MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

A.I. MEMO 574 March 1980

AGAINST DIRECT PERCEPTION

S. Ullman

Short Abstract: Central to contemporary cognitive science is the notion that mental processes involve computations defined over internal representations. This notion stands in sharp contrast with another prevailing view -- the direct theory of perception, whose most prominent proponent has been J.J. Gibson.

The publication of his recent book [*The Ecological Approach to Visual Perception* Boston, Houghton Mifflin Company, 1979] offers an opportunity to examine the theory of direct perception and to contrast it with the computational/representational view.

In this paper the notion of direct perception is examined primarily from a theoretical standpoint, and various objections are raised against it. An attempt is made to place the theory of direct perception in perspective by embedding it in a more comprehensive framework.

This paper is also to appear in the *The Behavioral and Brain Sciences, 1980*

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for this research was provided in part by National Science Foundation Grant MCS77-07569

© MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1980

Against direct perception

Abstract

Central to contemporary cognitive science is the notion that mental processes involve computations defined over internal representations. This view stands in sharp contrast with the "direct approach" to visual perception and cognition, whose most pronounced proponent has been J.J. Gibson. In the direct theory, perception does not involve computations of any sort, it is the result of the direct pickup of available information.

The publication of Gibson's recent book [Gibson, 1979] offers an opportunity to examine his approach, and more generally, to contrast the theory of direct perception with the computational/representational view.

In the first part of this paper (Sections 2-3) the notion of "direct perception" is examined from a theoretical standpoint, and a number of objections are raised against it. Section 4 is a "case study": the problem of perceiving the three-dimensional shape of moving objects is examined. This problem, which was studied extensively within the immediate perception framework, serves to illustrate some of the inherent shortcomings of that approach. Finally, in Section 5, an attempt is made to place the theory of direct perception in perspective by embedding it in a more comprehensive framework.

Against direct Perception

1. *Introduction*

Gibson's recent book [Gibson, 1979] is his third in thirty years devoted to the development and exposition of the theory of direct perception. The interest in Gibson's influential theory often transcended the interest in perception alone. One reason is that his approach to cognition in general stands in sharp contrast with another prevailing approach, the computational/representational one. According to the latter view, (of which generative grammer, theories in cognitive psychology, and some of the work in artificial intelligence are current examples) mental processes involve computations defined over internal representations. In the direct theory of perception mediating constructs are unnecessary, and in the early stages of his theory Gibson expressed hope that the direct approach, if successful, would extend to other areas of psychology as well:

> [The theory of direct perception] "...if successful, will provide a basis for a stimulus-response psychology, which otherwise seems to be sinking in the swamp of intervening variables" [Gibson, 1960].

In this paper the concept of direct visual perception (abbreviated as DVP) will be examined. The overall plan of the paper is as follows. First, a brief description of the concept will be given. It intends only to state the main points of relevance to the ensuing discussion, not to summarize Gibson's theory. For a comprehensive presentation of the theory in different stages of its evolution see Gibson [1950, 1966, 1979]. These books describe different approaches to direct perception, not all of which (especially the 1950 formulation) are retained in the current formulation of the theory. The notion of DVP is then examined primarily from a theoretical standpoint (for discussions of empirical evidence against direct perception see [Epstein & Park, 1964; Gyr, 1972a,b; Epstein, 1977]). Section 2 examines what it means for perception to be direct, and Section 3 raises general arguments against the plausibility of direct perception. Section 4 is a "case study": the application of the theory to a particular problem, the perception of moving objects, is discussed to highlight some of the inherent shortcomings of the direct approach. Finally, Section 5 tries to put the DVP approach in the perspective of a more comprehensive framework, and to identify some of its missing ingredients.

1.1 *Direct visual perception*

Visual perception and its relation to the structure of the environment is viewed by the theory of direct visual perception as a sequence of two direct and unambiguous mappings: "stimulation is a function of the environment, and perception is a function of stimulation" [Gibson, 1959, p. 459]. The first mapping is between various aspects of the environment and some spatio-temporal patterns of the visual array, sometimes called "higher order stimuli" (the more recent formulations of the theory emphasize the transformations and invariants in these patterns). The second mapping is between stimuli and percepts. When an observer moves in the environment, some aspects of the light array that reaches his eyes change, while others remain unchanged. The information in these transformations and invariances specify the environment: its layout, changes of layout, and the occurence of events therein. The theory emphasizes the first relation, between the environment and the light array. Its branch of "ecological optics" is aimed at describing the information available to be sampled, and the way it specifies the environment. The second relation is established in the current version of the theory by an "immediate pickup of information" that requires no processing of any sort on the part of the perceiver.

2. *What does it mean for perception to be "immediate"?*

The DVP theory contends that the relation between stimuli (or information in the array of light) and percepts is direct and immediate. To evaluate this claim we shall first examine what it means for percepts and stimuli to be "immediately related". More specifically, we shall ask under what conditions the theory of perception can view stimuli and percepts as directly related, and what would be the criteria for abandoning this view in favor of a different kind of relation.

The term "immediate" has several meanings and connotations; in particular, the qualifications for being "immediate" may be relative to the system under investigation. If a system *S* is investigated, then any signal that reaches *S* from the outside can be considered "immediate". For the psychologist, for example, signals of heat or touch produced by peripheral receptors might be thought of as immediate in this sense, since they are external to the system under investigation. For the physiologist, on the other hand, who studies for instance the internal mechanisms of Meissner's corpuscle (a touch receptor), the relation between touch and the receptor output cannot be dismissed as immediate.

In this sense, the term "immediate" does not serve to describe the signal or operation under consideration, but to express a point of view that regards them as lying outside the domain of interest. Viewing the relation between stimuli and percepts as immediate in this sense would imply that regardless of how percepts are actually related to stimuli, we simply hold this relation to be outside the scope of the theory of perception, which is an unlikely position.

Let us accept, therefore, the view that the relation between stimuli and percepts does not lie outside the domain of the theory of perception, and is not immediate in this sense. Describing the stimuli-percepts relation as "immediate" would still be justified if the relation *has no meaningful decompositions into more elementary constituents.* To clarify these notions of "immediate relation" and "meaningful decomposition" let me first discuss them in the context of performing a simple computation, e.g., the addition of two integers. Computations in general can be described in terms of elementary relations together with some schemes for combining them into more complex operations. As the basic operation for the addition of two integers (in decimal notation) one can use an "addition table" that lists the results of adding any two digits between 0 and 9. Together with the appropriate rules for proceeding from right to left, and for handling the carry (which may be either 0 or 1), any two integers can be added. The operation 2 + 7 in this scheme is immediate and amounts to a table lookup. It is a primitive operation that cannot be elaborated or decomposed within the framework described above. The computaion of 312 + 57, on the other hand, is non-immediate, but can be described in terms of the basic rules and operations. In decomposing a complex operation into its more elementary constituents it is required that these constituents be meaningful in the domain of discourse. The basic operations, on the other hand, can be decomposed and elaborated only outside the scope of the theory. If, for example, the above integer addition scheme is implemented in a machine, the lookup table underlying the computation will have some physical realization, using, for instance, electronic components. These components with their associated currents and voltages can be analyzed and described further, but this description would no longer be in terms of algebraic entities and operations.

The adding machine example is intended to illustrate the relevant terms in a simple situation, not to suggest that the perceptual system resembles an artificial electronic device. The brain does not have to resemble an adding machine, however, for the above distinctions to carry over to the domain of psychology. In explaining perceptual processes, the theory of perception will also employ primitive concepts and operations whose explanation lies outside the scope of the theory. Such primitives may ultimately be further elaborated in a different domain, e.g. physiology and anatomy.

