Acta Psychologica 67 (1988) 59–93
North-Holland

INVITED PAPER

MODULES IN VISION: A CASE STUDY OF INTERDISCIPLINARITY IN COGNITIVE SCIENCE *

Johan P. WAGEMANS
University of Leuven, Belgium

Accepted August 1987

The usefulness of interdisciplinarity in cognitive science (CS) is demonstrated by an overview of recent algorithms for recovering 3-D aspects from a 2-D input. The recovery processes are considered to be modular in Fodor's (1983) sense. This tenet, based on a thesis from philosophy of mind (PhM), proves to have serious impact both on computer vision (CV) as such and on computational theories of perception (CTP), which provide a compromise between classic indirect and direct theories (ITP and DTP) in perceptual psychology (PP). Both existent and possible mutual interchanges between CV, PP, and PhM are specified in current and future research on modular recovery processes such as shape from shading, depth from stereo, and structure from motion. Also, attention is paid to (meta) criticisms of PP and PhM on concrete hypotheses of CTP and on the CTP approach as a whole. For example, the relative independence of modular low-level vision processes is questioned, and the lack of an explanation of intentionality is highlighted. A plea is made to attempt to solve these fundamental (meta)criticisms within the CTP paradigm, since there are no logical arguments against this possibility, and because recent CTP theories are tackling these problems seriously.

1. Introduction

Cognitive science (CS), as a research program, is interdisciplinary by definition. The truism of this tenet is apparent from journals such as *Cognitive Science*, being the official organ of the Cognitive Science Society and bearing 'A Multidisciplinary Journal' as a subtitle, and from discussions on the foundations of the research program (e.g.,

* The writing of the article was supported by a grant from the National Fund for Scientific Research (N.F.W.O.) to Johan P. Wagemans. The author is indebted to K. Lamberts, H. Roelants, K. Verfaillie, C. de Weert, G. d'Ydewalle and some anonymous reviewers for critical reading of earlier drafts of this article and their thoughtful comments.

Correspondence concerning the article should be addressed to Johan P. Wagemans, Laboratory for Experimental Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium.

Chomsky 1980; Fodor 1980; Norman 1981; Pylyshyn 1980; 1984). Three cornerstones of CS, often acknowledged as such, are Artificial Intelligence (AI), psychology, and philosophy. It will be demonstrated here that these members of the CS triumvirate are not only supposed to be cooperating according to the articles of faith of the CS community, but that they are effectively doing so in the practice of daily scientific research. As an example of this actual interdisciplinarity, the contributions of computer vision (CV), perceptual psychology (PP), and philosophy of mind (PhM) to the current study of modules in vision are highlighted and commented. [1]

One could, in principle, unravel the interrelations among the three disciplines by first distinguishing among, on the one hand, influences from one discipline on another (see fig. 1A), and, on the other hand, common influences from two disciplines on the third and vice versa (see fig. 1B). Then, all the interchanges under these headings could be summed up and discussed. Although such a scheme could be justified for didactical purposes, I think the unraveling would cause a loss of the extra value of the interlace of the different branches. It would be as if one teased an Oriental carpet (handmade, 100% wool, a century old, at least) into separate fibres, to discover the way in which the fibres together constitute the carpet.

In order not to neglect the Gestalt adage that the whole is more than the sum of the parts, an introduction to the current research on modules in vision will be provided, without all the interactions mentioned in fig. 1 being disentangled.

Nevertheless, I do point at existent as well as desirable mutual interchanges between CV, PP, and PhM, when overviewing both the recent history and the near future of the studies on modular processes in visual perception. Also, in the end of the paper, the twelve possible interactions will be summed up in a more or less schematic way, only, however, after one was given the opportunity to see the three disci-

[1] The fact that the contribution of the neurosciences is not considered here, does not mean that there is none. It does reflect the opinion of Marr (1982) and his tradition of research on modules in vision that something important is missing in the neurophysiology of vision. Often, the behavior of cells is merely described, not explained. However, it must be admitted that, recently, some important steps have been taken to bridge the gap between the computational and neurophysiological approach (e.g., Arbib 1987; McClelland and Rumelhart 1985; Rumelhart and McClelland 1985; Wagemans 1987). One is even inclined to talk about the route from neuroscience to artificial intelligence (Koenderink 1987). An extensive account of this recent trend could be the topic of another full-fledged paper.

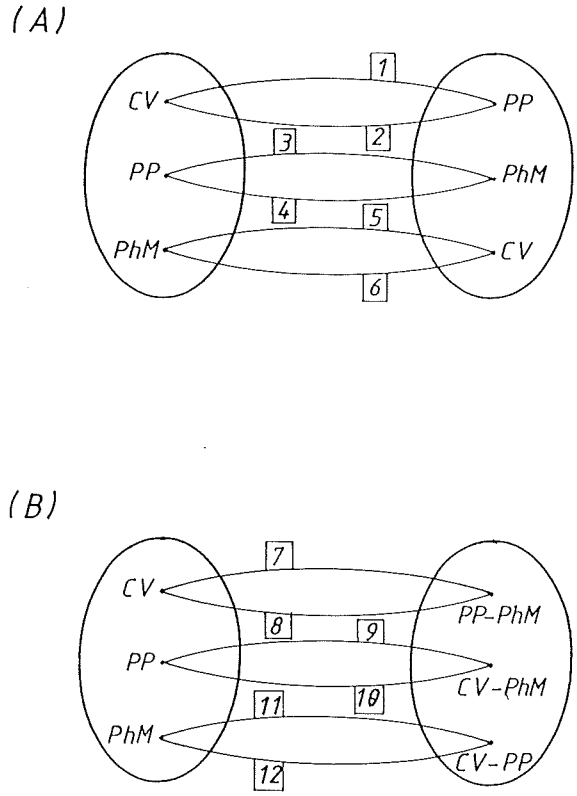

Fig. 1. Schematic summary of all possible interactions between the disciplines in CS studying vision. (A) The influences of one discipline on another. (B) The influences of one discipline on two others, and vice versa.

plines working together in solving a difficult scientific puzzle, viz. the recovery problem.

2. The recovery problem

2.1. Definition

In principle, a two-dimensional (2-D) pattern can result from any of an infinite number of three-dimensional (3-D) scenes (see fig. 2). Both human and machine vision are, therefore, confronted with the funda-

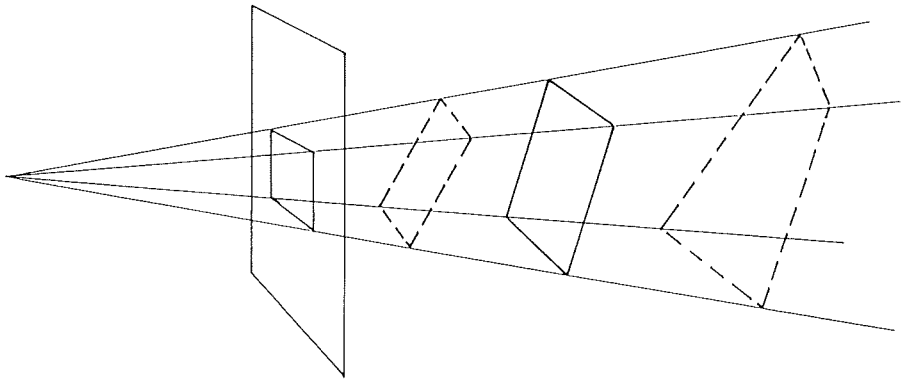

Fig. 2. The problem of underdetermination, central to the process of recovering 3-D structure from 2-D patterns: A 2-D image can result of a principally unlimited number of 3-D structures (adapted from Barrow and Tenenbaum 1981).

mental problem of reducing this number, ideally to one, in trying to recover the 3-D structure of a scene from the 2-D input images.

Despite the fact that PP and CV research have fundamentally different goals, viz. explaining vs. building a perceiving system, and, as a consequence, have different evaluation criteria, viz. corroborating of predictions vs. working of practical tools, it can be expected, on logical grounds already, that PP and CV studies can mutually benefit from each other's efforts, because they both have to face the recovery problem. Historically, PP and CV have been developed relatively independent from one another, but more recently, the link between the two disciplines has become very tight indeed.

2.2. PP solving the recovery problem

In traditional PP, two broad classes of theories have attempted to explain the human perceiver's ability to recover the 3-D structure of a scene from principally ambiguous 2-D patterns on the retina. In the first, often called 'indirect' or 'constructivist', the percept is regarded as the conclusion from an unconscious inference, not unlike deductive reasoning, the premises of which are both the impoverished visual input and the supplementary information provided by more cognitive means such as memory (e.g., Gregory 1970; Helmholtz 1910/1962; Hochberg 1981; Rock 1977, 1983).

Gibson (e.g., 1979) has firmly reacted against this view, and he has pled for another approach, the 'direct' or 'ecological' one. He and his followers emphasize the richness of the visual information, whose so-called ambiguity they regard as a consequence of contrived laboratory situations and as non-existent in the real world. Whereas the indirect theories of perception (ITP, used as a plural noun, since there is more than one variant of it) have tried to solve the recovery puzzle by invoking additional information from elsewhere in the information processing system, the direct theory of perception (DTP, used as a single noun, since there is a considerable amount of agreement on Gibson's proposals) has argued for the study of the information being available to a (moving) perceiving organism in a real, non-reduced environment. It would then, so they claim, soon become evident that there is a wealth of information waiting to be picked up directly, without any need of further (indirect) processing.

