it. But the assumption might mean (and has been taken to mean) that the information for perception must come over the sensory nerves. This is a different doctrine, that of Johannes Müller, and with this we need *not* agree. To assume that visual information comes through the visual sense is not to assume that it comes over the optic nerve, for a sense may be considered as an active system with a capacity to extract information from obtained stimulation. The visual system in fact does this. Retinal inputs lead to ocular adjustments, and then to altered retinal inputs, and so on. It is an exploratory, circular process, not a one-way delivery of messages to the brain. This hypothesis is elaborated in Chapters 2 and 3 of *The Senses Considered as Perceptual Systems*.

The Main Principles of Ecological Optics

The term *ecological optics* was introduced in a paper (Gibson, 1961) and the subject was further developed in a chapter on environmental information (Gibson, 1966, Ch. 10). But the concepts and postulates are not yet wholly established, and what follows must be regarded as tentative.

Ecological optics attempts to escape the reductionism of physical and geometrical optics. It introduces a new concept, *ambient light*, which goes beyond the physicist's conception of radiant light, and it postulates a notion of space-filling illumination that extends the classical meaning of illuminance.

The Unlimited Reflecting of Light Waves

In a medium of water or air, in which animals live and move and have evolved, light not only propagates as it does in empty space but also reverberates. It is rapidly reflected back and forth between earth and sky, and also between the facing surfaces of semi-enclosed spaces. Given the speed of light and the fact of sunlight, it almost instantly reaches an equilibrium in the medium, that is, a steady state. The light moves in all directions at once. This steady state of multiply reflected light has very interesting properties. First, at every point in the medium there is ambient light and, second, the ambient light at every such point will be structured by the reflecting surfaces facing that point.

Projection to a Point

At any point in a medium there will exist a bundle of *visual solid angles* corresponding to components or parts of the illuminated environment. The *faces* and *facets* of reflecting surfaces are such components; what we call *objects* are others; and the *patches of pigment* on a flat surface are still others. Note that the bundle of *solid angles* postulated above is not the same as a pencil of rays, which is concept of *geometrical* optics. The cross section of a solid angle always has a "form," no matter how small, whereas the cross section of a ray is a formless point. And the cross section of a *bundle* of solid angles always has a pattern whereas the cross section of a pencil of rays does not.

The Ambient Optic Array

A bundle of visual solid angles at a point (a point of observation) is called an *ambient optic array*. Such an array is invariant under changes in the illumination from noon to sunset. It is an arrangement of components, not an assemblage of points, and the components are nested within others of larger size. It is analyzed by topology or perspective geometry, not by analytic geometry. The array can be said to exist at a point of observation whether or not an eye is stationed at that point. In this respect the array is quite unlike a retinal image, which occurs only if a chambered (vertebrate) eye is put there and aimed in a certain direction. The array is also unlike an image inasmuch as the image is usually said to be an assemblage of focus points each corresponding to a luminous radiating point (presumably an atom) in the environment.

Projected Surfaces and Occluded Surfaces at a Point of Observation

Given that surfaces are in general *opaque*, not transparent, some of the surfaces of the world will be hidden at a given point of observation (occluded) and the remainder will be unhidden (projected at the point). This holds for any layout of surfaces other than a flat plane unobscured to its horizon. But any hidden surface may become unhidden by a change of the point of observation. The occlusion of one surface by another entails an *occluding edge*.

Connected Sets of Observation Points

A path of locomotion in ecological space consists of a connected set of observation points. To each connected set of observation points there corresponds a unique family of perspective transformations in the ambient optic array. In short the changing optic array at a moving point of observation specifies the movement of the point (i.e., the path of locomotion of the observer).

The optical transition between what I call two "vistas" of the world (as when an observer goes from one room to another) entails the progressive occlusion of some parts of the world and the disocclusion of others. The transition, however, arises from a path of locomotion which is reversible, and the transition is itself reversible. What went out of sight in going comes back into sight on returning. This reversible optical transition is to be distinguished from an *irreversible* transition such as occurs when an object is melted or dissolved or destroyed. The study of the two different ways in which an object can go out of sight, by being hidden or by being destroyed, suggests that they are clearly distinguishable on the basis of optical information.

The Family of Perspectives for an Object

Given an illuminated object with several faces (a polyhedron for example) it will be surrounded by an unlimited set of points of observation. Each *perspective* of the object (its projection in each optic array) is unique at each point of observation. The family of perspectives is unique to the object. An observer who walked around the object (looked at it "from all sides") would obtain the whole family.