In the theory of perceived colors, for example, the spectral absorption functions of the retinal receptors may play a primitive role. Within the theory, certain regions of the light spectrum can be "immediately registered" by the retinal cones. This does not exclude, however, an explanation of these absorption curves, for instance, in molecular terms. Similarly, the theory of perception would be justified in claiming that the shape of an object is "directly picked up" if a further elaboration of this "picking up" operation would only be possible in physiological, but not in psychological terms. If, however, the perception of shape has a meaningful decomposition, if it can be further decomposed and explained in terms of more elementary concepts and operations, than such an explanation would be more satisfactory than the "immediate registration of shape".

Another example of interest where the notion of a "meaningful decomposition" plays an important role concerns the distinction between molar and molecular descriptions. The ideal gas law $PV = NRT$ is an example of a molar description, stating the relation between the pressure, volume, and temperature of N moles of ideal gas. This molar equation can also be derived from more elementary phenomena, but a description in terms of the elementary phenomena would involve a shift from the domain of gas containers, their volume, pressure, temperature etc., to the domain of molecules in random motion. Gibson employed the molar/molecular distinction to argue that "immediate perception" is justified since psychology studies phenomena at a molar level. In describing the movements of an animal, for example, we are interested in a "molar" description, not in the detailed contractions of individual muscles [Gibson, 1960]. Analogously, he argues that on the molar level, stimuli and percepts should be described as immediately related. This claim implies that a meaningful elaboration of the stimulus-percept relation, and the process of information pickup, would require a shift to the molecular level, or, in the case of perception, to the physiological and anatomical level. In other words, the relevance of the molar-molecular analogy hinges on the feasibility of decomposing the relation between stimuli and percepts in psychologically meaningful terms.

This problem lies at the heart of the dispute between the theory of direct perception and the computational/representational approach: if the extraction of visual information *can* be expounded in terms of psychologically meaningful processes and structures, then it cannot be considered immediate. Much of the ensuing discussion will focus on problems pertaining to this controversy. For additional controversies related to direct perception that will not be emphasized here, see [1].

2.1 *A note on direct perception and direct realism*

Discussions of direct perception have often been related to the problem of realism in philosophy. It has been argued [Gibson, 1967; Yolton, 1968-9; Gibson, 1968-9; Metzger, 1972; Henle, 1974; Turvey, 1977] that the DVP theory has significant ramifications for the problem of realism in that it lends new and sophisticated support for direct realism.

Both realism in general and direct realism in particular are claimed to be supported by the theory of direct information pickup. If we are endowed with mechanisms that can directly register aspects of the environment, then such an environment must exist (which is a case for realism), and we have a direct knowledge of it (which is a case for direct realism). A detailed examination of these issues would require too long a digression. I shall therefore make only two brief comments that bear on the issues at hand, one related to realism in general, the second to direct realism.

In voicing his skepticism, the non-realist does not have to deny the self-consistency of the realist's position. The existence of external objects, and of perceptions that reflect them faithfully, is one possible state of affairs. It is not the only conceivable one, however, and the non-realist sees no compelling reason to except it. I see no significantly new argument in the theory of immediate perception that will force the non-realist to abandon his position. As far as the non-realist is concerned, the view that we possess mechanisms that are directly sensitive to patterns and invariances, and that these patterns in turn specify the external reality, is still not the only irrefutable position. The DVP theory is consistent with realism, but does not seem to offer a significantly "new and sophisticated support" for it.

To examine the relation between direct perception and direct realism, it would be useful to distinguish between two notions of directness. The first is the direct awareness of objects, as held by direct realism. The second, which has to do with direct perception, makes a claim about the psychological theory of perception. It implies that the perceptual process has no psychologically meaningful decomposition, in the sense defined in the previous section.

Now it may be argued that direct realism implies that a perceptual theory of the direct kind should be preferred. Even if this argument holds, however, it would mean that direct realism lends support to the theory of direct perception, rather than the other way around. If the psychological theory of direct perception is to lend new support to direct realism, it has to be evaluated on its own, independent of direct realism. This brings us back to the problem raised in the last section, concerning the analysis of the perceptual process in psychologically meaningful terms, and in particular the adequacy of "direct information pickup" as a primitive construct in the theory of perception.

In arguing for direct perception it has often been suggested [Gibson, 1966, 1967, 1972, 1979, p. 54, 60] that the alternative to direct perception is the indirect sense-data theories of the kind advocated by Locke. These sense-data theories view the perception of objects as composed of two stages. First, elementary stimuli such as homogeneous patches of color give rise to elementary sensations (or "ideas", as they where called by Locke) in the mind; then, the perception of objects is derived from composites of elementary sensations. The DVP theory rejects the "mental chemistry" of elementary sensations, and concludes that perceptions of objects and events are the direct result of "higher order" stimuli (or, in a later formulation, the information in the visual array):

> "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." [Gibson, 1972, p. 215].
>
> [The direct theory] ..."is therefore not obliged to postulate any kind of operation on the data of sense, neither a mental operation on units of consciousness nor a central nervous operation on the signals in nerves. Perception is taken to be a process of information pickup." [Gibson, 1967, p. 162].

Gibson argues against theories of perception that rely on the mental chemistry of "units of consciousness". The implication from this argument is that since percepetion cannot be so decomposed, a direct theory of perception is required. But the argument that a Gibsonian theory of direct perception is required simply because the above sensation-based theories are considered untenable suffers the fallacy of "argument by selective refutation". That is, only one of the alternatives to "direct perception", not all of them, is refuted. Association of sensations is not the only conceivable form of mediating perceptual processes. Rejecting the combination of sensations by the mind does not by itself justify, therefore, the conclusion that processes such as inference, interpretation, computation, categorization, assimilation, or stabilization [Gibson, 1959, p. 460], or copying, storing, comparing, and matching [Gibson, 1966, p. 39], have no place in the theory of perception.

3. *Can perception have an "immediate" theory?*

In the preceding section, certain aspects of "directness" in the theory of perception were examined. This discussion will now be applied to the question of the plausibility of direct visual perception. Section 3.1 raises the argument that the richness of stimuli and percepts prevents a satisfactory theory of a direct mapping between them. In Section 3.2 the notion of information pickup by the sense organs and its use as a primitive construct in the theory of perception are examined.

3.1 *The richness of stimuli and percepts*

The DVP theory describes perception in terms of a family of percepts coupled with their specific stimuli. When a stimulus (or even sufficient information) is present, it can be "directly registered" by an appropriate mechanism tuned for its detection, thereby giving rise to a specific percept. The registration of information is a primitive construct, that has no elaboration within the theory. According to this view the perceptual system performs only the most elementary kind of computation (if it can be called computation at all). Direct registration is thus essentially equivalent to a basic "table lookup" operation in the sense that functionally it relies mainly on a single construct whose further elaboration lies outside the scope of the theory. A direct registration is not the only sort of operation available, however, nor is it necessarily the most appropriate one. Some insight into the appropriateness of the "immediate" sort of theory can be gained by considering, in general terms, under what condition one can expect a system to be adequately considered in "immediate" terms, and in what systems would intermediate processes be necessary.

Let us first return to the elementary example of integer addition. We have seen how the addition of any two integers can be based on a restricted lookup table, augmented by the right-to-left processing rule and handling of the carry. Is this mode of addition better than a large-scale table that lists directly the results of adding pairs of integers? The large-scale table has an advantage: it does not require intermediate steps and therefore offers simplicity and possibly speed. The indirect method offers a different advantage: employing only a restricted table it was able to handle an unbounded set of inputs. The question of whether direct pairing or indirect computation is preferable thus depends on the task at hand. The direct approach is advantageous when the set of input-output pairs is small (compared with the capacity of the system), and when speed is of the essence. For example, our inborn repertoire of reflexes can probably be thought of as a pre-wired, immediate coupling between stimuli and responses. When an exhaustive enumeration becomes prohibitive, processes and rules of formation would offer an advantage over the direct coupling of input-output pairs.

