SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

REPORT DOCUMENTATION PAGE

DTIC FILE COPY

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SEC	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DEC	Approved for public release; distribution unlimited.
4. PERF ... REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
AD-A204 490	AFOSR-TR- 89-0002

6a. NAME OF PERFORMING ORGANIZATION: The Research Foundation of State University of New York	6b. OFFICE SYMBOL (If applicable):
6c. ADDRESS (City, State and ZIP Code): Attn: John Boehler, Sponsored Programs, 516 Capen Hall, Amherst, NY 14260	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
7b. ADDRESS (City, State and ZIP Code): AFOSR/NL, Building 410, Bolling AFB DC 20332-6448	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: University of Minnesota	8b. OFFICE SYMBOL (If applicable): NL
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-NL-86-0106	
8c. ADDRESS (City, State and ZIP Code): Attn: Merlin Garlid, Office of Research and Technology Transfer, 1919 University Avenue, St. Paul, MN 55104-3481	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313	A5	

11. TITLE (Include Security Classification): Human Image Understanding

12. PERSONAL AUTHOR(S): Irving Biederman

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 86 Feb 1 TO 88 May 31	January 1989	23 + reprints

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Pattern Recognition, Perception, Vision, Image Understanding.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This is the final progress report for AFOSR grant 86-NL-0806 HUMAN IMAGE UNDERSTANDING. The goal of the effort is to develop and empirically evaluate a theory (Recognition-by-Components (RBC)) of real-time human target identification which assumes that objects are represented as an arrangement of simple generalized-cone volumes. The fundamental assumption of RBC is that a particular set of these convex components, called geons, can be derived from invariant properties of edges in a 2-D image. If an arrangement of three geons can be recovered from the input, objects can be quickly recognized even when they are occluded, rotated in depth, novel, extensively degraded, or embedded in a scene. The report describes the research on consequences of various forms of image degradation, the exploration of the role of surface features, the attentional demands of object recognition, formal modeling of object recognition, and extensions to scene perception and expert identification.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐	21 ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. John F. Tangney	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5021
	22c. OFFICE SYMBOL: AFOSR/NL

DTIC ELECTE 3 FEB 1989 S D

12 JAN 1989

This final report will document theoretical, empirical and methodological developments on AFOSR grant No. 86-NL-0086, HUMAN IMAGE UNDERSTANDING. The general background for this effort can be obtained from the reprint of the *Psychological Review* article, "Recognition by Components: A Theory of Human Image Understanding" and the reprint of the chapter "Aspects and Extensions of a Theory of Human Image Understanding" in Z. Pylyshyn (Ed).

CONTENTS

I. INTRODUCTION AND BACKGROUND

Humans can typically recognize an object even when it is viewed from a novel orientation, or it is a novel exemplar, or its image is extensively degraded. Moreover, most often

only a single, brief fixation is all that is required to achieve quick and automatic understanding. The fundamental problem addressed by Recognition-by-Components (RBC) theory is how this is accomplished. Because a line drawing of an object can be classified as rapidly and as accurately as a full colored, textured photograph of the object (Biederman & Ju, 1987) the problem can be stated as one of determining how the edges extracted from an image of an object can activate--in real time--an appropriate representation of that object in memory.

RBC assumes that an image of an object is segmented at regions of deep concavity into an arrangement of simple convex generalized cone primitives, such as cylinders, bricks, wedges, and cones (Biederman, 1987a) as illustrated in Fig. 1a. The central assumption of the theory is that the members of a particular set ($N \leq 24$) of primitives, called geons (for geometrical ions), are distinguishable on the basis of dichotomous or trichotomous contrastive viewpoint-invariant properties of image edges, such as curved vs straight, parallel vs nonparallel, and cotermination of edges (for defining vertices) (figure 1b). These image properties can be determined from a general viewpoint and are highly resistant to degradation. Consequently, the geons, which are derived from these edge contrasts, themselves will be determinable under degradation and variations in viewpoint (Figure 1c). An analysis of the representational capacity of the geons and their relations leads to the expectation that the basic level classification of most single visual entities can be achieved from an arrangement of only two or three geons (Biederman, 1987a).

Stages of Processing

Figure 1d presents a schematic of the subprocesses posited by RBC. The stages are assumed to be arranged in cascade whereby partial activation (processing) at one level is sufficient to initiate activation at the next. An early edge extraction stage, responsive to differences in surface characteristics, viz., sharp changes in luminance or texture, provides an edge-based description of the object.