The features of the object that make it different from other objects have corresponding features in the family of perspectives that are *invariant* under perspective transformations. These invariants constitute information about the object. Although the observer gets a different form-sensation at each moment of his tour, his information-based perception can be of the same object. This hypothesis provides new reasons for realism in epistemology (Gibson, 1967).

Correspondence of Structure between an Ambient Optic Array and the Environment

There is evidently some correspondence between the structure of the environment and the structure of the ambient light at a stationary point of observation. It is by no means a simple correspondence. It is not point-to-point but component-to-component. There are subordinate and superordinate components of the world and corresponding subordinate and superordinate forms in the array, each level of units being nested within larger units. But some components of the environment are missing from a frozen array, because of occlusion. All components of the environment, however, could be included in the changing array over time at a moving point of observation.

Invariant Information in an Ambient Optic Array

A list of the *invariants* in an array as the amount of illumination changes, as the type of illumination changes, as the direction of the prevailing illumination changes, and (above all) as the point of observation changes cannot yet be drawn up with any assurance. But a few facts seem to be clear. The *contours* in an array are invariant with most of the changes in illumination. The *textures* of an array are reliably invariant with change of observation-point. The property of a contour being *closed* or *unclosed* is always invariant. The *form* of a closed contour in the array is independent of lighting but highly variant with change of observation point. A great many properties of the array are *lawfully* or *regularly* variant with change of observation point, and this means that in each case a property defined by the law is *invariant*.

Summary

Eight main principles of ecological optics have been outlined. They are perhaps enough to show that the new optics is not just an application of the accepted laws of physical and geometrical optics, inasmuch as different laws emerge at the new level. And it should now be clear why ecological optics is required for a theory of direct visual perception instead of what is taught in the physics textbooks.

The Sampling Process in Visual Perception

The theories of sensation-based perception presuppose the formation of a retinal image and the transmission of it to the brain. The theory of direct perception presupposes the sampling of the ambient array by the ocular system. What is this sampling process?

No animal has wholly panoramic vision (although some approximate to having it) and therefore no animal can perceive the whole environment at once. The successive sampling of the ambient array is carried out by head-movements, the eyes being stabilized to the structure of the array by compensatory eye-movements (see Gibson, 1996, Ch. 12, for an explanation of head-movements and compensatory eye-movements). The point to be noted is that vertebrate animals with chambered eyes must perform *sample-taking* in order to perceive the environment. Invertebrates with compound eyes probably do the same, although very little is known about visual perception in arthropods. The sampling of the optical environment is a more general process than the fixating of details. The latter arises in evolution only when the eyes develop concentrated foveas.

Along with the taking of stabilized samples of the spherical array there goes a process of optimizing the pickup of information in the sample. Accommodation of the lens, the centering of the retinal fovea on an item of the sample, and the adjustment of the pupil for an optimal level of intensity, together with the adaptation of the retina, are all cases of the adjustment of the ocular system to the requirements of clear vision.

From the earliest stage of evolution, therefore, vision has been a process of exploration in time, not a photographic process of image registration and image transmission. We have been misled about vision by the analogy between eye and camera. Physical optics, and the physiological optics that depends on it, do not now conceive the eye in any way except as a camera. But a camera is not a device with which one can perceive the whole environment by means of sampling, whereas an eye does perceive the environment by sampling it.

If the visual system is exploratory we can assume that it extracts the information in successive samples; we do not have to speculate about how the brain could "store" the sequence of images transmitted to it and combine them into a total image of the world. The experience of the visual world is not compounded of a series of visual fields; no one is aware of the *sequence* but only of the total *scene*. Presumably this is because the ocular system detects the invariants over time that specify the scene.

I once assumed (Gibson, 1950) that the only way one can be aware of the environment behind one's back is to remember it, in the sense of having *a memory image* of it. Similarly, I supposed that, when I look out of the window, my lawn, only part of which is projected through the window to my eyes, must be filled out by images of the remainder. But I no longer believe this theory. Awareness of the room behind my back and the lawn outside my window cannot depend on imagery. I doubt if it depends on *memory*. I apprehend part of the room as *occluded by my head*, and part of the lawn as *occluded by the edges of the window*. And the perception of occlusion, it seems to me, entails the perception of *something* which is *occluded*.

A memory image of a room or of a lawn is something quite different from the perception of surfaces that are temporarily hidden from sight. I can summon up a memory image of the house and the lawn where I lived as a child. This is not at all like the awareness I have of the room behind my back and the lawn outside my window. The theory of information-based perception differs from the theory of sensation-based perception in many ways but in none more radical than this: it does not require the assumption that memories of the past must somehow be mixed with sensations of the present.