2.3. CV solving the recovery problem

Early CV studies provided the additional constraints, required for solving the difficult problem of underdetermination, by restricting the world itself. They made use of the so-called blocks world, in which there are well-established relations between lines, junctions, and regions in the 2-D image and edges, vertices, and surfaces in the 3-D scene. Different classification systems were suggested to list these relations and programs were written that made use of them to derive a plausible interpretation of the 3-D world on the basis of the 2-D image (e.g., Clowes 1971; Guzman 1968; Huffman 1971).

Waltz (1975) demonstrated that the inclusion of more detailed information, namely shadows, does not complicate the interpretative process, but, rather, facilitates it. This finding, which seemed contra-intuitive in those days, as well as the fact that humans are able to solve the recovery problem in an even much more complex and realistic world, made CV researchers think that there should be some, until then unknown, factors in the real world constraining the number of possible solutions to the puzzle of recovery. This idea was in remarkable accordance with Gibson's basic conjecture, a fact that did not go unnoticed by the CV community (e.g., Braddick 1980; Hinton 1980; Marr 1982; Zucker 1980).

From that moment on, almost all CV research efforts went into the

problem of finding the general mathematical (foremost, geometrical) and physical (foremost, optical) regularities in the world that might be used during the recovery. It is, therefore, now, contrary to the time before Waltz and both within and outside the CV community, not such a surprise anymore that more information facilitates rather than complicates the perceptual processing of it. For example, to Kanade (1981), it is not surprising that a solution to the recovery problem is facilitated, when more complexities are involved, taking into consideration that more constraints are also available then. In the DTP also, one adheres to 'Johansson's conjecture', that mathematically complex displays tend to yield simple and stable perception, suggesting that what is more complex for mathematicians may not be so for the casual perceiver (see, e.g., Warren and Shaw 1985).

Furthermore, CV researchers soon noticed that their algorithms and programs, incorporating the constraining factors in a mathematically rigid manner, could be conceived as specific and detailed models of the human recovery process. As a consequence, CV has, since then, announced itself as a part of both practically oriented AI and theoretically based PP. This latter part of the CV research program, henceforth called 'the computational theory of perception' (CTP), can be considered a third class of perception theories.

3. CTP as a third theory in PP

According to the CTP, perception is a process of computations on representations: The information that is only implicit in the retinal image is, through computations, transformed into subsequent representations that make certain aspects of the information more explicit and accessible for further computation. For example, Marr (1982) – whose work will further on be discussed in more detail – distinguishes the primal sketch, making information about intensity changes explicit (including length, position, orientation, and contrast of line fragments), the 2,5-D sketch, and the 3-D model representation, successively making information explicit about surface orientation and object shape.

The CTP can be viewed as a compromise between the two previous approaches (ITP and DTP), since it bears both resemblances and differences to them. It can, therefore, be called 'Gibholtzian', as Haber (1983) did.

3.1. CTP and ITP

On the one hand, CTP resembles ITP in arguing that knowledge about the regularities of the natural world is indispensable for computing representations. On the other hand, however, it differs from them in the kind of knowledge it considers necessary: CV incorporates *general* constraints provided by geometric and physical laws in the world, whereas ITP stresses top–down influences depending on high-level knowledge of *specific* objects. Also, there is a difference in what can be called 'the level of knowing'. I believe that CTP would hold that the visual system does not really know or apply the laws of optics and geometry, but simply functions according to the principles on the basis of which it is programmed (see Kanizsa 1985). Therefore, it is more appropriate to speak of 'knowledge' (between quotes), to contrast it with knowledge (without quotes) in the more traditional sense of ITP. A third divergence between CTP and ITP lies in their view on the use that is made of knowledge: According to CTP, knowledge helps to make use of what is given *in* the visual input, whereas in ITP knowledge is considered to be used to go *beyond* the information given.

3.2. CTP and DTP

CTP resembles DTP in emphasizing the richness of the visual input, but it differs from it in unraveling the different computations and representations needed to process the rich visual information. The DTP explicitly ignores the role of computations and representations in perception, which they consider as information *detection* instead of information *processing* (for the difference between the two, see, e.g., Wagemans 1986).

3.3. CTP and the ITP–DTP debate

The uprise of CTP can be seen as a factor causing a loss of intensity and furiousness in the controverse between ITP and DTP. Indeed, the CTP view on the importance of natural constraints has provided perceptual psychologists with an excellent opportunity to bridge the seemingly unbridgeable gulf between ITP and DTP. It is no surprise, therefore, that even 'classic' indirect theorists have taken pains, recently, to analyze the available stimulus properties rigorously and to

take physical and geometric constraints of the natural world into account (e.g., Haber 1983; Hochberg 1986; Palmer 1983; Shepard 1984). The other way around, direct theorists are paying considerable attention to the CTP research and they have designed some experiments in close interaction with it (e.g., Todd 1981; 1982; 1985).

A lot of technical results attained within the CTP and much of the general framework behind it, is due to the late David Marr of the Massachusetts Institute of Technology. It is not a great luxury, therefore, to consider his approach in the next paragraph. This can be done briefly because good introductory articles on CV and CTP exist and are easily accessible to workers in PP and PhM (e.g., Brown 1984; Cohen and Feigenbaum 1982; McArthur 1982; Rosenfeld 1984). Even more elaborate and sophisticated collections of research in CV and CTP are also available in abundance (e.g., Ballard and Brown 1982; Beck et al. 1983; Brady 1982; Hanson and Riseman 1978; Marr 1982; Pentland 1986a; Ullman and Richards 1984; Winston 1975; Winston and Brown 1979).

3.4. CTP and Marr's approach

The work of Marr (1982) deserves somewhat more attention, for it can be considered as the prototypical work in CV research of the 'second generation' (i.e., more relevant to PP), and since it is acknowledged as such by the scientific CV community (e.g., Brady 1982; Cohen and Feigenbaum 1982; Pentland 1986a; Winston and Brown 1979).

Marr clearly distinguished three levels of explanation, that should be separated when investigating information processing systems such as vision and that should be tackled in successive order, because a logical hierarchy exists between them. (These levels of explanation should not be confused with the levels of representation shortly mentioned above.)

First, one has to consider what is the general goal of the computation, why it is appropriate, and what is the general rationale of the strategy by which it can be carried out. A central theme of the inquiries of the computational theory is 'the business of isolating constraints that are both powerful enough to allow a process to be defined and generally true in the world' (Marr 1982: 23). The second level of analyzing vision involves choosing the representations for the input and the output, and the algorithms for transforming the one into the other.

The third and last level of explanation concerns the manner in which the representations and algorithms can be realized physically.

According to this view, the interactions between CV and PP are not primarily situated on the third level: the hardware implementation of the representations and algorithms will surely differ in humans and machines. It should be noted, however, that there is, recently, a serious attempt to bridge this gap by considering (neural or computer) networks that perform the computations largely in parallel (e.g., Ballard 1986; Ballard et al. 1983; Feldman 1985; Grossberg 1983; Sabbah 1985).

Nevertheless, psychophysical studies on human subjects prove to be indispensable in specifying the exact representations and algorithms used to solve the computational problem (i.e., the second level), and in clarifying which constraints are taken into account by a human perceiver (i.e., the first level). This will be demonstrated by a summary of the work on the algorithms which help to transform the primal sketch into the 2,5-D sketch (the so-called shape-from methods).

4. Shape-from methods

These methods are proposed as solutions to the critical problem of getting 'from images to surfaces' (Binford 1981; Grimson 1981), or 'from pixels to predicates' (Pentland 1986a). In this way, they are, in fact, the central part of the computational solution to the recovery problem. They are called 'shape-from methods', after the pioneering work on shape from shading by Horn (e.g., 1977).

4.1. Shape from shading

The problem of using shading (i.e., smooth intensity variations across the image) to recover shape (viz. gradual changes in surface orientations) is that one needs to know how the image intensity at a pixel (i.e., a picture element) is determined. In general, this image intensity depends on the positions of the surface, the illumination, and the viewer, as well as on the surface's material and orientation. In order to derive the surface orientation from the image intensity, some assumptions must be made on the other possible causal factors, or, in other words, restrictions or constraints must be imposed on the other potential determinants.

Horn (e.g., 1977) has shown that one can arrive at a unique solution to the recovery problem, when one assumes that (i) the relative positions of surface, illumination, and viewer, and the surface's material are known, that (ii) each image point is assigned to maximally one surface orientation, and that (iii) orientations vary smoothly almost everywhere except at boundaries. These constraints can be incorporated into a model of the geometry of image projection and the photometry of intensity formation. When this model is implemented as a data base, artificial vision systems are able to recover shape from shading. In industrial settings, the former assumption (i) is justified, since the positions of the three constituents involved can be determined and since the surface reflectance properties can be measured. The latter constraints (ii) and (iii) are mathematically rather robust, so that they can be supposed to hold even in a more natural environment.

From the moment, however, that CV workers want to consider the shape-from-shading algorithm as a plausible model of (part of) the human perceiver's process of recovering the 3-D structure from the 2-D patterns (hence, CTP instead of practical CV), one has to face the fact that, under natural viewing conditions, the illumination and reflectance properties are unknown, so that at least assumption (i) is unjustified. To remedy against this shortcoming, Pentland (1982; 1986b) has recently proposed an alternative to Horn's method that is more appropriate as a model for human perception, because it requires no prior knowledge about the direction of illumination. Furthermore, it can be implemented in a physiologically more plausible way. However, it still assumes illumination being constant in the regio under examination and surfaces being Lambertian (i.e., reflecting light equally in different directions).