Following the determination of the components, a structural description specifying the components and their relations is then matched against a like representation in memory. It is assumed that the matching of the components occurs in parallel, with no loss in capacity when matching objects with a large number of components. Partial matches are possible with the degree of match assumed to be proportional to the overlap in the componential descriptions of a representation of the image and the memorial representation.

II. A Connectionist Implementation of RBC

Hummel, J. E., Biederman, I., Gerhardstein, P. C., and Hilton, H. J. From image edges to geons: A connectionist approach.

Hummel, Biederman, Gerhardstein & Hilton (1988) have implemented a connectionist model of geon recognition. The model is a five-layer network (Figure 2) that takes as input an activation vector representing the configuration of edges in the image of a geon. As output (Layer 5), the model produces an activation vector representing the geon defined by that configuration of edges. The connections that perform the mapping from image edges in the first layer to geons in the fifth were derived through error back propagation. (The Hummel et al. paper presents an earlier four layer version. The within-column architecture and representation and organization among the columns are virtually identical to the version presented here.)

The major goals of this effort are to determine: (1) whether the constraints imposed by the edge-to-geon mapping were sufficient to force the model to discover the non-accidental (or viewpoint-invariant) image properties posited by RBC as the basis upon which geons are

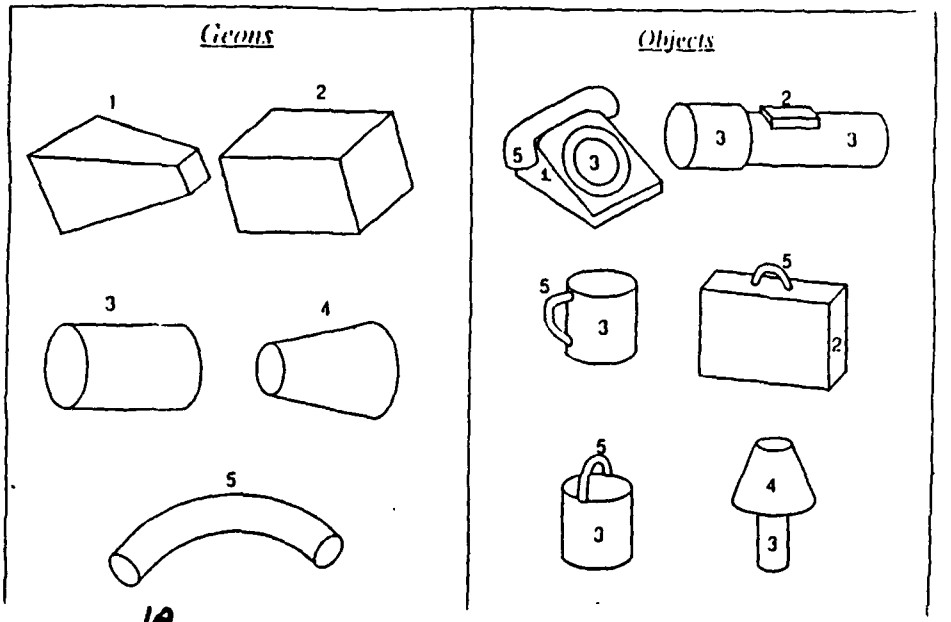


Figure 10. Left panel. Five geons. Right panel. Only two or three geons are required to uniquely specify an object. The relations among the geons matter, as illustrated with the pail and cup.