The production and recognition of the cricket's song is an elegant biological example of signal production and recognition that can be reasonably thought of as having an immediate nature [Zaretsky, 1971; Bentley & Hoy, 1974]. The cricket song is a train of sound pulses of a fixed temporal pattern. The generation of such a predetermined pattern requires no formation rules, and can be explained directly in terms of the underlying physiology and anatomy. As Bentley and Hoy comment, "the correct pattern arises from the neural connection established during development" (p. 41). Their work is aimed therefore at identifying the neural mechanisms responsible for the song production, and their genetic origin. Similarly, the recognition of the song is carried out directly by a neural "song-responding mechanism" that can "resonate" to the appropriate pulse-sequence [Zaretsky, 1971].

In contrast, the view raised by generative grammer theories is that the production and recognition of grammatical sentences in a natural language does not have an "immediate" theory in this sense. Rules of formation and recognition are incorporated in the system in order to handle the unbounded set of possible sentences. Similarly, if we consider all distinguishable perceptions (such as the perception of all different shapes) as distinct percepts, the number of possible stimuli and percepts becomes too large to succumb to a direct pairing.

To reduce the number of possible percepts one might try to lump them into groups or families. For example, "three-dimensionality" may be suggested as a single percept (Such percepts were suggested e.g. by Wallach and O'Connell [1953] and Braunstein [1962], though not in the context of supporting direct perception.) A percept of three-dimensionality would require a set of parameters associated with it, since we are not only able to distinguish whether an object is flat or three-dimensional, but can also perceive its particular three-dimensional shape. The required associated parameters have still to be retrieved, and therefore the problem of immediate perception is only hidden, not solved, by introducing such percepts as "three dimensionality". A plausible method for dealing effectively with problems that are too large and complex to be handled by direct pairing alone is to employ processes or rules of formation. A system that incorporates such processes is therefore a more likely candidate for coping with the enormously complex tasks of visual perception.

3.2 *The immediate registration of information and object properties*

The basic operation performed by the visual system according to the DVP theory is the registration or detection of information. The information in the ambient light array constitutes the stimulus to the sense organ, which picks it up and thus produces the awareness of objects and events:

> "...there can be direct or immediate awareness of objects and events when the perceptual system resonates so as to pick up information." [Gibson, 1967; p. 168]

All the observer has to do in the process is "to pick up information by looking" [Gibson, 1966; p.3]. The abstract information that the sense organ directly "resonates to" [Gibson, 1966; p. 267] is conveyed primarily in the form of invariants and transformations in the array of light [2]. For example, we correctly perceive the unchanging shape of a rigidly moving object

> "...not because we have formed association between the optical elements, not even because the brain has organized the optical elements, but

because *the retinal mosaic is sensitive to transformations as such.*"
[Gibson, 1957; p. 294, italics added.]

A general question raised by the above description is what sort of stimuli can be registered directly, and what sort of primitive operations can be consigned to the sense organs. Can information, transformations (as in the above paragraph) and invariants [Gibson, 1979; p. 178] be considered the direct stimuli for the visual system, as proposed by the theory of information pickup? Physiology tells us that the retinal receptors register light energy in various regions of the visible spectrum. Gibson raises two arguments for why we can nevertheless accept abstract information, rather then spatio-temporal distribution of light energy, as the direct stimulus for the sense organs. The first argument relies on the distinction between sensation and perception, and the second on the availability of patterns for immediate pickup. I shall consider each in turn.

3.2.1 *Sensation versus perception*

DVP parallels the sensation-based theories of perception in distinguishing between sensation and perception. According to this view physical stimulation by light causes sensations, not perception [Gibson, 1966, 1979]. What gives rise, then, to perceptions? The sensation-based theories suggest that they are produced from collections of sensations. Gibson rejects this idea and concludes that perceptions and sensations are produced along parallel tracks: stimulation at the receptors level gives rise to elementary sensations, while stimulation of the perceptual system by relevant information directly produces percepts of objects and events [Gibson, 1966, 1967].

The above implication (that abstract information constitutes the stimuli for perception) depends on accepting the theory of immediate perception as the only alternative to the sensation-based view of perception. If percepts are indeed directly coupled with stimuli, then these stimuli are necessarily highly complex and abstract. But if direct perception is not admitted, the notion of information as stimulation does not follow. If the possible role of mediating processes is appreciated, then the light distribution at the receptors can be accepted as the input to the visual system. The gap between the physical stimulus and the perception of objects can be bridged, at least in part, not by associating sensations, but by an elaborate process that constructs a representation of the environment on the basis of the incoming light distribution. The key point is not whether the latter view is correct, but that the immediate registration of abstract information is not the only alternative to the sensation-based theories of perception.

To summarize the above point: the argument for abstract stimuli claimed that (a) the sensation-based view is false, and therefore (b) immediate perception and (c) abstract stimuli follow. But the implication is actually that (a) and (b) together imply (c). Hence, the notion of abstract information as the stimulus for perception is implied primarily not by the rejection of the sensation-based view, but by accepting the theory of immediate perception.

3.2.2 *The availablity of patterns for immediate pickup*

A second argument that supports, according to the DVP approach, the existence of abstract stimuli and their registration, is that patterns of light distribution in space and time are directly available to the visual system. As far as I can see, this availability of patterns as stimuli is supported in the direct theory by two arguments: (i) the existence of neural interconnections and (ii) the locomotion of the observer. Neural interconnections create higher-order units in the nervous system that can register spatial patterns directly [Gibson, 1967]. When, in addition, the observer moves about in the environment, the interconnected network of photo-receptors and higher order "resonators" can register the information in the spatio-temporal patterns. Perception is therefore "not supposed to occur in the brain but to arise in the retino-neuro-muscular system as an activity of the whole system" that moves in the environment and resonates to the available information [Gibson, 1972; p. 217].

Let us first clarify the point of contest in this argument. The controversy does not concern the relevance of spatio-temporal patterns to visual perception. It is granted that information about objects is carried by patterns of light distribution and their changes over time. The debate concerns the nature and complexity of the processes that "register" the information in the spatio-temporal patterns. That is, whether the registration of information should be taken as a primitive construct, or should it have an explanation within the theory.

The fact that spatio-temporal patterns of light carry sufficient information for visual perception does not by itself entail, however, the immediate registration of the information in these patterns. It has recently been shown, for example, [Ullman, 1979a; 1979b; Longuet-Higgins & Prazdny, in press] how the rigidity and three-dimensional shape of moving objects can in principle be recovered from their changing images. These results are applicable both to continuous and discrete (movie-like) stimuli, and to perspective as well as parallel projection. For simplicity, let the case of discrete presentation and parallel projection (such as the image of a distant object) serve as an example. As it turns out, the three-dimensional structure of an object containing at least four non-coplanar elements can be recovered completely if it is viewed from three distinct viewing points. This result guarantees that under simple restrictions there is indeed sufficient information in the changing image to specify the rigidity and shape uniquely [3]. The information is encoded in "high order patterns" in the sense that extended patterns in space and time are required. The recovery of the rigidity and correct three-dimensional shape is possible in this scheme, but it is far from immediate, for two main reasons. First, the shape recovery cannot be broken down into a collection of percepts, each one associated with its specific, independent stimulus, invariant, or transformation. Second, the particular process by which the available information is utilized by the visual system has direct psychological implications. It is evident that in the recovery of structure from motion the visual system does not make full use of the information available to it. For example, if the number of elements in view is small, or if the presentation time is short, humans will fail to perceive the correct three-dimensional structure although sufficient information is in fact available. It seems, therefore, that for a satisfactory explanation of visual perception the "pickup of available information" will have to be studied and analyzed, rather than taken as a primitive construct. In both the direct and indirect theories, then, visual perception relies on the information in spatio-temporal patterns of light. The underlying question on which they disagree is whether the information in these patterns is indeed picked up immediately.