This means that even Pentland's improved method cannot account for surface interactions such as specular highlights, indirect illumination, transparency, and cast shadows, that are all real factors in a natural environment and that have proven to play an important role in human visual perception (Beck 1972; 1975). Also, Todd and Mingolla (1983) have found out experimentally that human perceivers tend to do worse with displays simulating the pattern of reflection for an idealized Lambertian surface, a fact that is diametrically opposed to what could be expected from the CV studies. It will be clear, therefore, from this single example already, that PP has correctly pointed out some of the failures of the shape-from-shading algorithms, when they are con-

sidered as possible models for the human recovery process. In the following paragraph, it will be demonstrated that PhM has an equally important role on the CS scene.

4.2. Modularity of the shape-from methods

As stated above, the 'knowledge' concerning the geometrical and the optical constraints of the image formation must be incorporated in a model to be usable during the recovery process. In man-made vision systems this model is implemented as a data base. The basic conjecture of the CTP is that the 'knowledge' of the constraints needed for solving the puzzle of recovery by nature's own perceiving machine is present within a module, [2] a notion that is defined and elaborated by Fodor (1983, 1985), one of the most prominent current representatives of the PhM.

Low-level vision processes, among which the shape-from algorithms, being considered as modular is a crucial part of Fodor's 'Modularity-of-Mind' thesis. The basic tenet is that an information processing system such as the human mind consists of two clearly distinguishable parts, a low-level, modular part of input analyzers (vertical faculties), and a high-level, central cognitive system (horizontal faculties). The details of this thesis can be founded in the original essay (Fodor 1983). Here, the discussion will be centered around Fodor's view on input systems, which is also the major focus of his BBS Précis (Fodor 1985).

What exactly is a modular system? Instead of giving a clearcut, simple definition, Fodor proposes a number of constituting characteristics, which are here specified for shape-from algorithms. A module is domain specific (i.e., restricted to visual aspects of the input), mandatory (e.g., one cannot help seeing a 2-D pattern as a particular kind of 3-D structure), fast (i.e., considerably faster than prototypical central

[2] Although Marr (1982) did not invent the notion of 'module', he was one of the first to use it in this context. He uses the term in a functional sense, in contrast with the older notion of 'channel' that is more often used in a structural sense, especially when it occurs in the context of a neuroscientific approach. In Marr's sense, a modular process is one that is independent from another ongoing process. It became necessary to coine a term for this independency, when Julesz (1971) found that a 3-D form could be perceived from two 2-D random dot stereograms. This finding was interpreted as indicating that at least some processes (i.e., stereopsis) are possible without object knowledge (cf. random dots), that is, independent from higher-order perceptual and cognitive processes.

processes like problem-solving, though hard to quantify exactly), and informationally encapsulated (i.e., impenetrable by higher level cognitive influence, such as specific object knowledge). Furthermore, modules have only limited central accessibility (i.e., as a consequence their processes cannot be reported verbally by introspection), 'shallow' outputs (i.e., the categorization of visual stimuli is not very specific), fixed neural architecture (i.e., hard-wired), specific breakdown patterns (i.e., agnosias cannot be explained by mere quantitative decrements in global capacities), and characteristic pace and sequencing during ontogeny.

The notion of modular input systems is a crucial one in CTP being a compromise between ITP and DTP. Whereas 'classic' ITP view perception as totally dependent upon cognitive processes, so that the distinction between perception and cognition even vanishes, DTP claims that everything is detected or picked up directly, without any help of computations on representations. CTP states that some properties are indeed picked up directly, viz. via transducers, but, also, that other properties are apprehended only indirectly, viz. via processes that involve inferences (see, e.g., Fodor and Pylyshyn 1981; Pylyshyn 1980, 1984).

Whereas DTP claims that features of the 3-D scene are directly perceived on the basis of picked up features of the light, CTP offers an explanation of how a perceiver (man or machine) gets from detected properties of the light to perceived properties of the scene: One inferes the latter from the former on the basis of 'knowledge' of the correlations that connect them, 'knowledge' that is present within the recovery module itself and which, therefore, need not be provided by the higher level cognitive system.

Relating Fodor's PhM thesis with Marr's CTP theory, one could say that transducers convert the physical energy of the light into the symbolic elements of the primal sketch, whereas the 2,5-D sketch is computed by means of modular shape-from algorithms (the shape from shading discussed above and the other ones to be discussed further on). This modularity of the shape-from algorithms has, in the view of CTP, as a consequence that different methods for inferring 3-D shape from 2-D images can be studied in relative isolation. The success of Horn's method has, therefore, sparked off a boom of investigations on other shape-from methods.

4.3. Shape from stereo

For example, one has studied an algorithm to infer depth from stereo images. Stereopsis requires a solution to three different problems: the definition of the primitives used in the matching process, the finding of the right matches of the corresponding primitives, and the computation of depth by making use of the disparity (i.e., difference in relative position) of the primitives. The fundamental problem here is also one of undetermination (see fig. 3): Without further constraints, different matches are equally possible.

Marr and Poggio (1976) tackled this puzzle by looking for constraining properties in the real world, and they found that (i) black dots can only match black dots, (ii) almost always, a black dot from one image can match only one black dot of the other image, and (iii) the disparity of the matches varies only smoothly over the image. Together, these constraints of resp. compatibility, unicity, and continuity make it possible to arrive at a single solution for the recovery of depth from stereo images.

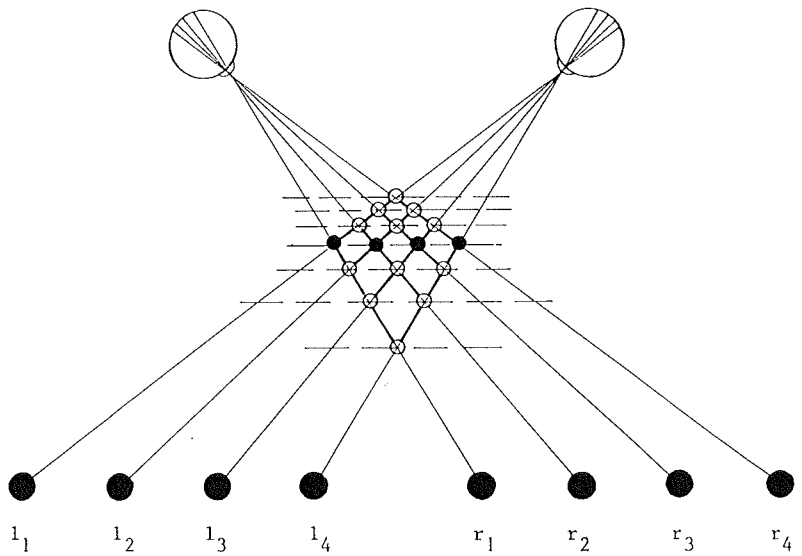

Fig. 3. The problem of underdetermination, applied to the process of recovering depth from stereo: there is a number of different possible matches even with only four elements specified in each image (adapted from Marr 1982).

Representing these constraints in a geometrical and physical model, Marr and Poggio (1976) embodied them in a cooperative algorithm (i.e., a specific way of propagating local constraints to attain a global solution, see, e.g., Davis and Rosenfeld 1981; Rosenfeld et al. 1976). However, a number of findings casted doubts on the relevance of the proposed algorithm as a model of human stereo vision: For example, the role of eye movements (Richards 1977), the ability to tolerate defocussed images (Julesz 1971), the ability to perceive depth in rivalrous stereograms (Mayhew and Frisby 1976), etc.

The main problems with the first algorithm (better performance than humans in some situations, worse in others) were resolved in a second algorithm, explicitly related to the notion of 2,5-D sketch and specifically designed as an account of how human stereo vision works (Marr and Poggio 1979). Other CV studies on shape from stereo have, since then, tried to base their models on PP findings on human subjects (e.g., Mayhew and Frisby 1981).

4.4. Shape from motion

Other examples of shape-from methods include shape from texture (Grimson 1981; Kanade and Kender 1983; Stevens 1980; 1984; Todd and Akerstrom 1987; Witkin 1981), shape from contour (Barrow and Tenenbaum 1981; Binford 1981; Ivry and Cohen 1987; Stevens 1981; 1986; Witkin 1981), shape from photometric stereo (Ikeuchi and Horn 1981; Woodham 1981), and shape from motion (Longuet-Higgins and Prazdny 1981; Prazdny 1980; Ullman 1979, 1983), the last example of which is now discussed into more detail. To recover the 3-D structure from the 2-D moving images, a perceiving organism (human or artifact) is confronted with the same critical problem of underdetermination as in the other cases: The same 2-D patterns can be caused by different 3-D structures.

Ullman (1979), analogously to previous solutions to similar problems, looked for a minimal set of constraining assumptions. He found that a unique 3-D form can be computed, if one assumes (i) a sufficient sample of the dynamic retinal pattern (i.e., three distinct views of four non-coplanar points), (ii) the possibility to determine which elements on the 2-D image arise from the same point on a moving 3-D stimulus across time (i.e., correspondence assumption), and (iii) the stimulus

elements being located on a rigid body, so that they maintain their relative 3-D distances (i.e., rigidity assumption).