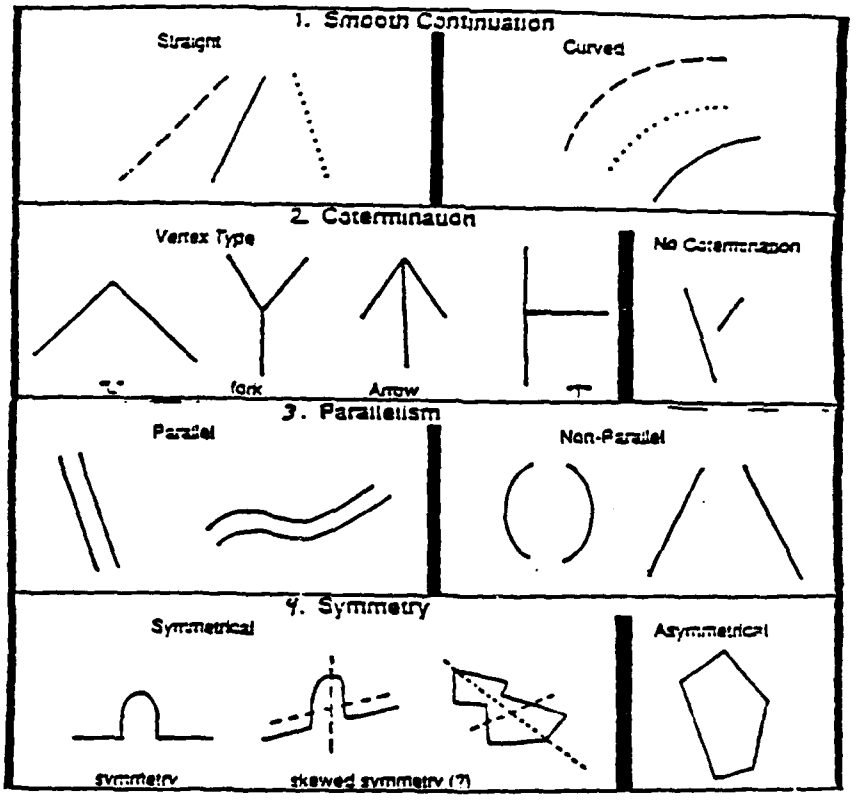


Figure 10. Contrasts in four viewpoint-invariant relations. In the case of Parallelism and Symmetry, biases toward parallel and symmetrical percepts when images are not exactly parallel or symmetrical are evidenced. [Adapted from Figure 5.2, *Perceptual organization and visual recognition*, [p.77] by David Lowe. Unpublished doctoral dissertation, Stanford University, 1984]. Reprinted by permission of the author.

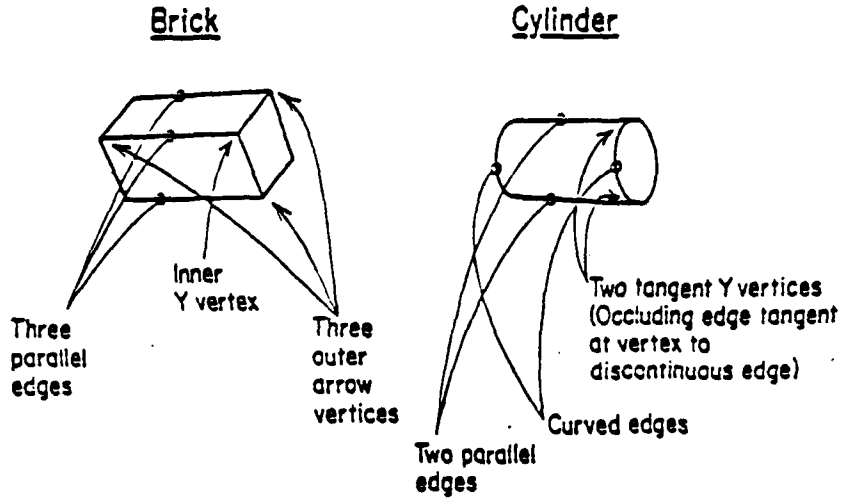


Figure 1C. Some nonaccidental differences between a brick and a cylinder. From Fig. 5, Recognition-by-Components: A theory of human image understanding, by Irving Biederman, 1987, *Psychological Review*, *94*, p. 121. Copyright 1987 by the American Psychological Association. Reprinted by permission of the publisher and author.

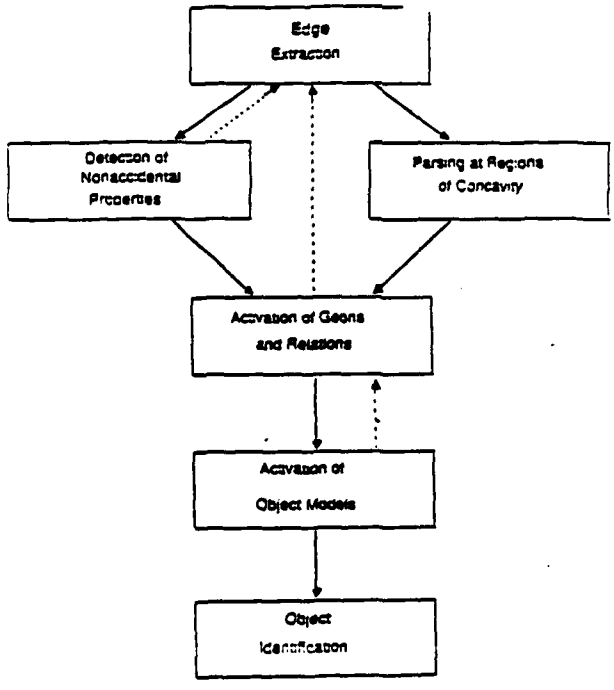


Figure 1D. RBC's processing stages for object recognition. Possible top-down routes are shown with dashed lines. From Fig. 2, Recognition-by- Components: A theory of human image understanding, by Irving Biederman, 1987, *Psychological Review*, *94*, p. 118. Copyright 1987 by the American Psychological Association. Reprinted by permission of the publisher and author.