The psychophysical investigation of frequency-tuned channels in human vision can illustrate some of the distinctions between immediate and non-immediate registration of information and patterns. Following the work of Campbell and Robson [1968], substantial evidence has been accumulated for the existence in human vision of a number of distinct channels, or mechanisms sensitive to different ranges of size and spatial frequency. It has been shown (e.g. in Richards & Polit, 1974; Julesz & Miller, 1975; Watson & Nachmias, 1977; Wilson, 1978; Marr & Poggio, 1979; Wilson & Bergen, 1979) that a variety of phenomena in pattern detection, pattern discrimination, and stereoscopic vision can be explained by the properties of the channels and non-linear interactions among them. It also appears that the basic properties of the channels themselves are a direct reflection of the receptive field properties in the retina and the lateral geniculate nucleus. These encouraging results illustrate a number of points concerning the immediate registration of patterns and information. In general, the "directness" of perceptual mechanisms may be a matter of degree, with no absolute boundary distinguishing the direct from the indirect. In the above example it appears that one can be comfortable with viewing the underlying channels as the basic mechanisms that register patterns of light more or less directly, since (a) the channels appear to be explicable in physiological terms, and (b) the detailed dissection of the channels does not appear to have significant perceptual implications.

More complex visual modules, such as stereopsis, can then be explained using the properties of the underlying channels and the interactions among them. The conclusion from this example is that a psychologically meaningful decomposition of, e.g., stereoscopic vision, seems possible. But if it is, then the explanation of stereoscopic vision as the immediate pickup of binocular information [Gibson, 1979, Ch. 12] would not be justified.

The same argument is relevant for other perceptual and non-perceptual domains. If meaningful decompositions are possible, then the psycholinguist, for instance, should be dissatisfied with the suggestion that we comprehend utterances in natural language simply because our auditory system is tuned to directly pick up their meanings. Similarly, the perceptual psychologist should be dissatisfied with the claim that a property like rigidity is directly picked up. The underlying reason is that an attempt should be made to elaborate these processes, rather then accept them as primitive constructs. If such an elaboration is possible, it would serve as an integral part of our understanding of the linguistic and perceptual processes. Even if such an elaboration may ultimately prove to be difficult or perhaps unattainable, the implication of the foregoing discussion is that the direct explanations should better be regarded as a 'last resort', rather then a starting point, for cognitive theories.

4. *Perceiving the three-dimensional structure of moving objects*

This section will examine the approach of the DVP theory to the problem mentioned above of perceiving the three-dimensional structure of a changing environment. This problem was one of the most extensively studied within the immediate perception approach, and its examination can serve to illustrate some of the shortcomings inherent in this approach.

Changes in the structure of the environment relative to the observer can be caused by the movements of the observer, by motion of objects in the environment, and by non-rigid transformations of objects. In the case of object motion relative to the observer, the visual system has a remarkable capacity for correctly recovering the three-dimensional shape of the moving objects, even when the objects are unfamiliar, and when each static view of the scene contains no information about the three-dimensional structure of the objects.

The first systematic study of this capacity was carried out by Wallach & O'Connell [1953] in the study of what they have termed the "kinetic depth effect". In their experiments, an unfamiliar object was rotated behind a translucent screen, and the shadow of its projection was observed from the other side of the screen. In most cases the observers were able to give a correct description of the hidden object's structure and motion even when each static view of the object was unrecognizable and gave rise to no three-dimensional impression.

In the original study of the kinetic depth effect, as well as in later studies [Wallach *et al.*, 1956; Jansson & Johansson 1973], the ability to perceive structure from motion was accounted for in terms of an "effect" produced by lines and contours that change simultaneously in both length and orientation [4]. This explanation which offers a direct coupling between a percept and a certain class of two-dimensional patterns is, however, highly unlikely. If only actual lines in the image were considered, the account is manifestly false, since the structure of unconnected dots can be recovered through their motion. Imaginary lines connecting identifiable points were therefore admitted as well [Wallach & O'Connell, 1953]. But the resulting condition (i.e. that the perception of three-dimensional structure is produced by lines, virtual lines, and contours that change in both length and orientation) is certainly insufficient. Consider for example the random motion of unconnected elements in the frontal plane. The virtual lines between them change constantly in both length and orientation, but no coherent three-dimensional structure is perceived. The above condition is also necessary in a trivial sense only: the only two-dimensional transformations of the image that violate Wallach and O'Connell's condition are rigid transformations (of the image, not of the three-dimensional objects) and uniform scaling. But if the structure of a three-dimensional object is not recoverable from a single projection, it is hardly surprising that a uniform displacement, rotation, or scaling of the image itself, are insufficient for revealing the unknown structure [5].

The perception of structure from motion was also addressed by Gibson and his collaborators. The first solution proposed in their studies was that kinetic depth phenomena are induced by gradients of velocities. This hypothesis was not confirmed, however, by empirical investigations (see a review in [Epstein & Park, 1964; Farber & McConkie, 1979]). A different hypothesis in later studies suggested that continuous perspective transformations are directly registered by the eye [Gibson, 1954; 1957; 1965; 1968; Gibson & Gibson, 1957; von Fieandt & Gibson, 1959]. But this hypothesis raises difficult problems: What singles out those two-dimensional transformations that originate from the motion of rigid objects, and how can these transformations be registered by the eye? Hay [1966], in an extension of Gibson's analysis, tried to provide some answers to these questions by using techniques from projective geometry. A major difficulty with applying projective geometry to the problem at hand is that the transformations induced by the projections of a moving object are not equivalent to the group of projective transformations studied in projective geometry. (Projective transformations are the projection of non-singular *linear* transformations. The motion of objects is not, in general, a linear transformation.) Hay tried to circumvent some of the difficulties by (a) restricting his analysis to planar objects, and (b) decomposing the problem, and treating the perception of moving objects as based on eight distinct stimuli that can be studied separately. It proved impossible, however, to extend the analysis to non-planar objects, nor was it possible to identify the relation between the eight basic stimuli and the various motion percepts [Hay, 1966; Gibson, 1968]. Additional problems with the hypothesis of continuous perspective transformations are that neither perspectivity nor continuity are required for the perception of structure from motion [Ullman, 1979a]. A later attempt at identifying the immediate stimuli for the perception of moving objects concentrated on the notion of invariants [Gibson, 1960; 1966; 1972; 1979]. This programme states that in the transformations induced by moving objects some aspects of the patterns change while others remain invariant. It is hypothesized that the invariants are directly registered by the eye, giving rise to the perception of objects in motion. In this latter formulation the notion of invariances assumes a pivotal role in motion perception: "The perceptual system simply extracts the invariants from the flowing array; it *resonates* to the invariant structure or it is *attuned* to it" [Gibson, 1979; p. 249]. More generally, "The extracting and abstracting of invariants are what happens in both perceiving and knowing" [p. 258].