This structure-from-motion theorem holds for both parallel or orthographic projection (i.e., perpendicular to the image plane) and the natural perspective projection of 3-D objects on the 2-D retina, because these are nearly identical under local analysis. Nevertheless, none of these simplifying assumptions seems justified, when the algorithm is proposed as a model for the human perceiver's ability to recover shape from motion. It is, for example, a well-established fact that human observers, contrary to what could be expected from the rigidity assumption (iii), sometimes see non-rigid motions, even when rigid motion can be derived from the input images, and that humans are able to detect non-rigid motions as accurate as rigid ones. Todd (1982) has, therefore, presented a geometric analysis that can account for both rigid and non-rigid motions and he designed some elegant experiments demonstrating the validity of his model.

Furthermore, Todd's model, contrary to Ullman's (i), does not regard the number of distinct frames as a critical variable for the recovery of 3-D shape from 2-D motion, an assumption that was clearly demonstrated to be invalid for human observers (e.g., Doner et al. 1984; Lappin et al. 1980). Another PP study casting doubts on the psychological relevance of Ullman's CV analysis of the perception of structure from motion, is the one of Todd (1985). He clearly demonstrated that projective correspondence of moving elements on the 2-D retina to identifiable moving points in 3-D space is not a necessary condition for the perception of structure from motion, which was assumed by Ullman (ii).

Although Ullman's (1979, 1983) adoption of the rigidity assumption was primarily motivated by an attempt to reduce the mathematical complexities involved in the structure-from-motion problem, it was also suggested by the human perceiver's tendency to prefer a rigid interpretation of moving elements (Gibson and Gibson 1957; Wallach and O'Connell 1953). Nevertheless, it is clear from the PP experiments mentioned above that the human visual system does not use the rigidity assumption in the same strict way as Ullman's (1979, 1983) previous CV studies have suggested.

Realizing this, Ullman (1984a) proposed a new method for deriving structure from motion, still based on a rigidity assumption, but in a more flexible way. According to the 'incremental recovery scheme', as

he has called it, an internal model of the 3-D structure is modified as the object moves with respect to the viewer. The transformations in the internal model, reflecting the changes in the environment, try to maximize the rigidity of the object by resisting changes in its shape as much as possible. Apart from the advantages that this scheme (note that Ullman (1984a), avoids the use of the notion 'algorithm' for his modified proposal) allows both rigid and non-rigid motion to be perceived, and that it provides a reliable recovery of structure in the presence of considerable amounts of visual noise in the images, it has other attributes that are consistent with human perceptual behavior, some of which are currently being tested (see, e.g., Hildreth and Hollerbach 1985; Ullman 1984a).

Ullman is not the only one having tried to solve the difficulties with his earlier method. Some researchers working in the same CV tradition look for other regularities of nature, different from the rigidity assumption, to base their model of interpretation of visual motion on, and they attempt to do this in better accordance with human perceptual abilities. Reuman and Hoffman (1986), for example, have started their study with the casual observation that even non-rigid motions are not completely arbitrary. For example, limbs in motion move in highly regular paths. The further part of their work consists of a mathematical specification of this regularity constraint (i.e., a planarity regularity), and a computer implementation of the algorithm embodying the constraint. Reuman and Hoffman (1986) have also shown that their recovery strategy works even in the face of noisy data, which is an important advantage when they would consider it as a model for the human recovery of shape from motion.

Other researchers working in a somewhat different tradition, namely PP, propose heuristic perceptual processes to derive depth from motion. According to Braunstein (1983), for example, these heuristics have important advantages over the use of more precise algorithms based on the rigidity assumption: They can easily account for the perception of non-rigid motion, they are easier to implement in a biological system, and they allow more rapid solutions, although some doubt on the latter proposals seems justified.

4.5. Conclusion

The cursory review of the current research on shape-from methods has illustrated that doing computational modeling in CV is one thing,

but accounting for human abilities in PP another. As Ramachandran (1985) has stressed recently, a full understanding of any visual process requires (at least) four questions being answered: (i) What is the information available in the visual scene?; (ii) which of the many possible sources of information does the visual system actually use?; (iii) what are the mechanisms by which it is possible to recover the information from the visual input?; and (iv) which of the many ways of recovering information is actually used by the visual system?

The first question clearly is one destined to be answered by ecological optics, a part of DTP (Gibson 1979). Regrettably, DTP has often forgotten to answer the second question (there are, however, some notable exceptions on this, e.g., Cutting 1986; Todd 1981, 1982). The third question concerns CV, but equally pity is the neglect by some CV workers of relevant PP findings concerning the human perceiving machine (i.e., the fourth question). Happily, however, these shortcomings have been adjusted in more recent years, when CV has done some efforts to adjust their algorithms in order to account as a CTP (being the theoretical part of CV and being part of PP, also).

It seems, thus, as if CTP must be considered a full-fledged theory, which promises to be a real breakthrough in PP, if, and only if, they take data on human subjects into account. In this way, CV would avoid what Grossberg (1983: 683) has called 'the unsettling confusion between means and ends, between wanting to understand human vision but hating to study human processes'. Or, would it be possible that other, fundamental criticisms on the CTP approach would cast some doubts on this?

5. (Meta)criticism on current CTP research

Since the approach taken in CTP is relatively new (late seventies), it can be expected to be largely inadequate and even wrong in detail. The crucial consideration, however, is that CTP is viable and can be progressively sharpened by recourse to experimental disconfirmation. Therefore, it is of vital importance to consider potential (meta)criticisms on the CTP approach as a whole, instead of particular details of some derived hypotheses. A division between two categories is made, relating to the relative weight of PP (criticism) and PhM (metacriticism).

5.1. Criticisms from PP

An essential property of a modular organization, according to Fodor (1983) is that the order and kind of computations within one module are not influenced by the ongoing or finished computations within another information processing module. The basic conjecture is, thus, that there is no information exchange between modules. Marr (1982) and other CTP scientists have concluded from this hypothesis that different recovery processes such as shape from shading, depth from stereo, and structure from motion, which are supposed to be modular, can be studied in relative isolation.

Others have some serious doubts about this possibility, ranging from general disapprovals of the whole approach to rather specific remarks, more or less based on PP research. Grossberg (1983: 683), as an example of the former, thinks 'the-independence-of-modules hypothesis is just a philosophical slogan to be used as a weapon at scientifically inconvenient moments'. Terzopoulos (1986), as an example of an intermediate position, admits that thinking of the early visual processes as a set of isolated modules ignores certain non-trivial interactions between the different processes, not the least of which is the combination or integration from multiple early visual processes. He, therefore, proposes an explicit account of this integration, namely a distributed and highly parallel computational process that accomplishes the integration cooperatively.

Cavanagh (1987), as an example of the latter, more empirically based criticism, demonstrated, in a series of experiments, that natural constraints do not always play a significant role in the recovery of depth from images defined by shading, binocular disparity, or motion. His findings imply the existence of a variety of mechanisms analyzing depth according to some very loose rules instead of precise algorithms, which is supported by others also (e.g., Braunstein 1983; Ramachandran 1985).

Furthermore, this variety of mechanisms challenges the plausibility of one general-purpose processing system that analyzes all aspects of depth in the image in the same manner (as is often supposed in classic ITP). Equally implausible is the alternative of a multitude of special-purpose modules (as part of CTP), because this would, according to Cavanagh (1987), entail a great deal of redundancy , namely duplicating similar processing mechanisms in many, if not all, of the

modules. He, therefore, argues for the less extreme alternative that all depth cues may be analyzed by a common set of processes which access different types of image data, depending on the cue in question.

In fact, these criticisms can be summarized as doubts about the relevance of the modular organization for human perceivers who do not always take natural constraints into account, who use a variety of some rather loose rules instead of rigid algorithms, and who have to integrate the 3-D information recovered from the 2-D images, somewhere. PP should, I think, direct its future research efforts to these presently neglected aspects of the CTP. It is clear that further studies with human perceivers are wanted to reveal which natural constraints are taken into account and how, the extent in which the recovery processes rely on algorithms or heuristics, and where exactly the integration of the different recovery processes takes place and how.

It is also clear, however, that none of the criticisms mentioned above seriously challenges the crucial assumption of independence of the modules. For it could be the case that the different recovery processes all work in parallel (whether they are heuristic rather than algorithmic remains an open question to be answered by empirical PP research), whereas their outputs are only integrated afterwards, so that information exchange during the initial phase would not be required. Indeed, this is the major tenet of Marr's view on the 2,5-D sketch as the representation where the information from the recovery modules is integrated (e.g., Marr 1982; Nishihara 1981; Stevens 1983).

5.2. Metacriticisms from PhM

5.2.1. Defining the problem

A still more fundamental criticism on the CTP approach as a whole is that it does not give an adequate explanation of the intentionality (i.e., the being-about or the aboutness) of perception. Representatives of PhM have argued that CTP has seriously neglected the intentional aspects of perception (such as meaning and reference), that are, nevertheless, the most crucial ones from the point of view of PhM. Perception is, indeed, the mind's window on the world, and, as such, it provides the link between the objective aspects of the world as it is and the subjective aspects of how it is experienced. A neglect of intentionality would, therefore, be a very serious shortcoming of the CTP.