recognized; (2) whether, using these non-accidental properties, the model could achieve translation, rotation and size invariance in geon recognition; (3) if the present model, replicated at different scales and locations over the visual field, could achieve parsing of a multi-geon image into its constituent parts, and (4) if the resulting representation could be used to derive inter-geon relations from the image of an object and, together with the descriptions of the geons themselves, be used to drive object recognition. The aspect of the model to be described here directly addresses the first two of these goals.

Architecture and Representation

The Input Layer. The model's input layer is divided into 19 identical clusters of cells. Each cluster contains 20 cells for the detection of image edges, and is located over a particular portion of the model's visual field. The cell clusters form an hexagonal lattice such that the center of a given cluster, *i*, is *r* units from the centers of i's six nearest neighbors (see Figure 2). Each cluster has a circular receptive field of radius r; image edges within a cluster's receptive field are recorded as activity in the cells of that cluster. As the receptive fields of adjacent clusters overlap, any given image edge will be registered in the cells of at least two input clusters.

The 20 cells within an input cluster respond to image edges in terms of straight and curved segments and terminations. There are four cells that respond to different orientations of straight segments, four that respond to different orientations of curved segments, and twelve that respond to the points at which edges of different orientations terminate (Figure 3). Thus, the cells within the model's input layer respond to edge features on the basis of: (1) location (defined by the location of the cluster to which a particular cell belongs), (2) whether the feature is an edge segment or termination, (3) curvature (in the case of segments only), and (4) orientation.

Whether a cell will respond to a given edge feature is defined by the compatibility between the cell type and the feature type. For example, edges tuned to respond to straight segments will not respond at all to the presence of a curved edge. However, given that a cell and feature are of compatible types, the strength with which the cell will respond to the feature is a non-linear function of the feature's location and orientation.

The Output and Intermediate Layers. As stated above, the model's output is an activation vector indicating the identity of the geon defined by the configuration of edges it is given as input. This output is produced in the eight cells of the model's fifth layer. Each cell at this level locally codes one of eight geon types: brick, wedge, cylinder, curved cylinder, cone, truncated cone, prism, and curved cone. The specifics of representation in the model's intermediate (second, third and fourth) layers were not designed a priori; a primary aim of this modelling effort is to observe what representations emerge naturally in these layers as a function of the model's mapping task. However, the architecture of the intermediate layers (including their inter-connectivity and connectivity to the input and output layers) is highly constrained on the basis of a priori considerations.

As in the first layer, the cells in the intermediate layers are organized into identical clusters. A given cluster in layer L has connections to only a subset of the clusters in layer L-1. This constrained pattern of connectivity between layers accomplishes two specific computational aims. First, it determines the degree to which retinotopic mapping is preserved as activation is passed between the layers; the size of a cluster's receptive field determines exactly how much of the visual field is represented by the activity in that cluster. Second, constraining the connections between layers to local subsets of the cells in those layers allows the connections to