The False Problem of Depth Perception and the True Problem of Environment Perception

For centuries, the problem of space perception has been stated as the puzzle of how "depth" or the "third dimension" could be seen when the sensations for the perception were depthless or two-dimensional. Three kinds of solution have been offered, one by nativism (intuition), one by empiricism (past experience), and a third by Gestalt theory (spontaneous orgainzation in the brain). But none of them has been convincing. In the light of the present theory the puzzle of depth perception is insoluble because the problem is false; we perceive the layout of the environment, not the third dimension of space. There is nothing special about "depth" in the environment. As Merleau-Ponty somewhere pointed out, "depth is nothing but breadth seen from the side." We have been misled by taking the third dimension of the Cartesian coordinate system to be a phenomenal fact of perception. And if the flat patchwork of visual sensation is not the basis of visual perception in any case, a third dimension does not *have* to be added to the two dimensions they already possess.

Perception of the *environment* differs from a perception of *space*. An environment implies points of observation in the medium, whereas a space does not. The points of geometrical space are abstract fictions, whereas the points of observation in an environment are the positions where an observer might be stationed. Perception of the environment is thus accompanied by an awareness of the perceiver's existence in the environment (and this is what I call proprioception) whereas a perception of space in its purest form need not be accompanied by any awareness of the thinker's existence in that space.

Geometrical optics is based, of course, on geometrical space. This is everywhere transparent, and it is composed of ghostly points, lines, and planes. It is impersonal and lifeless. Ecologocal optics is based on a space of solid opaque surfaces with a transparent medium in which living animals get about, and which permits the reverberation of reflected light. The surfaces are textured and pigmented. They are flat or curved. They have corners and occluding edges. There are objects and the interspaces between objects. In short, the environment has a layout.

The so-called *cues* for the perception of depth are not the same as the *information* for the perception of layout. The former are called *signs* or *indicators* of depth, or *clues* for an inference that depth exists in the world. Their meaning has to be learned by association. They are sensations in the visual field of the observer, noticeable when he introspects. The latter, the available kinds of information, are *specifiers* of layout, not signs or indicators or clues. They have to be distinguished or discriminated, but their meaning does not have to be learned by association. They are not sense impressions or sense data. When the information for occlusion of one surface by another is picked up there is no sensation for the occluded surface but it is nevertheless perceived. And the information for the occlusion of one surface by another *is* picked up by vision.

The surface layout of the world is thus perceived *directly* when the information is available and when the cycle of action from retina to brain to eye to retina again is attuned to this information. The information must be *attended to*, of course, and this may depend on the maturation of the system, and on practice in looking, and even on the education or training of attention. But the meanings of an edge, of a falling-off-place, of an obstacle in one's path, or of the solid ground under one's feet are given in the ambient optic array and do not have to be memories of past experience attached to present sense-data, or memories of touching aroused by sensations of seeing.

False Questions in the Perception of the Environment

We have seen that the old question of why the phenomenal environment has depth whereas the retinal images are depthless is a false question. There are other false questions of this same sort. One is the question made famous by Stratton's experiment in 1897, *why is the phenomenal world upright whereas the retinal image is inverted on the retina*? Another, going back at least to Helmholtz, is *why is the phenomenal world stationary when the retinal image continually moves with respect to the retina*? Still another (connected with the fact of sampling) is, *why is the phenomenal environment unbounded when each retinal image is bounded by the margins of the retina*? In another form, this is the question, *why does the phenomenal world seem to persist when the retinal images are impermanent*? The answer to all the above questions is this: we do not *see* our retinal images. We see the environment. The doctrine of Müller that all we can see is our retinal images (or at least all we can ever see *directly*) is quite false. If we saw our retinal images we would perceive two worlds, not one, since there is a separate image of it in each eye.

The False Puzzle of the Constancy of Phenomenal Object

The so-called "constancy" of objects in perception despite changing stimulation and changing sensation has long been considered a puzzle. For the past century, experimenters have studied the perceived size of an object with retinal size variant, the perceived form of an object with retinal form variant, and the perceived surface-color of an object with variation of the intensity and wavelength of the light in the retinal image. There is always some tendency to perceive the "real" size, form, and color of the surface of the object, the amount of constancy depending on experimental conditions. Explanations of this result differ with different theorists but they all begin with one assumption, namely, that the perceived size, form, and color are based on retinal size, form, and color respectively—that the process of perception must *start with* these stimulus variables of the image.

According to the present theory this assumption is mistaken. There is information in the optic array for the size, shape, and color of a surface in a layout of other surfaces. The information is a matter of complex invariant ratios and relations; it is not easy to isolate experimentally. But the size, the form, and the color of the image impressed on the retina, when they are experienced at all, are not relevant to and not commensurable with the dimensions and slant and pigmentation of the surface. If I am right, a whole century of experimental research on the *amount* of constancy obtained by an observer is pointless. Insofar as these laboratory experiments have impoverished the stimulus information for perception they are not relevant to perception.

