

The diagonal axis is therefore being constructed during the analysis of this pattern; it influences and therefore probably precedes the description of the shapes of the local elements.

Third, Warrington and Taylor (1973) drew attention to the difficulty experienced by their patients with right parietal lesions in interpreting certain views of common objects, which Warrington and Taylor called unconventional views. For example, these patients would fail to recognize the top view of a bucket (Figure 5-9c), denying that it was a bucket even when told it was. The patients were relatively unimpaired on views like Figure 5-9(a). As Warrington and Taylor pointed out, this difference cannot be easily explained in terms of familiarity or impaired depth perception, because both views of a bucket are common and depth is just as important to the three-dimensional structure of Figure 5-9(a) as it is to that of Figure 5-9(c). However, if the internal shape representation used for recognition was based on a shape's natural axes, the second figure would be more difficult to describe correctly, since its major axis is foreshortened. If this explanation were correct, Warrington and Taylor's unconventional views would correspond to views in which an important natural axis of the shape is foreshortened in the image, making it difficult for the patient to discover or derive a description in the shape's canonical coordinate system.

---

CHAPTER 6

---

## Synopsis

Our survey of this new, computational approach to vision is now complete. Although there are many gaps in the account, I hope that it is solid enough to establish a firm point of view about the subject and to prompt the reader to begin to judge its value. In this brief chapter, I shall take a very broad view of the whole approach, inquiring into its most important general features and how they relate to one another, and trying to say something about the style of research that this approach implies. It is convenient to divide the discussion into four main points.

The first point is one that we have met throughout the account—the notion of different levels of explanation. The central tenet of the approach is that to understand what vision is and how it works, an understanding at only one level is insufficient. It is not enough to be able to describe the responses of single cells, nor is it enough to be able to predict locally the results of psychophysical experiments. Nor is it enough even to be able to write computer programs that perform approximately in the desired way. One has to do all these things at once and also be very aware of the

additional level of explanation that I have called the level of computational theory. The recognition of the existence and importance of this level is one of the most important aspects of this approach. Having recognized this, one can formulate the three levels of explanation explicitly (computational theory, algorithm, and implementation), and it then becomes clear how these different levels are related to the different types of empirical observation and theoretical analysis that can be conducted. I have laid particular stress on the level of computational theory, not because I regard it as inherently more important than the other two levels—the real power of the approach lies in the integration of all three levels of attack—but because it is a level of explanation that has not previously been recognized and acted upon. It is therefore probably one of the most difficult ideas for newcomers to the field to grasp, and for this reason alone its importance should not be understated in any introductory book, such as this is intended to be.

The second main point is that by taking an information-processing point of view, we have been able to formulate a rather clear overall framework for the process of vision. This framework is based on the idea that the critical issues in vision revolve around the nature of the representations used—that is, the particular characteristics of the world that are made explicit during vision—and the nature of the processes that recover these characteristics, create and maintain the representations, and eventually read them. By analyzing the spatial aspects of the problem of vision, we arrived at an overall framework for visual information processing that hinges on three principal representations: (1) the primal sketch, which is concerned with making explicit properties of the two-dimensional image, ranging from the amount and disposition of the intensity changes there to primitive representations of the local image geometry, and including at the more sophisticated end a hierarchical description of any higher-order structure present in the underlying reflectance distributions; (2) the 2½-D sketch, which is a viewer-centered representation of the depth and orientation of the visible surfaces and includes contours of discontinuities in these quantities; and (3) the 3-D model representation, whose important features are that its coordinate system is object centered, that it includes volumetric primitives (which make explicit the organization of the space occupied by an object and not just its visible surfaces), and that primitives of various size are included, arranged in a modular, hierarchical organization.

The third main point concerns the study of processes for recovering the various aspects of the physical characteristics of a scene from images of it. The critical act in formulating computational theories for such processes is the discovery of valid constraints on the way the world behaves

that provide sufficient additional information to allow recovery of the desired characteristic. We saw many examples of this in Chapter 3, and they were summarized in Table 3-3. The power of this type of analysis resides in the fact that the discovery of valid, sufficiently universal constraints leads to conclusions about vision that have the same permanence as conclusions in other branches of science.

Furthermore, once a computational theory for a process has been formulated, algorithms for implementing it may be designed, and their performance compared with that of the human visual processor. This allows two kinds of results. First, if performance is essentially identical, we have good evidence that the constraints of the underlying computational theory are valid and may be implicit in the human processor; second, if a process matches human performance, it is probably sufficiently powerful to form part of a general purpose vision machine.

