Some Issues for Discussion at the "Pre-Psychonomics Meeting on Object Representation & Recognition"

Thursday, November 4, 1993
3:30 pm Omni Shoreham Hotel
Washington DC

Figure-Ground Segregation

Bradley Gibson
Johns Hopkins University

1.) Must figure-ground organization necessarily be computed prior to object recognition? Why has this been assumed?

2.) What serves as the substrate for pre-figural object recognition processes, if not figural regions?

3.) Are there different routes to object recognition? Are pre-figural object recognition processes different than post-figural object recognition processes?

4.) Must shape always be given to the figural region?

5.) Can prefigural object recognition processes be incorporated into current models of object recognition?

6.) Can pre-figural object recognition processes be distinguished from alternative accounts of top-down effects on figure-ground organization (e.g., Rock's preconscious oscillations)?

Segmentation and Object Recognition

Johan Wagemans

University of Leuven, Belgium
(presently at the University of Virginia, Charlottesville, VA)

Before any object recognition system could start doing its work, it must receive input from an earlier system which has localized the region of interest by segmenting it from a cluttered visual scene (e.g. a blob-like region).

Question 1: Is it justified to assume that this scene segmentation problem is solved before object recognition begins? (Recent evidence by Peterson & Bradley has shown that some processes of object recognition might start before even figure/ground segregation has been achieved)

Question 2: Many different cues are known to play a role in this scene segmentation process (motion, texture, depth, luminance contrast etc.). A common assumption is that all of these cues are processed in different modules working in parallel. Do we have to assume that all of them will produce the same output (i.e. the same list of "blobs" to recognize) or do we need a scheme for evaluating and/or combining these outputs?

Assuming that we have the blobs of interest, many object recognition systems would just start working on the contours. Again, we know that there are other routes to derive aspects of shape from these 2-D inputs (stereo, structure-from-motion, shading, etc.). Therefore, similar questions arise in this context.

Question 3: Is it justified to assume that these modules recovering "shape-from-something" are doing their work before any contact with memories of specific objects is made?

Question 4: Do we have to assume that all of them will produce the same output (e.g. depth map or 2.5-D sketch) or do we need a scheme for evaluating and/or combining these outputs? Most research that I know of (e.g. by Bülthoff) shows that many different schemes are used depending on the source of information and the task at hand.

Assuming that we just focus on shape-from-contours for the time being, I think here the issue arises whether we should start deriving 3-D information or just continue working in 2-D. As far as I can tell this is one of the most critical factors distinguishing different theories of object recognition (e.g. Lowe vs. Marr). In relation to this very general question, I would like to point out two specific issues where I think we should be more careful than we are.

Question 5: Some research programs tend to overlook the difference between the theoretical claims and the specific research materials used. In Biederman's seminal Psych Rev paper (1987), for example, most of the claims are about the relevance of a set of 3-D building blocks (called "geons") but most of the experiments mentioned have used simple outline drawings in which specific regions would often be compatible with several different geons. How relevant are experimental results obtained with outline drawings for theories about 3-D object recognition?

Question 6: Another difference that is often overlooked is between outline drawings and silhouettes. Obviously, both are similar in that internal information which might also be relevant to derive 3-D shape (like internal contours or shading gradients) are absent, but both differ dramatically in their behavior under different filtering schemes. Some experiments are done with silhouettes, but most of our research on object recognition has used outline drawings. Should we expect essentially the same results, or do we know they are different?

Assuming that we work with outline drawings and want to focus on an object's parts to recognize it, there are essentially three different approaches (see Ullman, 1989 or Vaina & Zlateva, 1990 for an overview): primitive-based (Biederman's RBC), axis-based (Blum's and Marr's generalized cylinders), and boundary-based (Hoffman & Richards' codons). A lot could be said about the comparison between these in terms of efficiency, feasibility, plausibility, etc., but the most general question is:

Question 7: Do these schemes produce essentially the same parts for most objects, or can they be distinguished empirically? In the latter case, is one of them more general or are they complementary?

Since Attneave's demonstration that curvature extrema are important for object recognition, surprisingly little empirical research has been devoted to comparing their role with the role of inflections, which are computationally much more interesting (because they are much more robust to noise and invariant under a larger set of transformations). Moreover, the few experiments in relation to this issue are flawed (e.g. Kennedy & Domander).

Question 8: Is it just too difficult to design fair experiments on this issue? For example, it might be that it critically depends on the shape itself. Some of the inflections are just very hard to localize (perhaps there is an interesting trade-off here).

Assume that we encode a shape's contour with something like Hoffman & Richards' coding scheme (minima for segmentation, maxima and inflections to describe the primitives). Although a codon description is very appealing in some respects, it has to deal with two rather detailed and technical issues which are nevertheless important.

Question 9: It is clear that a codon description provides only a very coarse summary of a shape's contour. In this sense, it might fail on the sensitivity criterion for object representation formats: Can two shapes be distinguished when they share the same codon representation? (yes, they can) On the other hand, from psychophysical results on curvature discrimination we know the visual system works with a rather restricted set of curvature primitives (e.g. Foster, 1989: straight, slightly curved, large amount of bent). Are these data relevant?

Question 10: To describe a contour's curvature as a sequence of singularities, one has to set a threshold, because curvature varies at a large number of different scales. Should we take just one scale, or a hierarchy of scales, or something computationally very demanding, like a scale-space?

Object-Centered and Viewer-Centered Theories of Representation, and Other Tales

Cassandra Moore
Columbia University

William Hayward
Yale University

1. Does the object recognition system utilize both viewer-and object-centered representations? With which stimuli or in which perceptual situations might each be useful? Are these characterizations opposite ends of a spectrum or two dissociable formats? Can viewer-centered representations become object-centered representations? If so, is representation necessarily a two-stage process in which objects are first represented as viewer-centered structures, then as object-centered structures?

2. Are there separate mechanisms for within- and between-category discriminations between objects? For example, Biederman and Tversky & Hemenway suggest that basic level categories differ in the particular parts they contain and the structural relations between those parts. Diamond & Carey propose that within category distinctions (at the subordinate level) are made on the basis of metric relations while the same parts and structural relations are shared by category members. Does the adoption of these views require us to maintain that all members of a basic level or subordinate level category necessarily have the same parts? Or that the basic level is defined perceptually in terms of the parts and their relations? Do we use different features of objects and/or different cognitive processes when we make within as opposed to between category discriminations?

3. Does Farah's multiple parts/complex wholes distinction map onto the object-centered/viewer-centered representation distinction? On the basis of this mapping, how is Farah's model related to within-category/between-category recognition?

4. In the recent object perception literature, two "structural description systems" have been proposed: Biederman's local relations (atop, beside, below) and Cooper & Schacter's global object axis representation. In what specific ways are the two structural description systems different from each other? What are the pros and cons of each type of structural description? Can one be subsumed by the other?

5. How does traditional priming specificity account for the non-specific aspects of visual object priming reported by Cooper & Schacter? Is priming caused by a single process (e.g. Tulving & Schacter's "perceptual representation system"), or multiple perceptual systems that might prime different aspects of the physical stimulus?

6. Research on face recognition has shown that (i) in some situations caricatures of faces are recognized faster than the line drawings of faces and (ii) deficits in brain damaged patients can be extremely specific to faces (e.g. a recent case (McNeill & Warrington, 1993, QJEP) where human faces, but not sheep faces or apparently any other stimuli, were unable to be identified). What does this research have to offer work in object representation? Are faces examples of a single basic level category? Or are they in some way special and hence only of limited interest to this group? In other words, how can the face recognition literature inform the study of object representation and object recognition?