The Effect on Perception of Impoverishing the Stimulus Information

If perception is a process of operation on the deliverances of sense, it has seemed obvious that one way of investigating the process is to *impoverish* the stimulation, to *minimize* the cues, and observe what happens. Visual perception is supposed to come into its own when the input is reduced. Perception then has more work to do. Experiments with a tachistoscope, or with blurred pictures, or with very faint images on a screen are therefore common in the psychology laboratory.

According to the present theory, however, this is not the best way of investigating the process, for perception is frustrated when the stimulus information is impoverished. If the visual system is not allowed to "hunt" for the external specifying information, all sorts of internal processes begin to occur. They are very interesting processes, worthy of investigation, but they should not be confused with the normal process of perceiving.

The situation is similar when contradictory information in the same display is presented to an observer, "conflicting cues." The ambiguous figures and reversible perspectives that have been so frequently studied are of this sort. Ink blots are a combination of impoverished and inconsistent information. I argue that the *guessing* that goes on in these experiments, the attempt to fill out or complete a perception by supplementing the almost meaningless data, is not indicative of what goes on in ordinary perception. The process does not reach an equilibrium state of *clarity* as it does in ordinary perception. And the achieving of precise awareness is the aim of perception.

Orthodox theories assume that there is always an "objective contribution" to perception (the sensations) and a "subjective contribution" to perception (innate ideas, or memories, or field-forces in the brain), the two contributions being combined in various proportions. I reject this assumption. If unequivocal stimulus information is made available to an observer in an experiment, his perception will be determined by it and by nothing else. When *ambient* stimulus information is available to an observer outside the laboratory he can *select* the information that interests him; he can give attention to one part instead of another, but his perception will be determined by the information he attends to.

When *no* stimulus information is allowed to reach the eyes of an observer, as when the eyes are covered by diffusing plastic caps (which can be made of halved ping-pong balls) he is *deprived* of visual perception, although not of sensation. The subject does not like the situation; it is worse than being blindfolded. The only visual experience is that of "nothing." His perceptual system acts a little like a motor running without a load. If he is not allowed to go to sleep, experiences resembling hallucinations may arise.

Summary and Conclusions

This theory of vision asserts that perception is direct and is not mediated by retinal images transmitted to the brain. Most theories assume that perception of the world is *indirect*, and that all we ever *directly* perceive is our retinal images.

Now it is perfectly true that when an observer looks at a painting, photograph, sculpture, or model, he gets an *indirect* visual perception, a *mediated* experience, an awareness at *second hand*, of whatever is represented. A human artifact of this sort is an *image* in the original meaning of the term. It is a light-reflecting object in its own right but it displays *information* to specify a quite different object (Gibson, 1966, Ch. 11). An image in this straightforward meaning of the term is something to be looked at, and it has to be looked at, of course, with eyes. Thus there can be a direct perception of a man's portrait accompanied by an indirect perception of the man himself.

The fallacy of the standard theories of perception consists of taking as a model for vision the kind of indirect visual perception that uses pictures as substitutes for things. The false analogy should now be evident. Direct perception of a retinal image implies an eye inside the head, in the brain, with which to look at the image. But there is no little man anywhere in the brain who can do this. We do not look at our retinal images and perceive the world in the way that we look at a portrait and perceive the sitter. Putting the objection another way, the so-called image on the retina is not an image at all, properly speaking, since it cannot be looked at, as a picture can be looked at, and cannot therefore mediate perception. The famous experiment of looking at the back of the excised eye of a slaughtered ox and observing an image is profoundly misleading. The eye is a biological device for sampling the information available in an ambient optic array. The vertebrate eye does it in one way and the insect eye does it in another way but both register differences of light in different direction at a point of observation.

The availability of information in ambient light and the possibility that it can be picked up directly have implications for epistemology. They lend sophisticated support to the naive belief that we have direct knowledge of the world around us. They support direct realism (Gibson, 1967). If these hypotheses prove correct, they justify our deep feeling that *the senses can be trusted*. At the same time they explain the seemingly contrary conviction *that the senses cannot be trusted*. For a distinction has been drawn between what might be called the *useful* senses, the perceptual systems, and the *useless* senses, the channels of sensation.

Bibliography

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

Gibson, J. J., Ecological optics. *Vision research*, 1961, 1, 253–262.

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

Gibson, J. J., New reasons for realism. *Synthese*, 1967, 17, 162–172.