In evaluating the invariance-based programme it is worth noting that the question of whether a given system follows some rules of invariance is often merely a matter of convenience. For instance, the physical rules governing the motion of a free-falling object can be expressed in terms of invariant total energy (potential energy is transformed into kinetic energy). Alternatively, they can be expressed in terms of the effect of gravitational forces. The rules of mechanical motion can be expressed in yet another formalism (also favored by some theories of perception), the formalism of minimum principles. In Hamiltonian mechanics, motion is governed by de Maupertuis' principle of least action. For formulations of minimum principles in perception see, e.g., Mach [1897], Hochberg & McAlister [1953], Attneave & Frost [1969], Attneave [1972], Restle [1979], and the Gestalt *Pragnanz* principle [Koffka 1935].

The question of which formalism is to be used, whether a minimum principle, an invariance, or otherwise, is of secondary concern to the theory of visual perception in its current stage. Since little is known about the rules governing perception, the primary concern is the discovery of these rules, rather then the feasibility of an invariance-based formulation. The definition of invariances in the theory of direct perception is in fact so broad that almost any rule, once discovered, can be reformulated in terms of invariances

[6]: "A great many properties of the array are *lawfully* or *regularly* variant with changing observation point, and this means that in each case a property defined by the law is *invariant*" [Gibson, 1972; p. 221].

The relevant problem for the perception of structure from motion is therefore not whether the information in the visual array and the perception of moving objects are expressible in terms of invariances, but what the information is and how it is utilized by the visual system. A formulation in terms of invariances would be advantageous for the theory of direct perception if invariances could be discovered in the changing visual array that would be (a) informative enough to specify the structure of the moving objects, and (b) simple enough so that it would be reasonable to suggest that they are picked up directly. A hypothesis along these lines has been made [Gibson, Owsley & Johnston, 1978] by suggesting that the cross-ratio, which is known from projective geometry to be an invariant of projective transformations, underlies the perception of moving objects [7]. Whether or not the cross-ratio invariance is indeed utilized by the perceptual system is an open question. But since it requires four collinear points, and cannot reveal the structure of moving objects in general, it cannot even begin to answer the problem of recovering the structure from the changing projection. As has been mentioned above [Section 3.2.2], alternatives do exist: there are schemes that can recover unambiguously the structure of moving objects. But these schemes are neither direct nor based on invariances [Johansson 1964; 1970; Ullman 1979a; footnote 8].

In summary, several inherent shortcomings of the direct perception approach are manifest in the attempt to apply the theory to the perception of moving objects. The direct approach leads to viewing the perception of moving objects as a collection of percepts or "effects" produced by characteristic stimuli. The decomposition of perception into simple, distinct percepts, and the search for stimulus characteristics that can reasonably be registered directly, did not prove very fruitful (at least in the sense that no direct scheme exists that can describe the three-dimensional shape that will be perceived from the changing stimuli in the Kinetic Depth demonstrations). The more promising indirect schemes suggest that this may reflect inherent problems in the direct approach, not merely a temporary failure to identify the relevant stimulus invariances.

4.1 *Mach's illusion and the possible role of internal representations*

The perception of moving objects can serve to illustrate an additional source of dispute between the theory of immediate perception and current "indirect" theories. A well known phenomenon in motion perception is the illusion named after Ernst Mach [9]. Mach's illusion can be demonstrated in the following way. Consider a sheet of paper folded to create a standing v-shaped figure. When viewed monocularly, this shape is ambiguous, the v-shape can reverse in depth [Eden, 1962; Lindsay & Norman, 1972]. An observer views the v-shaped object monocularly, and waits for a depth reversal to occur. The reversal having occurred, he slowly moves his head left and right, up and down, forward and backward. The result is startling: the object seems to move whenever the head does. (Similar illusions can be produced by other constructions, e.g. a wireframe cube, and by motion of the object rather than the observer.) This illusory motion arises despite the observer's knowledge of the true situation, and it often contradicts shading information, stability criteria, and touch information [Eden, 1962].

The perception of structure and motion in this example is a function of two variables: the incoming image, and the current interpretation of the observer. The perception cannot be predicted on the basis of the stimulus alone. If, however, the current interpretation of the observer is known as well (the observer might report, for example, the perceived shape before he starts to move), then the perception can be predicted accurately. (For additional support for the pertinence of "internal states" to perception see Attneave, 1972; Gyr, 1972, 1979; Hochberg, 1974; Epstein, 1977; Gilchrist, 1977; Rock, in press.)

The theory of direct perception sometimes dismisses misperceptions and ambiguities as "non-ecological" and irrelevant to the theory of perception. Gibson argues [Gibson, 1972; 1979; Ch. 9] that if these irrelevant cases are dismissed, then perception becomes a function of the stimulus and nothing else. This is, of course, nothing but a tautology: if only stimuli that give rise unambiguously to unique perceptions are considered, then stimuli and percepts are related by a one-to-one mapping. Such a mapping does not disprove the existence or the irrelevance of internal states. It does restrict the analysis, however, to situations that make the internal states less accessible [10].

The perception in Mach's illusion evidently depends on the internal state of the observer. One current approach to the internal states of the perceptual system is to suggest that a certain representation of the environment is constructed during the perceptual process. This representation can mediate the consistent integration of information from a variety of sources, and make it explict and accessible. In Mach's illusion the misperception of motion is consistent with the changes in the image together with the misperceived structure. Perception is then determined by the incoming image together with the current state of the internal representation [11].

4.2 *Empirical investigations of internal representations*

The above discussion considered the "internal states" of the perceptual system in the case of Mach's illusion. If, however, something like an internal representation of the environment exists in this case, it is unlikely that it is constructed in this case only; it is more likely to be a part of the perceptual process in general. In addition, there has been in recent years a growing body of evidence regarding the existence and nature of the internal representations in a variety of situations. Although the emphasis here is on a theoretical analysis, I shall describe briefly some of this evidence, as it bears directly on the problem of internal representation.

The current research into the nature of the internal representations in perception received much of its thrust from the experiment of Shepard and Metzler [1971]. In this experiment, subjects were presented with 1600 images, each one depicting a pair of three-dimensional objects. In all cases the two objects were separated by rotation in space, i.e., they had a different orientation with respect ot the viewer. Half of the pairs depicted two objects of identical three-dimensional shape. In the other pairs the two objects were not identical, but a mirror image of each other. The subject's task was to decide as quickly as possible whether the two objects were identical in shape.

The main finding of the experiment was that response time to the identical pairs varied linearly with the angular separation between the objects. Furthermore, it did not matter whether the portrayed objects were separated by rotation in the image plane or in depth. These findings were subsequently replicated and extended (see [Shepard, 1975; 1978] for a summary of results). One noteworthy variant of the experiment established that when the two objects are presented successively, and the subject is given sufficient advance information concerning the object to be presented and its orientation, then the response time becomes uniform, i.e., independent of the orientation difference.

Shepard and Metzler's interpretation of the data was that the perceived identity in the experimental situation required a transformation of internal representations. This transformation has an effect equivalent to rotating one representation in an attmept to bring it to registration with the other. The linear dependence is then explained in terms of a constant rate of this rotation-like operation. In the case of sufficient prior information the transformation can be performed prior to the presentation of the second object, thus reducing the response time in the observed manner.

The particular scheme suggested by Shepard and his co-workers has been the source of much debate, and alternative theories have been proposed (e.g., Pylyshyn, 1976; Marr & Nishihara, 1978; Hinton, 1979; Sutherland, 1979; Kosslyn, in press). Common to all the alternative explanations, however, is the suggestion that a reasonable account of these and related phenomena would involve processes operating on internal representations. It is conceivable that a different kind of explanation that does not employ internal representations may be offered. It may also be argued that the above tasks are not "purely perceptual," and that while internal representation may underlie these tasks they play no role in other aspects of perception.