The final point concerns the methodology or style of this type of approach, and it involves two main observations. First, the duality between representations and processes, which is set out explicitly in Figure 6-1, often provides a useful aid to thinking how best to proceed when studying a particular problem. In the study both of representations and of processes, general problems are often suggested by everyday experience or by psychophysical or even neurophysiological findings of a quite general nature. Such general observations can often lead to the formulation of a particular process or representational theory, specific examples of which can be programmed or subjected to detailed psychophysical testing. Once we have sufficient confidence in the correctness of the process or representation at this level, we can inquire about its detailed implementation, which involves the ultimate and very difficult problems of neurophysiology and neuroanatomy.

The second observation is that there is no real recipe for this type of research—even though I have sometimes suggested that there is—any more than there is a straightforward procedure for discovering things in any other branch of science. Indeed, part of the fun is that we never really know where the next key is going to come from—a piece of daily experience, the report of a neurological deficit, a theorem about three-dimensional geometry, a psychophysical finding in hyperacuity, a neurophysiological observation, or the careful analysis of a representational problem. All these kinds of information have played important roles in establishing the framework that I have described, and they will presumably continue to contribute to its advancement in an interesting and unpredictable way. I hope only that these observations may persuade some of my readers to join in the adventures we have had and to help in the long but rewarding task of unraveling the mysteries of human visual perception.

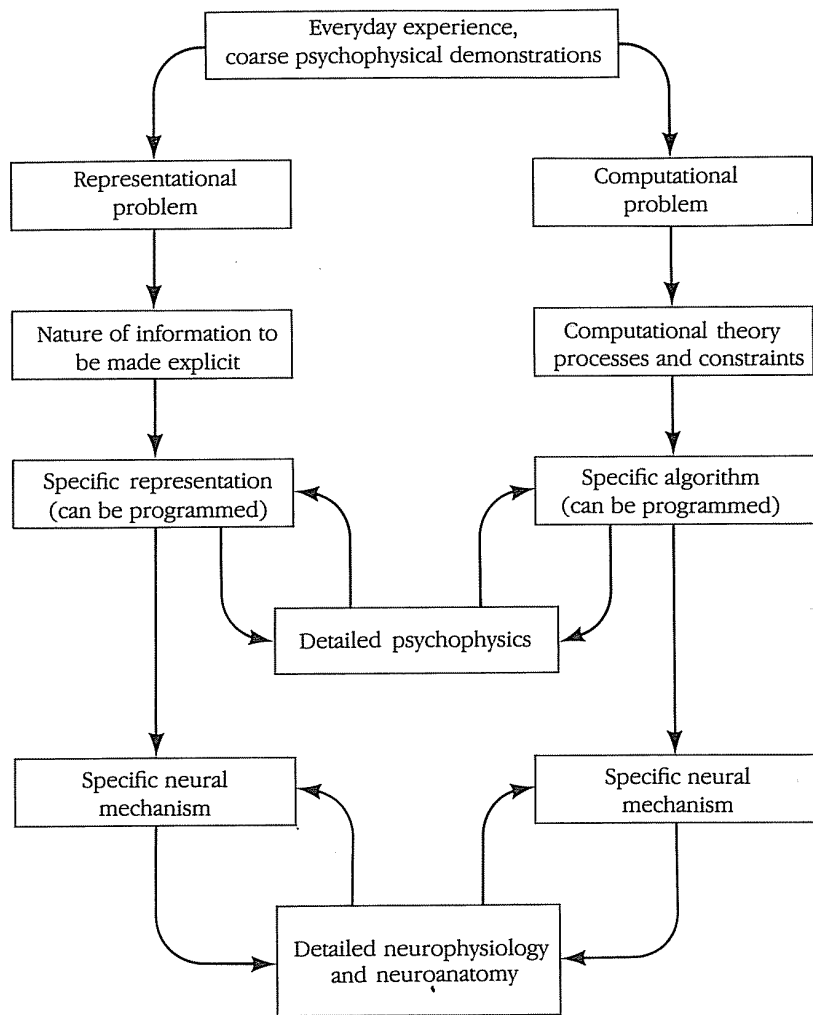


Figure 6-1. Relationships between representations and processes.

PART III

Epilogue