This form of metacriticism (I call it 'meta' since it is not aimed at a specific hypothesis of the CTP, but at CTP in general) is not new. Indeed, for example, Searle (e.g., 1980) and Dreyfus (e.g., 1972) are sceptics of the early hour. However, the metacriticism on the lack of an account of intentionality in theories such as Marr's, is becoming more widespread (e.g., Calis 1984; Dodwell 1985; Russell 1984; Sayre 1986; Winograd and Flores 1986), and, for that reason, deserves renewed attention.

One could say that the bottom-up information processing in modules only results in form detection, that is, the establishing of a particular shape (e.g., cylindrical, cubical, etc., or some combination of these). Object and scene perception, that is, the recognition of a particular thing or situation (e.g., bottle, box, etc., or a supermarket), however, requires top–down influence of object-specific information. Somewhere in the perceptual process, a turning point between bottom–up and top–down, a transition from modular input systems to central cognitive systems must necessarily exist.

Marr (1982) and, in fact, the whole CTP approach is very silent about this (some notable exceptions are discussed further on). One usually does not go farther than stating the problem. It remains, therefore, an interesting problem to determine where bottom-up processing has come to an end and where top–down influences intrude, and how object recognition then takes place exactly. It might seem very odd to workers outside the PP field that object recognition remains an unknown part of the perceptual process, for it must be considered the crucial part of it. Nevertheless, we must be fair in admitting that very little is known about object recognition indeed. All existent theories (ITP, DTP, and CTP) prove to be unsatisfactory.

5.2.2. Refining the problem

PhM is extremely well-suited for helping to refine the problem of the difference between form detection and object perception. As is often the case with the refining of problems, some distinctions must be made, viz. between 'seeing' and 'seeing as', projectable and non-projectable properties, referential transparency and opacity, and extensional and intensional contexts, a cluster of closely interrelated concepts, hard to define exactly. The interdependence of the notions will be demonstrated by means of some examples.

A first distinction that must be made, according to PhM (e.g., Fodor

and Pylyshyn 1981), is the one between 'seeing' and 'seeing as'. For example, if you just *see* the murderer of your wife, you would not necessarily kill or congratulate him (the choice between them depending, among other things, on your view on the quality of the relation with your wife). The retaliation of the murderer or the expression of fecilitations would, at least, require the *seeing* of a certain person *as* your wife's killer. To predict the cognitive and behavioral activity of a human subject on the basis of his or hers perception (which is a crucial task of scientific psychology in general), implies PP giving an adequate account of how a perceiver gets from seeing to seeing as, an account that is clearly missing in CTP, according to its critics.

Second, a distinction is made between projectable properties, i.e., properties involved in physical laws, and non-projectable properties, i.e., that cannot function in nomological propositions. Projectable properties such as 'gold' can simply be detected, whereas non-projectable ones such as 'my girlfriend's favorite metal' necessarily require to be perceived with the help from higher level knowledge. There is, for example, a difference in stating that 'my girlfriend was looking at some piece of gold, without realizing what she was looking at (because of some dust on it, probably)' and stating that 'she saw her favorite metal with a little twinkle in her eyes'. (I would prefer the first, I think, because I know what time it is when the second event occurs: time to get some hard cash from my bankaccount.)

Third, referential opacity and transparency are distinguished. When information processing cannot be influenced by changing the non-projectable aspects of the input, the process involved is called referentially opaque. For example, I always flicker my eyes, when you fastly move your finger in my eye's direction, although you may tell me hundreds and thousands of times that you are a nice fellow and never even hurt a fly. When, on the other hand, changing the non-projectable aspects of the input has a demonstrable influence on a process, this process is called referentially transparent. There will be, for example, a change in the way I perceive my neighbour, when somebody, some day, would tell me that he is, actually, 'Jack the Ripper'. The truth value of the sentence 'I like my neighbour' would, probably, change when I would be given this information. Detection is supposed to be referentially opaque, whereas perception is considered referentially transparent.

Fourth, referential opacity occurs in extensional contexts, transparency in intensional ones. Examples of extensional properties are

fluidity/solidity, occluding/occluded, 2-D/3-D, moving in relation to a fixed point or not, supporting or not, etc. Intensional properties are, for instance, being drinkable, edible, throwable, being bottles, boxes, walls, etc. (see, e.g., Russell 1984). Extensional properties can be detected bottom–up, whereas intensional ones require top–down influence to be recognized or perceived as such.

In summary, detection or 'seeing' of projectable or extensional properties is referentially opaque and can be explained by mere bottom–up processing of the input, relying on 'knowledge' of general constraints in the world, present and used in a modular recovery process. Perception or 'seeing as' of non-projectable or intensional properties is referentially transparent and must entail top–down influence of object-specific knowledge. This cluster of concepts helps to clarify some of the contrasts between ITP, DTP, and CTP.

On the one hand, classic ITP have always tried to explain perception of intensional properties by stressing top–down influences from memory, expectation, motivation, emotion, etc. But, excessive forms of ITP, such as New Look theories (e.g., Bruner 1957) and theories of perception as problem-solving (e.g., Gregory 1970), have too often forgotten to give an account of the detection of the input on the basis of which higher level processes can operate. In other words, ITP have, generally, neglected the bottom–up processing of extensional properties.

According to DTP, on the other hand, intermediating processes are never necessary. Properties such as sit-on-ability, climb-on-ability, fall-of-ability, etc., 'affordances' as they are called (see, e.g., Gibson 1979), are all directly perceived. Gibson and his ecological associates have, in this manner, classed these clearly intensional properties under the heading of extensional properties, that can be picked-up directly. As a consequence, they need not have any theory of how to come from detected extensional properties to perceived intensional properties.

Marr's CTP theory has, in my opinion, given a fairly adequate account of the first part of the perceptual process, viz. the bottom–up detection of extensional properties. In its current state, however, CTP has not (yet) paid sufficient attention to the necessary second part, the recognition of intensional properties which requires top–down influence. The metacriticism of some representatives of PhM (e.g., Russell 1984; Sayre 1986) is that CTP is not only lacking that part of a full-fledged perceptual theory now, but that CTP will, in principle, never be able to explain the capturing of the meaning of a stimulus, or,

in other words, the intensional properties of it ('intensional' with an 's'), or, in still other words, the crucial aspect of the intentionality of perception ('intentionality' with a 't').

It is my opinion, however, that we should, at least, give CTP a fair change to try, since there is, I think, no logical necessity that a scientific theory of recognition could not be possible in a CTP framework (when they are prepared to admit that not all processing is bottom–up, of course). ITP and DTP have had (and are, in fact, still having) a huge amount of research time and money to puzzle about it, so, why should not CTP receive the same opportunity? (Some) CTP workers realize the shortcomings of their theories and, in the next section, it will be demonstrated that they are seriously doing some efforts to improve their models by trying to bridge the gap between detection and recognition.

5.3. CTP attempts to overcome the (meta)criticisms

The connection between the (PP) criticism on the relative independence of the recovery modules and the (PhM) metacriticism on the neglect of intentionality, is apparent from the stating of the problems as well as from the attempts to find solutions to them. Concerning the connection between the definition of the problems, for example, Calis (1984: 212) clearly states that 'the contemporary working programs seem to be too specialized modules, which are moreover too linear (...), to be intentional themselves. Perhaps, it only makes some sense to talk of intentional aspects of these working modules if they are viewed as a part of a larger process.' This solution is exactly the one CTP has tried to attain with their recent proposals.

Ullman (1984b), for example, makes a distinction between the bottom–up creation of early representations of the visible environment, such as the primal sketch and the 2,5-D sketch, and the subsequent top–down application of, what he calls, 'visual routines' to the representations constructed in the first stage. The early visual representations, on the one hand, are fixed and unchanging (i.e., always the same properties are represented, e.g., surface orientation, depth, and direction of motion), unarticulated (i.e., essentially local representations of these properties), spatially uniform (i.e., the same properties are extracted and represented across the visual field), viewer-centered (i.e., not with respect to the environmental coordinates), and bottom–up

driven (i.e., representations depend on the visual input alone and on the 'knowledge' within the module).

The further representations, needed to attain object and scene recognition, on the other hand, are open-ended (i.e., the extraction of newly defined properties and relations is permitted), articulated (i.e., more globally attained), spatially non-uniform (i.e., the computations by the visual routines are not applied uniformly over the visual field), object-centered (i.e., with respect to the environmental coordinates), and top–down driven (i.e., for the same visual input different aspects will be made explicit at different times, depending, therefore, not only on the input, but also on the goals of the computation and on object-specific knowledge).

Visual routines, proposed as the computational processes to establish these later representations, are composed of sequences of elemental operations. Depending upon the task at hand, the visual system can assemble different routines from a fixed set of basic operations to extract an unbounded variety of shape properties and spatial relations, needed to attain recognition. Although the details of Ullman's (1984b) theory remain to be worked out, the general framework of his proposal is quite clear: The first stage of visual information processing is the bottom–up creation of the early representations, on which, in the next stage, visual routines are applied. In the absence of specific expectations or prior knowledge, universal routines are applied first, followed by the selective application of specific routines. Intermediate results obtained by these visual routines are summarized in a kind of incremental representation which can be used by subsequent routines.

This theory is an explicit attempt to give, within the CTP approach, an adequate explanation of what happens between detection and recognition, and as such, it deserves attention. Patience will be needed to see what kind of empirical research will be sparked off by this thesis. Also, one has to look for similar attempts in the same direction, which do, in fact, exist in the current CTP literature.