My own view is that the border line between pure and non-pure perception is somewhat artificial in this case. If internal representations of some sort will be shown to play a role in the tasks of the type studied by Shepard and his co-workers, they are likely to play a role in the theory of perception in general. Some fundamental differences between this representational view and the theory of immediate perception are discussed in the next section.

5. *From function to mechanisms*

In the theory of direct visual perception, the visual process is to be understood on two levels that can be roughly labeled "information content" and "mechanism". On the first level the information content of the visual array, e.g., the "ecologically valid" transformations and invariants, and the way they specify object and events is to be analyzed. The second level belongs primarily to the realm of physiology, and its task is to unravel the neural mechanisms that register the information explored at the first level.

A different approach, described by Marr & Poggio [1976], distinguishes three main levels in the understanding of information-handling systems: the levels of function, algorithm, and mechanism [12]. Although the border lines between the levels are not always clear, the distinctions are useful in examining the relations between various aspects of information-handling systems. The first and last of these levels roughly correspond to the analysis of information content and mechanisms respectively. The intermediate algorithmic level is indispensable in bridging the gap between the levels of function and mechanism. A simple example may illustrate this role. Suppose that an investigator tries to unravel the internal workings of the electronic calculator we have considered in Section 2. One possible approach would be to investigate the mechanism by probing the currents and voltages of the various components. If the function of the calculator is unknown to the investigator, he would face a difficult, perhaps impossible, task. Understanding the function of the system as performing arithmetic operations would facilitate the study of the mechanism, and would also serve an integral part in the theory of the system [c.f. Ullman 1979b; p. 1-4]. The theory of arithmetic is, however, insufficient for the mapping of arithmetic operations onto the mechanisms within the system. For the theory of arithmetic, the particular representation of numbers, for instance, is immaterial. It can be binary, decimal, or any other representation. Knowledge of the particular representation employed would become, however, instrumental in trying to identify the roles of particular mechanisms within the system. This conclusion is not restricted to simple artificial devices. The general point is that if representations are employed, then a detailed study of the representations and the operating processes is required to relate the level of function to the level of the physical mechanisms.

The dismissal of the middle level, which includes processes, representations, and the integration of information, as immaterial "intervening variables" leads to three deficiencies in the theory of perception. First, as we have seen, the algorithmic level plays an indispensible role in bringing together the studies of function and of mechanism. Second, the elucidation of the participating representations and processes constitutes an integral part of the theory of perception. The behaviorist might object to this notion and question whether representations and processes "really exist". Thus Neff [1936] in a review of theories of motion perception, concludes that "the assumption of an active mind is one of the most primitive beliefs of mankind" [p. 39], and Gibson dismisses perceptual processes as "old-fashioned mental acts" [1979; p. 238]. But a distinction has to be drawn between "symbolic" and "mental" [13]. The mediating processes in the computational/-representational theory do not operate on subjective experiences [Gibson, 1979; p. 238], nor are they intended to account for their origin. Subjective experience remains for the computational/representational approach (as it is for the direct approach) a complete mystery. Gibson's objection to the computational approach on the grounds that "no one has suggested that a computer has the experience of being here" [Gibson, 1972; p. 217] cannot serve therefore to refute the computational approach. In fact, the perceptual processes are not necessarily open to conscious introspection. Consequently, the introspective impression that the perception of objects is immediate and unanalyzable cannot be taken as evidence supporting the theory of immediate visual perception [c.f. Gibson, 1972; p.222].

The calculator example examined above illustrates in what sense processes and representations are amenable to an empirical investigation: certain events and components within the calculator can consistently be interpreted as having their meaning in the domain of numbers and operations on numbers [14]. There is nothing mysterious or mentalistic, then, in accepting and studying these intermediate representations and processes. Analogously, although the brain mechanisms may be very different from electronic ones, it is perfectly conceivable that certain events and components within the brain constitute (or can be consistently interpreted as) visual representations and processes that are amenable to empirical study, and are instrumental in explaining perception. The dismissal of the algorithmic level as immaterial is therefore unjustified in either sense of the word (i.e., "fictitious" on the one hand and "insignificant" on the other).

The third inadequacy in ignoring the algorithmic level is that it leads to oversimplifications of the theory. If processing is trivial or non-existent, then one is lead to search for "immediately registerable" information, such as the simple cross-ratio in the perception of three-dimensional structure in motion. If the role and complexity of the processes that "pick up" the information is appreciated, then it would be possible to realize that the information can assume less direct forms. The complexity of these underlying processes may be veiled by the subjective ease and immediacy of perception. But this subjective impression should not serve to underestimate their complexity. Schrodinger [1958] argued that as a process is perfected in the course of evolution, it "drops out of consciousness", and becomes inaccessible to introspection. If he is right, we can actually expect some of the most elaborate and perfected processes to be inaccessible to introspection. In any event, the possibility that perceptual processes may be highly complex has to be confronted. The process of stereopsis, i.e., the combination of information from the two eyes, exemplifies this hidden complexity in visual perception. Subjectively, it seems that all we have to do is to use both eyes, and binocular fusion occurs. We can "pickup information by looking" [Gibson, 1966, p. 3], or so it seems. The actual process turns out, however, to be highly complex. See Julesz [1971] for much of the empirical data, and Marr & Poggio [1979] for a recent theory of human stereopsis. In one respect Marr and Poggio's analysis agrees with Gibson's: it capitalizes on "ecological" properties such as the opacity and continuity of objects. But the Gibsonian view that what remains to be done is to pick up the invariances in the inputs to the two eyes turns out to be too simplistic. The information is extracted by an intricate interplay of filtering, matching, and eye movements [15]. This process establishes that there is sufficient information in the visual arrays to allow for the reliable extraction of stereo disparity. I doubt, however, that the method by which the stereo information is encoded can be revealed by examining the two inputs in search of the relevant immediate invariances, independent of the processes that extract this information [Gibson, 1961; 1979, Ch. 12].

Recently, Neisser [1976] expressed an uneasiness with what he called the information-processing view that describe cognition in terms of processing and "still more processing" [*ibid*, Figure 1]. He suggested that if Gibson is correct in his information-content analysis, perhaps we should play down the role of information processing and adopt an approach closer to the Gibsonian view:

> "If percepts are constructed, why are they usually accurate? ...The answer must lie in the kind and quality of optical information available to the perceiver...But if this is admitted the notion of 'construction' seems almost superfluous. One is tempted to dispense with it altogether, as J. J. Gibson has done" [p. 18].

It seems to me that this discontent is justified, but somewhat misguided. The crucial point is to appreciate the distinct roles of the first and the second levels of description. Some theories of the information processing approach have disregarded the theory level, substituting "processing and still more processing" for an underlying theory

[Marr, 1976; Pylyshyn, 1978; Ullman, 1978]. Processing models do not dispense with the information-content analysis. But the converse is also true: the fact that reliable information exists in the light array does not entail that processing is unnecessary. The role of the processing is not to create information, but to extract it, integrate it, make it explicit and usable [c.f. Marr, 1976; Ullman, 1979b, Ch. 5]. In conclusion, it would be misleading to pose the problem as a trade-off between "ecological optics" on the one hand and "information processing" on the other, since they play largely distinct roles. On the top level the functions of the visual system have to be understood. This level includes the information-content analysis of ecological optics. On the second level, the particular representations and processes employed by the visual system are to be explored. The third level includes physiological and anatomical studies of the neural mechanisms of the visual system, and the relation of these mechanisms to the representations and processes employed by the system.

I think that viewing the theory of immediate perception in light of the above three levels helps to put it in a proper perspective. The parts of the theory regarding the information content of the visual array, and its relation to the "ecology" are likely to make a lasting contribution to the theory of perception. The immediate approach, on the other hand, would have to be extended by a more comprehensive theory, that will draw an integrated picture of the perceptual systems on the levels of function, process, and mechanism.