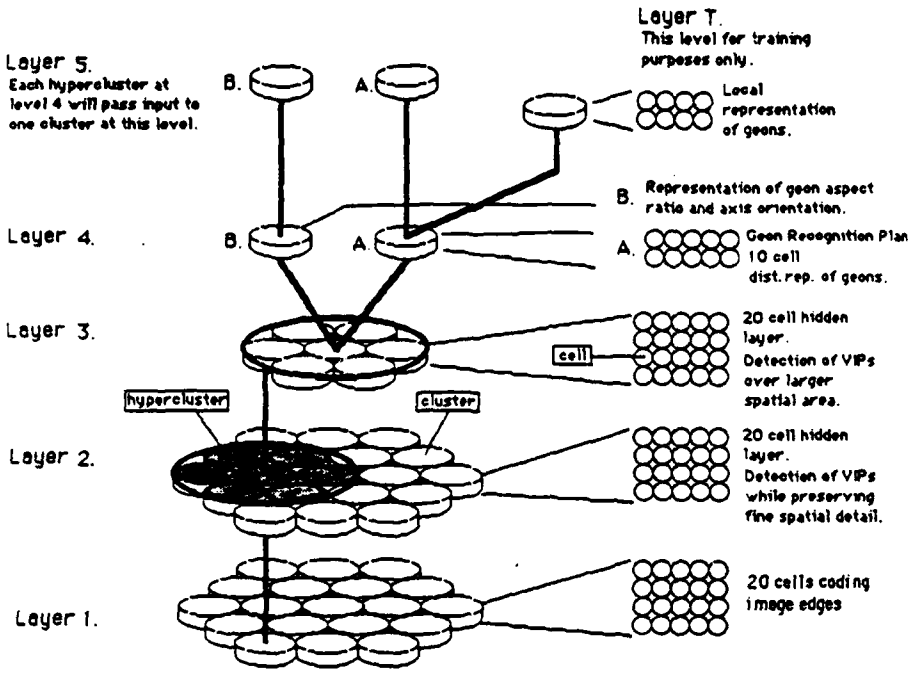


FIG. 2. The Five Layer Column for Geon Recognition

Layer 1 : 19 Clusters each with 20 cells - edges are represented here
Layer 2: 19 clusters each with 20 cells - each cluster takes its input from one layer 1 cluster
Layer 3 : 7 clusters each with 20 cells - each cluster takes its input from one layer 2 hypercluster
Layer 4 : 1 cluster with 10 cells - this cluster takes its input from the single layer 3 hypercluster
Layer T : 1 cluster with 8 cells - this cluster takes its input from the single layer 4 cluster - geons are locally represented here - this layer is only for training

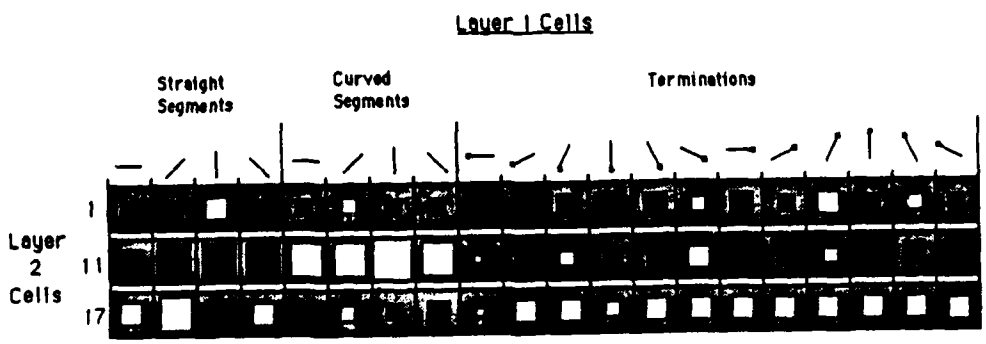


Figure 3. A Subset of the Connections from Layer 1 to Layer 2.

Above are shown the connections from all 20 first layer cells (within one cluster) to three cells in a second layer cluster. All second layer clusters are identical. Therefore, this pattern is replicated throughout all 19 second layer clusters. The black boxes indicate inhibitory connections, and the white excitatory connections; the size of a box is proportional to the magnitude of the connection.

be "reused". That is, if several separate connection matrices perform a given layer-to-layer mapping, these connections matrices can be duplicated. The advantage of this type of "matrix duplication" is that it insures translational invariance in mapping between separate regions of adjacent layers.

The mapping between the first (input) and second layers occurs on a one-cluster-to-one-cluster basis. Each of the 19 clusters in layer two contains 20 cells that are fully interconnected to the 20 cells in one layer-one cluster. Thus, retinotopic mapping is completely preserved in mapping from the first to second layers. The motivation behind this design was to allow the training regime discover highly localized VIP features, viz., vertices, in the second layer clusters.

The model's third layer contains seven clusters, each with 20 cells. Each third layer cluster takes its input from an hexagonal lattice of seven clusters in the second layer (see Figure 2). This seven-to-one mapping was designed to allow the clusters in the third layer to discover important combinations of firing patterns in the clusters of the second layer that may represent viewpoint invariant properties occurring over an extended region, such as parallelism or symmetry.