For example, Pentland (1986c) has proposed a new kind of representation format, because the currently available ones (e.g., Marr's), according to him, have been developed for purposes other than the ones required for a theory of how the visual system produces meaningful descriptions of 3-D objects and scenes on the basis of a 2-D array of image intensities. The requirements for representations as part of an adequate CTP theory are that the elements of the representation are

lawfully related to important physical regularities, and that the representation corresponds with the perceptual organization a human perceptual system imposes on the stimulus, whereas the existent representational formats have been developed for other purposes (viz. physics or engineering).

Pentland (1986c), therefore, argues for a representation of the scene structure at a scale that is similar to the naive notion of a perceptual part and he claims to have found that format in 'fractals', i.e., adequate geometrical descriptions of an extensive variety of both regular and irregular natural forms such as clouds, mountains, coasts, trees, etc. Indeed, it is now being discovered that these divergent forms are also constrained by physical laws to a limited number of basic patterns, although not by the generally known laws of physics and material science (see, e.g., Mandelbrot 1982; Stevens 1974; D'Arcy Thompson 1942).

Fractal-based descriptions are intermediate-level representations, explicitly suggested in order not to be forced to bridge the gap between the initial low-level representations based on general models of image formation and later high-level representations based on object-specific models of boxes or bottles. The details of this theory, again, are not very well articulated, but, as in Ullman's case, the general idea is rather straightforward: Fractal-based descriptions are, in fact, models of parts, lawfully associated with image regularities as well as with parts discovered by perceptual organization. The task of the perceptual system is, then, to recognize the content of the image as a combination of these fractals. Other CV workers equally stress the importance of parts as an extremely relevant notion in a CTP theory of human perceptual organization and recognition (e.g., Hoffman 1983; Hoffman and Richards 1984). Also, from within PP, Biederman (1985, 1987) has recently proposed a theory of object recognition on the basis of component parts.

5.4. Conclusion

Although a considerable amount of derived hypotheses and general suggestions of CTP theories are demonstrated either to be false (by PP research) or to be inadequate (by PhM scrutiny), the hard core of the research program (see Lakatos 1970) is not yet blown up. As is apparent from the recent studies mentioned above, CTP scientists are

still situated within the general approach instantiated by Marr. As long as this is the case for the largest part of the CS community working on vision, there is no need to leave this kind of approach, as has been argued for recently. Some, for example, defend another source of inspiration on which to base a perception theory, namely information theory (e.g., Leyton 1986a and b; Sayre 1986). Others make a plea for a general reconsideration of the human subject, who has a completely different 'Dasein' and who can, therefore, not even be compared with a computational system, let alone be considered to be one (e.g., Dreyfus 1972; Russell 1984; Winograd and Flores 1986). I, however, argue that the research program of CTP, as a whole, is still alive and well, no matter how affected some of its body parts might be from criticisms and metacriticisms of PP and PhM.

6. Other possible interactions between CV, PP, and PhM

Apart from the mutual interchanges between the three constituent branches of CV research on vision, that have, hopefully, become evident from the discussion so far, a number of other interactions that are currently insufficiently explored, might be possible and fruitful for scientific progress. All members of the CS triumvirate have their own task they need to take up.

6.1. Tasks of CV and PP

CV with theoretical aspirations (i.e., CTP wanting to account for human perception) necessarily has to rely on PP studies on human subjects. Not only, as has been the case in current research mostly, as the possible refutation or corroboration of the predictions on algorithms and representations used, but, also, to find some inspiration for new suggestions of computational processes.

For example, concerning the finding of contours and their curvatures (a problem that is related to Marr's primal sketch), current CV work is in trouble about the so-called locality problem (e.g., Asada and Brady 1986; Fischler and Bolles 1986). Mathematically, curvature is determined on an infinitesimally small surrounding. Practically, however, one has to rely on a finite piece of the contour. CV, I argue, should rely their solutions of the locality problem (i.e., how local is local?) on the

ones suggested by experimentally established findings on how human perceivers solve it. Do people take pieces of a well-determined length (probably depending on the sensitivity of the retinal receptors), or is the length of the contour part maybe dependent on the size of the whole or on the importance of the part in the whole?

Another example of PP research possibly relevant to CV, concerns the use of physical constraints of color mixing for the recovery of 3-D shape from 2-D patterns (a problem related to Marr's 2,5-D sketch). In an experiment reported elsewhere (Wagemans and de Weert 1987), we designed some colored variants of the 'Necker cube' (in fact, transparent boxes) and we tried to disambiguate the 2-D patterns by choosing the colors of the different parts such that there was only one 3-D interpretation physically possible. By forcing the subjects to choose which orientation of the box they were seeing, we were able to demonstrate that human perceivers incorporate some 'knowledge' on physical constraints of color mixing, since they preferred, to a statistically significant extent, the physically possible variants of the boxes. Nevertheless, CV researchers seldomly use color information as an input and they have never, as far as I know, suggested algorithms that can perform the recovery of 3-D shape of 2-D colored images on the basis of models of color-mixing laws.

In conclusion, CV researchers should not only try to publish their vision models in the PP literature, because they consider them relevant to it (and I demonstrated they are indeed), but they should also consult that literature to base their CTP models on, because, in my opinion, the possible reversed transfer (from PP to CV) is as relevant and fruitful as the currently dominant one (from CV to PP).

6.2. Tasks of PhM

There are three goals for philosophy in general and PhM in particular, being considered as a potentially valuable member of the CS triumvirate studying vision: (i) to give an adequate characterisation of the explicanda of (i.e., what needs to be explained by) CTP (as part of both CV and PP); (ii) to provide fundamental criticisms on CTP for not having reached a full-fledged account of these explicanda; and (iii) to inspire the formation and canalizing of the interpretation of CV and PP data that are all the same relevant to the theories of the explicanda.

Some of these tasks (foremost, ii and iii) have been tackled already, others (foremost, i) might better deserve somewhat more attention. As

was probably apparent from the discussion in 5.2.2., much remains to be done in explaining pairs of terms such as intensional/extensional, projectable/non-projectable, referentially opaque/transparent, etc. and the relations between them. Also, PhM should give a clear account of intentionality, usable by PP and CV. Furthermore, PhM should elaborate on the modularity thesis (which is an excellent example of doing iii) and all the problems that are related to it: for example, differences between bottom–up and top–down processing, general and specific knowledge, implicit and explicit information, 'knowledge' and knowledge, etc. Although PhM, especially the one in the analytical tradition has a rather extended experience on these topics, the CS community in general, and research on vision in particular, would clearly benefit from further work along these lines, that is accessible for non-philosophers such as PP and CV scientists.

Related to objective (iii) of PhM, is the interesting task of integrating data relevant to the general mind-body problem. Clearly, PP and CV studies on vision could be of extreme relevance to this age-long fundamental puzzle (or mystery). Recently, some steps in this direction were set (e.g., Pribram 1986; Thagard 1986), but, surely, there is a lot that remains to be done.

7. Summary of both existent and possible interactions in CS

In 1978, Pylyshyn could, in rather general terms, speak of the courtship between AI and (experimental) psychology, 'a loose but symbiotic relation in which each supplies a source of heuristic inspiration and ideas to the other' (Pylyshyn 1978: 99). Now, in 1987, we are confronted with the marriage between the two (CS) and, indeed, the children (CTP). It was, therefore, necessary to consider how this happened (without going into intimate details, of course) and what relations and interactions can be specified exactly. Influences that were implicitly present in the forgoing exposition, are now summed up explicitly. (Numbers refer to arrows in fig. 1, representing influences from one discipline on another, or from one on two others or vice versa.)

(1) The influence of CV on PP. CV has an extensive impact on theorizing in PP, namely the rise of CTP as a third alternative, as well

as on the experimental research with human perceivers, viz. studies by Todd and others, explicitly designed as tests of CV predictions or derived hypotheses.

(2) The influence of PP on CV. In historical perspective, Gibson (1979) has had a serious role in the breakthrough of CV by pointing at the richness of the information available in the natural world. In current research, PP experimental results, established by Todd and others, are given serious attention by CV workers, who try to improve their models by taking these corroborating or refuting findings into account. In future studies, CV should also search for PP findings as a source of inspiration when there is a lack of knowledge on constraints to algorithms and representations used by humans.

(3) The influence of PP on PhM. Important are the PP data, which PhM needs to take into account. This has been done by, for instance, Fodor, when theorizing about the modularity of mind, and by others, when thinking about the mind–body problem.

(4) The influence of PhM on PP. This PhM theorizing (e.g., Fodor's) is influencing PP research at present and will continue to do so in the future. Furthermore, PhM can help to distinguish terms as detection and recognition, vitally important for PP theories.

(5) The influence of PhM on CV. Fodor's modularity-thesis is very influential in current CV work. Furthermore, the metacriticisms of PhM on CV in general are not going unnoticed. Recent CV scientists (e.g., Ullman, Pentland) are seriously taking them into consideration.

(6) The influence of CV on PhM. The other way around, Marr can be regarded as one of the instigators of Fodor's view on modularity. Also, the notion of 'module' is one that is being frequently used in structured programming languages in AI. Finally, representatives of PhM can be supposed to know the CV literature quite well, since they write a lot about it.

(7) The influence of CV on PP–PhM. CV has provided PP with a third possible theory, the status of which is discussed in close interaction with PhM.

(8) The influence of PP–PhM on CV. Both the criticisms of PP research and the metacriticisms based on PhM scrutiny (which are closely interrelated, of course) are clearly influencing CV as a discipline. As a consequence, CV noticed their relevance to PP as well as their current lack of a well-established theory of intentionality.