FOOTNOTES

1. To avoid possible confusions, it may be helpful to list a number of related controversies that will not be in the focus of the discussion here, either because they have been discussed in detail in the past, or because they are not central to the arguments examined in this paper. These are:

(1) The role of past experience in perception [e.g. Gibson, 1972; Pittenger, Shaw & Mark, 1979].

(2) The interactions between non-visual modalities and visual perception [Gyr, 1972a, 1979; Errikson, 1974, Turvey, 1977].

(3) The degree to which the environment is specified by static images, and by changes in the visual array [Gibson, 1966; 1979; Neisser, 1976; Turvey, 1977].

(4) The differences between continuous optical flow and discrete sampling of the visual array [Gibson, 1972; Turvey, 1977].

2. If the "resonator" or "tuning-fork" metaphor used to describe the process of information pickup is taken too literally, it raises an additional difficulty: a tuning-fork is basically a linear device, while our visual system incorporates essential non-linearities (see, e.g, [Caelli & Julesz, 1978; Julesz & Caelli, 1979]). The term "resonator" will be interpreted therefore in a broader sense, i.e., any mechanism that can register information directly, not necessarily linearly.

3. The analysis of visual motion described in these schemes applies equally well to continuous and to discrete presentation. I do not wish to suggest that the human visual system employs a discrete sampling (in time) of the visual array. These schemes stand in contrast, however, with the claim that the interpretation of visual motion is unattainable on the basis of discrete sampling, which is central in [Turvey, 1977].

4. It should be noted that Wallach and O'Connell, as well as Johansson, do not subscribe to the direct approach in general. The explanation of the KDE as an "effect" produced by the simultaneous change in length and orientation is, however, "direct" in nature.

5.. Scaling can be used to indicate motion in depth [Marmolin, 1973] and time-to-collision [Lee, 1976], but not to recover structure from motion.

6. Similarly, Shaw, McIntyre and Mace [1974] emphasized the role of symmetries in direct perception. But the notion of symmetry in their formulation is broad enough to include, e.g., the rules of entropy, homeostasis, adaptation, and the attainment of knowledge.

7. The cross-ratio is defined in projective geometry for four collinear points (a, b, c, d) to be (ac * bd)/(bc * ad). The cross-ratio of four distinct points is invariant under projection.

8. The perceived structure is, of course, an invariance. But the registration of *this* invariant is simply equivalent to the original problem.

9. The depth reversal of Mach's figure, but not the motion effects, are discribed in Mach, [1897].

10. Chomsky [1959] makes a similar argument against the mapping between stimuli and responses in behaviorism. For details see [Chomsky, 1959; p. 551].

11. See also the discussion of the integration of size and orientation information in [Hochberg, 1974].

12. In announcing the establishment of a Center for Cognitive Studies at MIT, the same three levels were described as the skeleton not only for the study of visual perception, but for the Cognitive Sciences in general. Tech Talk, 23(28), March 21 1979.

13. While Neff, Gibson and others view symbolic events as mental, others have committed the opposite error, reducing subjective experiences to symbolic processes. For example, E. R. John claims that "consciousness itself is a representational system" and can be explained in terms of information processing [Thatcher & John, 1977], and G. J. Taylor contends that the study of conscious experience is a legitimate branch of natural science [Taylor, 1962]. For more discussion of this point see Griffin [1979] and Ullman [1979c]. More generally, I do not wish to claim that the computational/representational theory is likely to encompass all aspects of perceptual phenomena, certainly not all aspects of the mind. The claim, however, is that it provides a more satisfactory psychological theory of perception than the DVP theory.

14. The interpretation is not necessarily unique, but this difficulty is not central to the argument here.

15. In Marr and Poggio's theory. But even if the theory is incomplete or incorrect, to fit the available data it seems likely that any competing theory would be at least as complex.

Acknowledgement: I wish to thank E. Hildreth, W. Richards, and K. Stevens for their invaluable help.

REFERENCES

Attneave, F. Representation of physical space. In: A. W. Melton & E. Martin (eds.) *Coding Processes in Human Memory*, Washington, D.C.: V.H.Winston & Sons, 1972.

Attneave, F. & Frost, R. The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psychophysics, 6(6B)*, 391-396, 1969.

Bentley, D., & Hoy, R. R. The neurobiology of cricket song. *Scientific American, 231*, 34-44, Aug. 1974.

Caelli T. M., & Julesz, B. On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics, 28*, 167-175, 1978.

Campbell, F. W. & Robson, J. G. Application of Fourier analysis to the visibility of gratings. *J. Physiol. 197*, 551-566, 1968.

Chomsky, N. A review of B. F. Skinner's *Verbal Behavior*, 1959, reprinted in J. A. Fodor & J. J. Katz (eds.) *The Structure of Language*, Englewood Cliffs, N. J.: Prentice-Hall, 1964.

Eden, M. A three-dimensional optical illusion. *Quarterly Progress Report No. 64* M.I.T R.L.E. 267-274, 1962.

Epstein, W. & Park, J. Examination of Gibson's psychophysical hypothesis. *Psychological Bulletin, 62*, 180-196, 1964.

Epstien, W. What are the prospects for a higher-order stimulus theory of perception? *Scandinavian Journal of Psychology, 18*, 164-171, 1977.

Eriksson, E. S. 1974. A theory of veridical space perception. *Scandinavian Journal of Psychology, 15*, 225-235, 1974.

Farber, J.M. & McConkie, A.B. Optical motion as information for unassigned depth. *Journal Exp. Psychol., Human Perception and Psychophysics, 5(3)*, 494-500, 1979.

Fieandt, von Kai and Gibson, J. J. The sensitivity of the eye to two kinds of continuous transformations of a shadow-pattern. *Journal of Experimental Psychology, 57*, 344-347, 1959.

Gibson, E. J., Owsley, C. J. & Johnston, J. Perception of Invariants by five-month-old infants: Differentiation of two types of motion. *Developmental Psychology, 14(4)*, 407-415, 1978.

Gibson, E. J., Gibson, J. J., Smith, O. W. and Flock, H. Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology, Vol. 8, 1,* 40-51, 1959.

Gibson, J. J. *The Perception of the Visual World.* Boston: Houghton Mifflin, 1950.

Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychological Review, 61(5),* 304-314, 1954.

Gibson, J. J. Optical motions and transformations as stimuli for visual perception. *Psychological Review, 64(5),* 288-295, 1957.

Gibson, J. J. and Gibson, E. J. Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology, Vol. 54, 2,* 129-138, 1957.

Gibson, J. J. Perception as a function of stimulation. In: S. Koch (ed.) *Psychology: A Study of a Science,* Vol. I, New York, Toronto, London: McGraw-Hill, 1959.

Gibson, J. J. The concept of the stimulus in psychology. *American Psychologist, 15,* 694-703, 1960.

Gibson, J. J. Ecological optics. *Vision Research, 1,* 253-262, 1961.

Gibson, J. J. Research on the visual perception of motion and change. In: Spigel, I. [1965] 125-146, 1965.

Gibson, J. J. *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin, 1966.

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

Gibson, J. J. What gives rise to the perception of motion? *Psychological Review, Vol. 75, 4,* 335-346, 1968.

Gibson, J. J. Are there sensory qualities of objects? *Synthese, 19,* 408-409, 1968-9.

Gibson, J. J. A theory of direct visual perception. In: J. R. Royce & Wm. W. Rozeboom (eds.) *The Psychology of Knowing.* New York, Paris, London: Gordon & Breach, 1972.

Gibson, J. J. *The Ecological Approach to Visual Perception.* Boston, Houghton Mifflin Company, 1979.