The mapping between the model's third and fourth layers is also a seven-to-one mapping. The single fourth layer cluster contains only ten cells and serves as a bottleneck in which the model was forced to discover a distributed representation of the geons represented locally in the output (fifth) layer. By virtue of the seven-to-one mappings leading up to the fourth layer, the cluster there summarizes over all spatial information within the original input image. Therefore, within the scope of the present model, retinotopic mapping is not preserved in the fourth layer. However, the complete model is assumed to consist of several duplicates of the present model distributed, at various spatial scales, over the visual field. As such, retinotopic mapping is preserved between fourth-layer clusters in the complete version.

Simulation Procedure. The following discussion of simulation results is based upon a simulation with a six-geon training set. The set included 48 stimuli in all: eight examples each of bricks, wedges, cylinders, curved cylinders, cones and truncated cones. The set was generated by creating two tokens of each geon type and presenting each token in four randomly chosen orientation/position conjunctions. The model was trained by back propagation and training proceeded until criterion performance (100% correct recognition and mean error per output cell less than .02 deviation from desired output) was achieved.

Most of the parameters employed during training and testing are unimportant to the present discussion. However, the effect of the activation rule was sufficiently striking as to be worthy of note. The activation rule employed was a (-1, +1) bounded version of the logistic activation function typically employed in back propagation models. An output threshold of zero was imposed on the cells, and no input bias was used. As a result, cells whose net input was negative or zero produced an output of zero, and cells whose net input was positive produced an output that approached 1.0 as the input approached infinity. The effect of this rule was to completely shut off cells whose net input was negative. This effect is to be contrasted with the typical rule in which a cell's output approaches zero as its input approaches negative infinity. By quieting cells whose inhibitory inputs exceeded their excitatory inputs, this activation rule greatly reduced the amount of noise propagated through the system. The effects of this noise reduction are discussed briefly below.

Results

Training proceeded very rapidly: the model required only 28 presentations of the training set to reach criterion performance. This rapid convergence can be attributed both to the limited training set employed and to the parameters used to govern the learning algorithm.

After training, the model was tested on four classes of stimuli: (1) novel translation/rotation conjunctions of trained tokens, (2) novel tokens of the trained geon types (e.g. another configuration of a cylinder), (3) novel geon types, and (4) scrambled tokens of trained types. The model's ability to recognize novel stimuli is best described as imperfect but sensible. It generalized perfectly to novel instances of familiar tokens (test set type (1)). That is, when test stimuli were constructed by modifying the positions and orientations of training stimuli, performance was perfect. The model thus appears to have developed translation/rotation invariance for the stimuli in its training set.

The model's ability to generalize to novel tokens of each of the trained types (test set type (2)) was somewhat less reliable, with a few stimuli classified as similar geons but most geons were correctly classified. Eight examples each of prisms and curved cones--geons that did not have representatives in the model's training set--were used to test the model's classification of novel geon types. Its classification of these stimuli was sensible, revealing the appropriate similarity structure, e.g., prisms were consistently classified as brick/wedge/cone combinations. Further, the cells activated tended to cluster closely around the characteristics of the stimulus presented.

The test set type 4 stimuli were created by rearranging (scrambling) the vertices and edges composing the bricks and cylinders in the training set. The vertices themselves, and any parallelism among the constituent edges, were preserved; only the spatial arrangement of these features was perturbed. Stimuli of this type were consistently classified correctly despite the stimulus features' incorrect relative positions: scrambled bricks were classified as bricks, and scrambled cylinders as cylinders or curved cylinders. We regard this "success" as problematic--a consequence of the model's insensitivity to spatial relations.

The insensitivity to relations characterizes many connectionist modeling efforts and it should not have been a surprise that our initial effort would reflect this shortcoming. The specific insensitivity to scrambling the features of a geon is a likely a consequence of the requirement that the model derive a viewpoint-invariant representation of the geons directly from their constituent features. Deriving viewpoint invariance directly from image features (such as vertices or parallelism) requires that many feature-sensitive cells have excitatory connections directly to the cell or cells that represent a particular geon. For example, if the cell representing a brick is to be activated by the features of the brick regardless of viewpoint, that "brick" cell must be connected to all cells representing the features of bricks from all viewpoints. As a consequence, any combination of "brick features" will result in the activation of the brick cell (provided only that a sufficient number of the feature cells are active). Thus, it seems that when a viewpoint-invariant representation of a geon must be derived directly from the features defining that geon, the relations among those features get ignored by necessity.

One possible solution to this problem is to incorporate a junction dictionary. We are currently exploring how such a dictionary could be incorporated into a connectionist account.