(9) The influence of PP on CV–PhM. PP research has, for example, demonstrated that not all aspects of the computational modules are applicable to the human recovery process. Both CV and PhM have to work out which aspects of modularity in vision do and which do not apply to human perception.

(10) The influence of CV–PhM on PP. The other way around, PP is able to make a lot of concrete predictions about human recovery processes, that are due to the fruitful interaction between CV and PhM on the modularity of low-level vision.

(11) The influence of PhM on CV–PP. PhM (or, better: the philosophy of science part of it) is very useful in making explicit the different scientific status (e.g., goals and evaluation criteria) of CV and PP, and can, therefore, be very helpful in determining the exact relation between the two. Furthermore, the PhM thesis about modularity is clarifying a lot about CTP (part of CV and PP) as being a compromise between ITP and DTP. Third, PhM should attempt to define the explicanda of CV and PP, and it should try to refine the terms needed in CV and PP, in such a way that they are understandable for non-philosophers. Finally, the PhM metacriticisms will surely stimulate further research along the CTP line, and it will, in this manner, promote scientific progress in CS as a whole.

(12) The influence of CV–PP on PhM. The empirical findings of both CV and PP are important data to integrate in PhM. These theoretical constructions can range from rather particular theses (such as modularity) to quite general speculations about the mind–body problem. Apart from integrating, PhM is also very busy criticizing the work in CV and PP, and, especially, the CTP as a result of their interaction.

References

Arbib, M.A., 1987. Levels of modelling of mechanisms of visually guided behavior. Behavioral and Brain Sciences, in press.

Asada, H. and M. Brady, 1986. The curvature primal sketch. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8, 2–14.

Ballard, D.H., 1986. Cortical connections and parallel processing: structure and function. Behavioral and Brain Sciences 9, 67–120.

Ballard, D.H. and C.M. Brown, 1982. Computer vision. Englewood Cliffs, NJ: Prentice-Hall.

Ballard, D.H., Hinton, G.E. and T.J. Sejnowski, 1983. Parallel visual computation. Nature 306, 21–26.

Barrow, H.G. and J.M. Tenenbaum, 1981. Interpreting line drawing as three-dimensional surfaces. Artificial Intelligence 17, 75–116.

Beck, J., 1972. Surface color perception. Ithaca, NY: Cornell University Press.

Beck, J., 1975. The perception of surface color. Scientific American 233, 62–75.

Beck, J., Hope, B. and A. Rosenfeld, 1983. Human and machine vision. New York, NY: Academic Press.

Biederman, I., 1985. Human image understanding: recent research and a theory. Computer Vision, Graphics, and Image Processing 32, 29–73.

Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. Psychological Review 94, 115–147.

Binford, T.O., 1981. Inferring surfaces from images. Artificial Intelligence 17, 205–244.

Braddick, O.L., 1980. Direct perception: an opponent and a precursor of computational theories. Behavioral and Brain Sciences 3, 381–382.

Brady, M., 1982. Computer vision. Amsterdam: North-Holland.

Braunstein, M.., 1983. 'Contrasts between human and machine vision: should technology recapitulate phylogeny?' In: J. Beck, B. Hope and A. Rosenfeld (eds.), Human and machine vision. New York, NY: Academic Press. pp. 85–96.

Brown, C.M., 1984. Computer vision and natural constraints. Science 224, 1299–1305.

Bruner, J., 1957. On perceptual readiness. Psychological Review 64, 123–152.

Cavanagh, P., 1987. Reconstructing the third dimension: interactions between color, texture, motion, binocular disparity and shape. Computer Vision, Graphics, and Image Processing 37, 171–195.

Calis, G., 1984. Concerning Gibson's 'on the face of it': immediate perception and single-glance face recognition. Acta Psychologica 55, 195–214.

Chomsky, N., 1980. Rules and representations. Behavioral and Brain Sciences 3, 1–62.

Clowes, M.B., 1971. On seeing things. Artificial Intelligence 2, 79–116.

Cohen, P.R. and E.A. Feigenbaum, 1982. 'Vision'. In: The handbook of artificial intelligence, vol. 3, Los Altos, CA: Kaufmann. pp. 125–321.

Cutting, J.E., 1986. Perception with an eye for motion. Cambridge, MA: MIT Press/Bradford Books.

Davis, L.S. and A. Rosenfeld, 1981. Cooperating processes for low-level vision. Artificial Intelligence 17, 245–263.

Dodwell, P.C., 1985. Theories of perception as experimental epistemology. Behavioral and Brain Sciences 8, 291–293.

Doner, J., Lappin, J. and G. Perfetto, 1984. The detection of three-dimensional structure in moving patterns. Journal of Experimental Psychology: Human Perception and Performance 10, 1–11.

Dreyfus, H.L., 1972. What computers can not do: a critique of artificial intelligence. New York, NY: Harper & Row.

Feldman, J.A., 1985. Four frames suffice: a provisional model for vision and space. Behavioral and Brain Sciences 8, 265–289.

Fischler, M.A. and R.C. Bolles, 1986. Perceptual organization and curve partitioning. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8, 100–105.

Fodor, J.A., 1980. Methodological solipsism considered as a research strategy in cognitive psychology. Behavioral and Brain Sciences 3, 63–109.
Fodor, J.A., 1983. The modularity of mind. Cambridge, MA: MIT Press/Bradford Books.
Fodor, J.A., 1985. Précis of 'The Modularity of Mind'. Behavioral and Brain Sciences 8, 1–42.
Fodor, J.A. and Z.W. Pylyshyn, 1981. How direct is visual perception? Some reflections on Gibson's 'Ecological Approach'. Cognition 9, 139–196.
Gibson, J.J., 1979. The ecological approach to visual perception. Boston, MA: Houghton Mifflin.
Gibson, J.J. and E.J. Gibson, 1957. Continuous perspective transformations and the perception of rigid motion. Journal of Experimental Psychology 54, 129–138.
Gregory, R.L., 1970. The intelligent eye. New York, NY: McGraw-Hill.
Grimson, W.E.L., 1981. From images to surfaces: a computational study of the early visual system. Cambridge, MA: MIT Press.
Grossberg, S., 1983. The quantized geometry of visual space: the coherent computation of depth, form, and lightness. Behavioral and Brain Sciences 6, 625–692.
Guzman, A., 1968. Computer recognition of three-dimensional objects in a visual scene. Doctoral dissertation, MAC-TR-59, Project MAC, MIT, Cambridge, MA.
Haber, R.N., 1983. 'Stimulus information and processing mechanisms in visual space perception'. In: J. Beck, B. Hope and A. Rosenfeld (eds.), Human and machine vision. New York, NY: Academic Press. pp. 157–235.
Hanson, A.R. and E.M. Riseman (eds.), 1978. Computer vision systems. New York, NY: Academic Press.
Helmholtz, H. von, 1910/1962. Treatise on physiological optics. New York, NY: Dover.
Hildreth, E.C. and J.M. Hollerbach, 1985. The computational approach to vision and motor control, AI Memo 846, MIT, Cambridge, MA.
Hinton, G.E., 1980. Inferring the meaning of direct perception. Behavioral and Brain Sciences 3, 387–388.
Hochberg, J.E., 1981. On cognition in perception: perceptual coupling and unconscious inference. Cognition 10, 127–134.
Hochberg, J.E., 1986. 'Representation of motion and space in video and cinematic displays'. In: R.K. Boff, L. Kaufman and J.P. Thomas (eds.), Handbook of perception and human performance: Sec. 4. Space and motion perception: Vol. 1. Sensory processes and perception. New York, NY: Wiley. pp. 22-1–22-64.
Hoffman, D.D., 1983. The interpretation of visual illusions. Scientific American 249, 137–144.
Hoffman, D.D. and W.A. Richards, 1984. Parts of recognition. Cognition 18, 65–96.
Horn, B.K.P., 1977. Understanding image intensities. Artificial Intelligence 8, 201–231.
Huffman, D.A., 1971. Impossible objects as nonsense sentences. In: B. Meltzer and D. Mitchie (eds.), Machine intelligence, vol. 6. New York, NY: Halsted. pp. 295–323.
Ikeuchi, K. and B.K.P. Horn, 1981. Numerical shape from shading and occluding boundaries. Artificial Intelligence 17, 141–184.
Ivry, R.B. and A. Cohen, 1987. The perception of doubly curved surfaces from intersecting contours. Perception & Psychophysics 41, 293–302.
Julesz, B., 1971. Foundations of cyclopean perception. Chicago, IL: University of Chicago Press.
Kanade, T., 1981. Recovery of the three-dimensional shape of an object from a single view. Artificial Intelligence 17, 409–460.
Kanade, T. and J.R. Kender, 1983. 'Mapping image properties into shape constraints: skewed symmetry, affine- transformable patterns, and the shape-from-texture paradigm'. In: J. Beck, B. Hope and A. Rosenfeld (eds.), Human and machine vision. New York, NY: Academic Press. pp. 237–257.
Kanizsa, G., 1985. Seeing and thinking. Acta Psychologica 59, 23–33.
Koenderink, J., 1987. Van neurowetenschap naar artificiële intelligentie [From neuroscience to artificial intelligence]. Lecture given in Leuven.

Lakatos, I., 1970. 'The methodology of scientific research programmes'. In: I. Lakatos and A. Musgrave (eds.), Criticism and the growth of knowledge. Cambridge, UK: Cambridge University Press.