Griffin, D. R. Prospects for a cognitive ethology. *The Brain and Behavioral Sciences, 1(4),* 527-538, 1978.

Gilchrist, A. Perceived lightness depends on perceived spatial arrangement. *Science, 195,* 185-187, 1977.

Gyr, J. W. Is a theory of direct visual perception adequate? *Psychological Bulletin, 77,* 246-261, 1972a.

Gyr, J. W. Comments on Gibson's paper. In: J. R. Royce & Wm. W. Rozeboom (eds.) *The Psychology of Knowing.* New York, Paris, London: Gordon & Breach, 1972b.

Gyr, J., Willey, R. & Henry, A. Motor-Sensory feedback and geometry of visual space: a replication. *The Behavioral and Brain Sciences, 2,* 59-64, 1979.

Hay, C. J. 1966. Optical motions and space perception -- an extension of Gibson's analysis. *Psychological Review, 73,* 550-565, 1966.

Henle, M. On naive realism. In: R. B. Macleod H. L. Pick, Jr. *Perception: Essays in Honor of James J. Gibson.* Ithaca and London, Cornell University Press, 1974.

Hinton, G. Some demonstration of the effects of structural descriptions in mental imagery. *Cognitive Science, 3,* 231-250, 1979.

Hochberg, J. Higher-order stimuli and inter-response coupling in the perception of the visual world. In: R. B. Macleod H. L. Pick, Jr. *Perception: Essays in Honor of James J. Gibson.* Ithaca and London, Cornell University Press, 1974.

Hochberg, J. and McAlister, A. A quantitative approach to figural "goodness". *Journal of Experimental Psychology, 46*, 361-364, 1953.

Jansson, G. & Johansson, G. Visual perception of bending motion. *Perception, 2,* 321-326, 1973.

Johannson G. Perception of motion and changing form. *Scand. J. Psychol., 5,* 181-208, 1964.

Johansson, G. On theories for visual space perception -- A letter to Gibson. *Scan. J. Psychol., 11.* 67-74, 1970.

Julesz, B. *Foundations of Cyclopean Perception.* Chicago: The University of Chicago Press, 1971.

Julesz, B., & Caelli, T. On the limits of fourier decomposition in visual texture perception. *Perception, 8,* 69-73, 1979.

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

Koffka, K. *Principles of Gestalt Psychology.* New York: Harcourt, Brace, and World, 1935.

Kosslyn, S. M. *Image and Mind.* Cambridge, Mass: Harvard Univ. Press, in press.

Lee, D. N. A theory of visual control of braking based on information about time-to-collision. *Perception, 5,* 437-456, 1976.

Lindsay, P. H. and Norman, D. A. *Human Information Processing.* New York and London: Academic Press, 1972. Longuet-Higgins, H. C. & Prazdny, K. The interpretation of a moving retinal image. *Proc. Roy. Soc. Lond.,* in press.

Longuet-Higgins, H.C. & Prazdny, K. The interpretation of a moving retinal image. *Proc. Roy. Soc. Lond.* (in press).

Mach, Ernst. *The Analysis of Sensations.* New York: Dover Publications, 1897.

Marmolin, H. Visually perceived motion in depth resulting from proximal changes. *Perception & Psychophysics, 14(1),* I 133-142, II 142-148, 1973.

Marr, D. Early Processing of visual information. *Philosophical Transactions of the Royal Society of London, Vol. 275,* 942, 483-534, 1976.

Marr, D. Artificial Intelligence -- A personal view. *Artificial Intelligence, 9,* 37-48, 1977.

Marr, D. and Poggio, T. From understanding computation to understanding neural circuitry. *Neuroscience Research Program Bulletin, 15(3),* 470-488, 1977.

Marr, D. & Nishihara, K. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Roy. Soc. London B, 200,*269-294, 1978.

Marr, D., & Poggio, T. A computational theory of human stereo vision. *Proc. R. Soc.-London B, 204,* 301-328, 1979.

Metzger, W. Critical remarks to J. J. Gibson's conception of "direct" visual percepetion, i.e., of revived prephysiological realism. In: J. R. Royce & Wm. W. Rozeboom (eds.) *The Psychology of Knowing.* New York, Paris, London: Gordon & Breach, 1972.

Neff, Walter S. A critical investigation of the visual apprehension of movement. *American Journal of Psychology Vol. XLVIII,* 1-42, 1936.

Neisser, U. *Cognition and Reality.* San Francisco: W.H. Freeman and Company, 1976.

Pittenger, J.B., Shaw, R.E. & Mark, L.S. Perceptual information for the age of faces as higher order invariances of growth. *Journal Exp. Psychol., Human Performance and Psychophysics, 5(3),* 478-493, 1979.

Pylyshyn, Z. Imagery and artificial intelligence. In: W. Savage (ed.), Minnesota Studies in the Philosophy of Science, Vol. IX, Minneapolis: University of Minnesota Press, 1976.

Pylyshyn, Z. Computational models and empirical constraints. *The Behavioral and Brain Sciences, 1(1),* 93-99, 1978.

Restle, F. Coding theory of the perception of motion configurations. *Psychological Review, 86(1),* 1-24, 1979.

Richards, W. &Polit, A. Texture matching. *Kybernetic, 16,* 1299-1302, 1974.

Rock, I. Alternative solutions to kinetic stimulus transformations. Journal of Experimental Psychology: Human Perception and Performance, in press.

Schrodinger, E. Mind and Matter, 1958. In: *What is Life & Mind and Matter,* Cambridge: Cambridge University Press, 1967.

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 James J. Gibson.* Ithaca and London, Cornell University Press, 1974.

Shepard, R. N. Form, formation, and transformation of internal representations. In: R. Solo (ed.), Information processing and cognition: the Loyola Symposium, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Shepard, R. N. The mental image. *American Psychologist, 33,* 125-137, 1978.

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

Sutherland, N. S. The representation of three-dimensional objects. *Nature, 278,* 395-398, 1979.

Taylor, J. G. *The Behavioral Basis of Perception.* New Haven, Yale University Press, 1962.

Thatcher, R. W. & John, E. R. *Foundations of Cognitive Processes.* Hillsdale, N. J., Lawrence Erlbaum, 1977.

Turvey, M. T. Contrasting orientations to the theory of visual information processing. *Psych. Review, 84(1)*, 67-88, 1977.

Ullman, S. Artificial Intelligence systems and human cognition: the missing link. *The Behavioral and Brain Sciences, 1(1)*, 117-119, 1978.

Ullman, S. The interpretation of structure from motion. *Proc. Roy. Soc. London B, 203*, 405-426, 1979a.

Ullman, S. *The Interpretation of Visual Motion.* Cambridge and London: MIT Press, 1979b.

Ullman, S. Mental representations and mental experiences. (A commentary on Griffin's paper.) *The Behavioral and Brain Sciences,* 1979c.

Wallach, H. and O'Connell, D. N. The Kinetic depth effect. *Journal of Experimental Psychology, 45, 4*, 205-217, 1953.

Watson, B. A., & Nachmias, J. Patterns of temporal interaction in the detection of gratings. *Vision Research, 17*, 893-902, 1977.

Wilson, H. R. Quantitative characterization of two types of line spread function near the fovea. *Vision Research, 18*, 493-496, 1978.

Wilson, H. R. & Bergen, J. R. A four mechanism model for spatial vision. *Vision Research, 19(1)*, 19-32, 1979.

Yolton, J. W. Gibson's realism. *Synthese, 19*, 400-407, 1968-9.

Zaretsky, M. D. Patterned response to song in cricket central auditory neurone. *Nature, 229*, 195-196, 1971.