Overall, given the early state of this modeling effort, the results of the simulation are encouraging. The model succeeded in learning translation and rotation invariance for the tokens in its training set, and it generalized reasonably well both to new tokens of familiar types and to novel types. Also, analysis of the weight matrices revealed the emergence of many of the VIPs posited by RBC.

This continues to be an active part of our research efforts. On a computational and descriptive level, RBC has also undergone some modifications and extensions. Many of these are described in Biederman (1988a).

III. Assessing Representation Through Priming

We have launched several experiments designed to assess the nature of the representation that results from the viewing of a picture of an object. It is well known that the prior viewing of a picture results in a facilitation in the speed of identifying that same picture on a subsequent exposure (e.g., Bartram, 1974; Biederman, Blickle, et al., 1988). The presentation of the first picture is then said to have "primed" the identification of the second picture.

Priming with Complementary Images (w. Eric Cooper). According to RBC, the representation of an object is in terms of its geons, which are activated by image features, such as vertices and edges. But if the geons are activated by image features--vertices and edges--why not just represent an object in terms of image features? To see why this may not be so, the reader is invited to identify the recoverable contour-deleted images shown in Figure 4. Now look at Figure 5. When viewed without the benefit of side-by-side comparisons, observers generally report that the images are the same. But figures 4 and 5 are actually complementary images, with each member of an object pair (e.g., the flashlight) having 50 percent of the contour of the original intact version. The images were produced by deleting, from each geon, every other vertex and edge. Each image of an object is thus a complement of the other so that if the two versions were superimposed they would make an intact picture with no overlapping of contour. (A small segment of edge was retained to define the vertex. Also, very long edges were divided between the versions.) If we were to represent objects in terms of image features, we would need a different representation for each arrangement of occluding contour.

An image feature representation would suggest, therefore, that the recognition of the original should show an advantage over the recognition of the complement. Because the same geons would be activated from either representation, according to RBC there should be no difference in the two versions.

We tested these possibilities in an experiment with 24 object pictures (and 32 subjects). In an initial priming block, subjects viewed and named one of the two deletion versions of each of the 24 objects for 500 msec. In the second (testing) block, the images were either the identical or complementary versions of the 24 object pictures that had been shown in the priming block. In both blocks subjects had to name the images as quickly as possible. The results confirmed the RBC account: recognition performance for complementary images (845 msec RTs and 8% errors) was virtually identical to performance on the identical images (832 msec RTs and 11.1% errors).

There is no doubt that people could code the individual image features in that they could learn to distinguish the various versions of the complementary images. But the expectation from RBC would be that the reliance on such coding would slow their identification performance. That is, subjects might more readily identify the complementary version of the camera if they did not attempt to determine if it contained the particular vertices and segments present in the original version.

We are currently running a control experiment in which new object classes and different geon models of the same class appear in the second block of objects. Because these images would have different geon models they should not be primed by the first block of pictures.

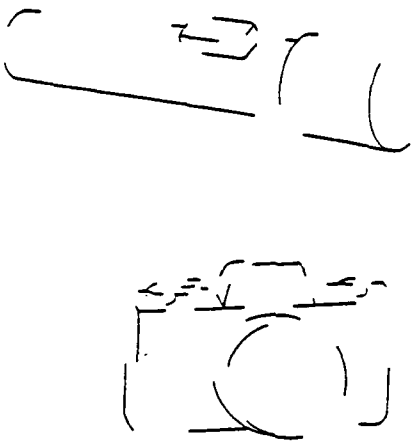


Figure 4 Contour deleted images of two objects.

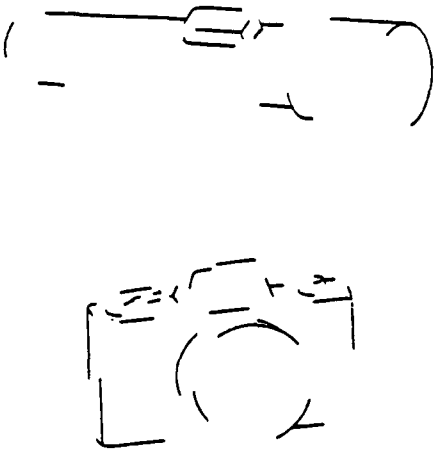


Figure 5 The complements to the contour deleted images of the two objects in Figure 4. These images contain almost all of the missing edges and vertices of the objects in 4 with almost no overlap between the two figures.

Bartram has already demonstrated that such pictures are indeed worse than repeated images but this needs to be demonstrated in our own paradigm. Worse performance on these images would document that there was, indeed, a priming effect.