Lappin, J.S., Doner, J.F. and B.L. Kottas, 1980. Minimal conditions for the visual detection of structure from motion in three dimensions. Science 209, 717–719.

Leyton, M., 1986a. A theory of information structure: I. General principles. Journal of Mathematical Psychology 30, 103–160.

Leyton, M., 1986b. A theory of information structure: II. A theory of perceptual organization. Journal of Mathematical Psychology 30, 257–305.

Longuet-Higgins, H.C. and K. Prazdny, 1981. The interpretation of moving retinal images. Proceedings of the Royal Society of London B208, 385–390.

Mandelbrot, B.B., 1982. The fractal geometry of nature. San Francisco, CA: Freeman.

Marr, D., 1982. Vision: a computational investigation into the human representation and processing of visual information. San Francisco, CA: Freeman.

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

Marr, D. and T. Poggio, 1979. A computational theory of human stereo vision. Philosophical Transactions of the Royal Society of London B290, 199–218.

Mayhew, J.E.W. and J.P. Frisby, 1976. Rivalrous texture stereograms. Nature 264, 53–56.

Mayhew, J.E.W. and J.P. Frisby, 1981. Psychophysical and computational studies towards a theory of human stereopsis. Artificial Intelligence 17, 349–385.

McArthur, D.J., 1982. Computer vision and perceptual psychology. Psychological Bulletin 92, 283–309.

McClelland, J.L. and D.E. Rumelhart, 1985. Parallel distributed processing: explorations in the microstructure of cognition. Vol. 2: Psychological and biological models. Cambridge, MA: MIT Press/Bradford Books.

Nishihara, H.K., 1981. Intensity, visible-surface, and volumetric representation. Artificial Intelligence 17, 265–284.

Norman, D.A., ed., 1981. Perspectives on cognitive science. Norwood/Hillsdale, NJ: Ablex/Erlbaum.

Palmer, S.E., 1983. 'The psychology of perceptual organization: a transformational approach'. In: J. Beck, B. Hope and A. Rosenfeld (eds.), Human and machine vision. New York, NY: Academic Press. pp. 269–339.

Pentland, A.P., 1982. Finding the direction of illumination. Journal of the Optical Society of America 72, 448–455.

Pentland, A.P., 1986a. From pixels to predicates: recent advances in computational and robotic vision. Norwood, NJ: Ablex.

Pentland, A.P., 1986b. Shading into texture. Artificial Intelligence 29, 147–170.

Pentland, A.P., 1986c. Perceptual organization and the representation of natural form. Artificial Intelligence 28, 293–331.

Prazdny, K., 1980. Egomotion and relative depth map from optical flow. Biological Cybernetics 36, 87–102.

Pribram, K.H., 1986. The cognitive revolution and mind/brain issues. American Psychologist 41, 507–520.

Pylyshyn, Z.W., 1978. Computational models and empirical constraints. Behavioral and Brain Sciences 1, 93–127.

Pylyshyn, Z.W., 1980. Computation and cognition: issues in the foundations of cognitive science. Behavioral and Brain Sciences 3, 111–169.

Pylyshyn, Z.W., 1984. Computation and cognition: towards a foundation for cognitive science. Cambridge, MA: MIT Press/Bradford Books.

Ramachandran, V.S., 1985. Guest editorial in special issue on human motion perception. Perception 14, 97–103.

Reuman, S. and D. Hoffman, 1986. 'Regularities of nature: the interpretation of visual motion'. In: A.P. Pentland (ed.), From pixels to predicates. Norwood, NJ: Ablex. pp. 201–226.
Richards, W.A., 1977. Stereopsis with and without monocular cues. Vision Research 17, 967–969.
Rock, I., 1977. 'In defense of unconscious inference'. In: W. Epstein (ed.), Stability and constancy in visual perception: mechanisms and processes. New York, NY: Wiley. pp. 321–373.
Rock, I., 1983. The logic of perception. Cambridge, MA: MIT Press/Bradford Books.
Rosenfeld, A., 1984. Image analysis: problems, progress, and prospects. Pattern Recognition 17, 3–12.
Rosenfeld, A., Hummel, R.A. and S.W. Zucker, 1976. Scene labeling by relaxation operations. IEEE Transactions on Systems, Man, and Cybernetics SMC-6, 420–433.
Rumelhart, D.E. and J.L. McClelland, 1985. Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1: Foundations. Cambridge, MA: MIT Press/Bradford Books.
Russell, J., 1984. Explaining mental life: some philosophical issues in psychology. London: MacMillan.
Sabbah, D., 1985. Computing with connections in visual recognition of Origami objects. Cognitive Science 9, 25–50.
Sayre, K.M., 1986. Intentionality and information processing: an alternative model for cognitive science. Behavioral and Brain Sciences 9, 121–166.
Searle, J.R., 1980. Minds, brains, and programs. Behavioral and Brain Sciences 3, 417–457.
Shepard, R.N., 1984. Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. Psychological Review 91, 417–447.
Stevens, K.A., 1980. The information content of texture gradients. Biological Cybernetics 42, 95–105.
Stevens, K.A., 1981. The visual interpretation of surface contours. Artificial Intelligence 17, 47–73.
Stevens, K.A., 1983. False dilemmas: confusion between mechanism and computation. Behavioral and Brain Sciences 5, 675.
Stevens, K.A., 1984. On gradients and texture 'gradients'. Journal of Experimental Psychology: General 113, 217–220.
Stevens, K.A., 1986. Inferring shape from configurations of contours across surfaces. In: A.P. Pentland (ed.), From pixels to predicates. Norwood, NJ: Ablex. pp. 93–110.
Stevens, P.S., 1974. Patterns in nature. Boston, MA: Little, Brown & Co.
Terzopoulos, D.W., 1986. 'Integrating visual information from multiple sources'. In: A.P. Pentland (ed.), From pixels to predicates. Norwood, NJ: Ablex. pp. 111–147.
Thagard, P., 1986. Parallel computation and the mind–body problem. Cognitive Science 10, 301–318.
Thompson, D'Arcy W., 1942. On growth and form (2nd ed.). Cambridge, UK: Cambridge University Press.
Todd, J.T., 1981. Visual information about moving objects. Journal of Experimental Psychology: Human Perception and Performance 7, 795–810.
Todd, J.T., 1982. Visual information about rigid and non-rigid motion: a geometric analysis. Journal of Experimental Psychology: Human Perception and Performance 8, 238–252.
Todd, J.T., 1985. Perception of structure from motion: is projective correspondence of moving elements a necessary condition? Journal of Experimental Psychology: Human Perception and Performance 11, 689–710.
Todd, J.T. and R.A. Akerstrom, 1987. Perception of three-dimensional form from patterns of optical texture. Journal of Experimental Psychology: Human Perception and Performance 13, 242–255.
Todd, J.T. and E. Mingolla, 1983. Perception of surface curvature and direction of illumination

from patterns of shading. Journal of Experimental Psychology: Human Perception and Performance 9, 583–595.
Ullman, S., 1979. The interpretation of visual motion. Cambridge, MA: MIT Press.
Ullman, S., 1983. 'Recent computational studies in the interpretation of structure from motion'. In: J. Beck, B. Hope and A. Rosenfeld (eds.), Human and machine vision. New York: NY: Academic Press. pp. 459–480.
Ullman, S., 1984a. Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion. Perception 13, 255–274.
Ullman, S., 1984b. Visual routines. Cognition 18, 97–159.
Ullman, S. and W.A. Richards (eds.), 1984. Image understanding 1984. Norwood, NJ: Ablex.
Wagemans, J.P., 1986. Direct theory of perception: an evaluation by representatives of indirect theories of perception. L'Année Psychologique 86, 261–273.
Wagemans, J.P., 1987. Schemas and bridging gaps in the behavioral and brain sciences. Behavioral and Brain Sciences, in press.
Wagemans, J.P. and C.M.M. de Weert, 1987. Shape from surface color: incorporating physical constraints of color-mixing for the recovery of 3-D structure (Psych. Rep. No. 65). Leuven: University of Leuven.
Wallach, H. and D.N. O'Connell, 1953. The kinetic depth effect. Journal of Experimental Psychology 45, 205–217.
Waltz, D., 1975. 'Understanding line drawings of scenes with shadows'. In: P.H. Winston (ed.), The psychology of computer vision. New York, NY: McGraw-Hill. pp. 19–91.
Warren, W.H. and R.E. Shaw (eds.), 1985. Persistence and change: proceedings of the first international conference on event perception. Hillsdale, NJ: Erlbaum.
Winograd, T. and F. Flores, 1986. Understanding computers and cognition: a new foundation for design. Norwood, NJ: Ablex.
Winston, P.H. (ed.), 1975. The psychology of computer vision. New York, NY: McGraw-Hill.
Winston, P.H. and R.H. Brown (eds.), 1979. Understanding vision. In: Artificial intelligence: an MIT perspective, vol. 2. Cambridge, MA: MIT Press. pp. 1-208.
Witkin, A.P., 1981. Recovering surface shape and orientation from texture. Artificial Intelligence 17, 17–45.
Woodham, R.J., 1981. Analyzing images of curved surfaces. Artificial Intelligence 17, 117–140.
Zucker, S.W., 1980. The computational/representational paradigm as normal science: further support. Behavioral and Brain Sciences 3, 406–407.