THE STRUCTURE OF VISUAL CONCEPTS

Joseph Psotka
Computer-Based Systems Development
Army Research Institute
Alexandria, Virginia.

The research reported here addresses the question: What is the visual equivalent of a word, idea, or higher-order concept? This research is a beginning attempt to analyze visual concepts in ways that are most appropriate for future richer computational environments that will be able to process images quickly and in great numbers: massively parallel machines like those envisioned by Feldman and Ballard (1982). Knowledge representation in artificial intelligence applications and other computer-based systems today uses many structuring schemes: feature lists, vectors, trees, networks, and production systems. In each, entities or nodes are filled with concepts representing words or lists of words. Most of these systems have developed out of a linguistic background, and reflect the symbolic processing limitations of present-day computers. This context has made images and pictures difficult to use, but it is changing.

The foundation for this research lies in the many experiments that demonstrate the great fidelity of the human pattern recognition system and its enormous capacity to store and process large numbers of visual representations in very short periods. For instance, recognizing the face of an acquaintance one has not seen for many years implies the existence of an ability to select one alternative from thousands (perhaps millions, given the many transformations a face may undergo) within a few moments. Computer-based systems cannot begin to rival this real-time feat, in part perhaps because no knowledge representation schemes appropriate to specifically visual concepts have yet been developed.

Several experiments have been conducted (and one will be demonstrated to the audience) that provide some support for the theoretical proposal that Galton's system of composite portraiture provides one model for the structure of visual concepts. This proposal also has practical implications and applications to computing. Galton used a photographic process to add several photographs of human faces onto the same picture, one on top of the other. The resulting single face may be a prototype for this particular set of faces (see Osherson and Smith, 1981 for a discussion of prototypes). His technique was used in these experiments and has been extended by Weil (1982) to a computer-controlled optical videodisc system funded by DARPA. Large scale parallel optical computers offer the possibility of dramatic future enhancements.

The experiments in this report are an impoverished beginning for the very large task ahead. They have established that:
1) Composite faces can be recognized after viewing the set of component faces they comprise;
2) The similarity of components to composites correlates with recognition performance;
and 3) Composites are more attractive than their component faces.
These findings are hardly earth-shaking in their own right, but they fit well with the general framework that composites are prototypes. Basically they provide an incentive for looking for more critical facts that might flesh out the critical features of visual concepts.

Galton's technique of superimposing whole faces in registration to develop an average face provides a wholistic procedure for creating prototypes and making them visible. The first question that the procedure raises is: can people recognize these wholistic composite faces after viewing their components? Second, do these prototypes accurately represent critical features and relations among the components they comprise? Third, is there anything unique or distinct about these prototypes, not found in their components?

Finally, the Galtonian process of composite portraiture provides a reduced analogue for convolution and crosscorrelation processes (as used in holography) that have provided the impetus for models of memory (cf. Psotka, 1977; Metcalfe and Murdock, 1981). Are these composites sensitive to common memory research manipulations?

SUBJECTS: Four groups of 25 undergraduates were used in the first half of the experiment. A fifth group of 20 undergraduates were use in the second half. All students were naive about the purpose of the experiment and participated as part of a course requirement at the University of Waterloo.

PROCEDURE: 100 full-face photographs were selected randomly from male yearbooks with the restriction that there should be no facial hair or eyeglasses. These faces were randomly grouped into ten lists of ten faces whose composites were photographed.

The first four groups were shown only one-half of the 100 faces in five lists of ten faces, each list followed by a forced choice pair of composites (See figure 1). They were asked to pick the more familiar of the two composites. Only one of the composites was composed of the list of ten faces previously shown. In order to ensure that each student examined each face carefully, they were asked to rate the attractiveness of each face on a ten point scale. The faces were presented on slides shown in groups with order of presentation of the lists counterbalanced.
The fifth group rated the similarity of the paired composites to each of the two sets of ten components on a ten point scale. This task was performed individually using prints of the slides.

RESULTS: Recognition performance showed that people were able to recognize a composite on the basis of viewing the prototypes. Percent correct ranged from 48 (chance) to 92. Eighteen out of 20 groups were above chance (sign test, $p < .01$). There was an effect of order of presentation. The first prototypes shown were recognized best.

The similarity ratings showed that a component face was more similar to the composite into which it entered than to the distractor composite ($t=5.23$; $df=99$; $p<.01$). A significant correlation was found between the similarity ratings and percent correct recognition ($r=.86$; $df=8$; $p<.01$).

The composites were each judged much more attractive than the component faces. All ten lists showed this effect. Only six of the 100 faces were judged more attractive than the least attractive composites. These faces were also the most similar to the composites.

CONCLUSIONS: The evidence is clear that people can recognize composites on the basis of their memory of previously viewed components. Of course, they could use distinctive features from each unique face stored in memory, or they could be creating a composite as they are viewing the set of faces. On this point, the evidence is not decisive; but then this is a venerable dispute that traces back to Locke and Berkeley. I would hardly expect one experiment to provide a resolution. Nevertheless, I believe that something like these composites is being formed in the visual/memorial system.

The high correlation between the difference in similarity ratings and percent correct recognition provides evidence that people base their recognition process on the entire group of faces they have seen. If the group of faces is remembered as a whole, then in some sense it can be represented best by a composite that is an average of the entire group. It seems impossible to distinguish between a set of separate and unique individuals and an amalgam --- schema or prototype --- that acts separately unless an emergent property of the amalgam can be found. This emergent property may be attractiveness. There seems to be no satisfying explanation for the increased attractiveness of the composites except that they are more representative of the whole set of faces (beyond this experiment) encountered in our experience. They are more like the prototypical face we have unconsciously created as a standard in its physical rather than personality characteristics.

In summary, there are four main pieces of evidence to leave with:
First, there exist longstanding conceptions and theories with elegance and considerable power. They address the existence of global concepts as unique ideas or associations, separate from individual stimulus items, events, or episodes in memory.
Second, the techniques reported here have uncovered several facts that may not provide critical support for these theories, but lend credence to them and offer avenues for research to discover facts that would distinguish them from competing theories.

Third, the findings are of some interest in their own right. The relations among motivation and cognition are still wide open to be explored by cognitive scientists for they have hardly advanced since Wolfgang Kohler's monumental effort in "The Place of Value in a World of Fact".

Finally, these techniques applied with greater computational power may be useful tools for other purposes: to create beautiful images, or for criminal identification as Weil (1982) began to explore.

REFERENCES

Feldman, J.A. and Ballard, D. H. Connectionist models and their properties. Cognitive Science, 1982, 6, 205-254.

Metcalfe, J. and Murdock, B. B. An encoding and retrieval model of single-trial free recall. JVLVB, 1981, 20, 161-189.

Osherson, D. N. and Smith, E. E. On the adequacy of prototype theory as a theory of concepts. Cognition, 1981, 9, 35-58.

Psotka, J. Syntely: Paradigm for an inductive psychology of memory, perception, and thinking. Memory & Cognition, 1977, 5, 553-560.

Weil, P. About Face: Computergraphic synthesis and manipulation of facial imagery. MIT, 1982, MS dissertation.