The effects of mirror reversal. In this experiment half the images on the second block of trials were in the original left-right orientation and the other half were in a reversed orientation. No special status is accorded to such orientation in RBC so there is no reason why any effect would be expected. It is not that people could not recognize the orientation of the object. It may be merely that speeded recognition might not depend on specification of orientation. And this was found. Objects in their original orientation (829 msec RTs and .08 percent errors) were just barely (but not significantly) better than those images in their reversed orientation (847 msec RTs and 11% errors).

The effects of rotation in depth. A variant of the complementary priming task will provide a strong test as to whether the representation is indeed 3D (rather than 2D). In this task, the test (second block) images will be presented as rotated in depth, but with the same geons present. (Often rotations of 10-20° for many objects provide the same geon descriptions as an original orientation.) A lack of an effect of such rotation and the complementary deletion would provide strong evidence that the representation was not of the image features (segments and vertices), but of volumetric units. In these experiments, conditions will be run with "nonsense" objects that did not conform to familiar object classes to insure that the priming was not from a familiar object model.

IV. Role of Surface Features in Object Recognition.

A. **Surface vs. Edge-Based Determinants of Visual Recognition.** Biederman, I., & Ju, G. (1987).

Two roles hypothesized for surface characteristics, such as color, brightness, and texture, in object recognition are that such information: a) can define the gradients needed for a 2 1/2-D sketch so that a 3-D representation can be derived (e.g., Marr & Nishihara, 1978) and b) provide additional distinctive features for accessing memory. In a series of five experiments, subjects either named or verified (against a target name) brief (50-100 msec.) presentations of slides of common objects. Each object was shown in two versions: professionally photographed in full color or as a simplified line drawing showing only the object's major components (which typically corresponded to its parts). Although one or the other type of picture would be slightly favored in a particular condition of exposure (duration or masking), overall mean reaction times (RTs) and error rates were virtually identical for the two types of stimuli. These results support a view that edge-based representations mediate real-time object recognition in contrast to surface gradient or multiple cue representations. A previously unexplored distinction of color diagnosticity allowed us to determine whether color (and brightness) were employed as additional features in accessing memory for those objects or conditions where there might have been an advantage for the color slides. For some objects, e.g., banana, fork, fish, camera, color is diagnostic as to the object's classification. For other objects, e.g., chair, pen, mitten, bicycle pump, color is not diagnostic, as such objects can be of any color. If color was employed in accessing memory, color diagnostic objects should have shown a relative advantage when presented as color slides compared to the line drawing versions of the same objects. Also, this advantage would be magnified when subjects could anticipate the color of an object in the verification task, particularly on NO trials when the foil was of a different color. Neither an overall advantage for color-diagnostic objects when presented in color nor a magnification of a relative advantage on the NO trials in the verification task was obtained. Overall, any advantage to depiction by color slides over a line drawing version was equivalent for diagnostic and

nondiagnostic objects. Although differences in surface characteristics such as color, brightness, and texture can be instrumental in defining edges and are powerful determinants of visual search, they play only a secondary role in the real-time recognition of an intact object when edges can be readily extracted.

B. **The Perception of Objects that Require Surface Feature Specification.** (w. John Hilton). Most of our prior empirical work concentrated on those objects whose representations could be completely specified by a volumetric description, such as a frying pan, horse, or nail. Some objects, such as a racquet, zebra, or screw, require a texture specification in addition to their volumetric specification. We compared the speed and accuracy of object naming of 100 msec exposures of line drawings of these two classes of objects. The objects were matched in their silhouette and general volumetric description but, in one case, a surface description was required as well, as shown in Figure 6. Some examples:

Volumetric Alone	Volumetric + Texture
Nail	Screw
Horse	Zebra
Knife	Nail File
Frying Pan	Racquet
Lock	Basket
Shovel	Broom
Bed	Accordion
Lion	Tiger

Care was taken to evaluate possible effects of familiarity and frequency. If the texture region was functioning as another component, then performance should have been facilitated through the additional geon, in that complex objects can be identified more rapidly than simple ones (Biederman, 1987a and described below). Alternatively, the detailed processing required to specify the texture field might not be completed in a brief exposure duration so such objects might prove to be less recognizable. The results supported the latter alternative and are consistent with the previously reported secondary status of surface features. Mean RTs and error rates for objects that could be specified by a volumetric structure alone were 858 msec and 7.5 percent, respectively. The corresponding values for the those objects requiring a texture field was 980 msec and 19.0 percent errors.

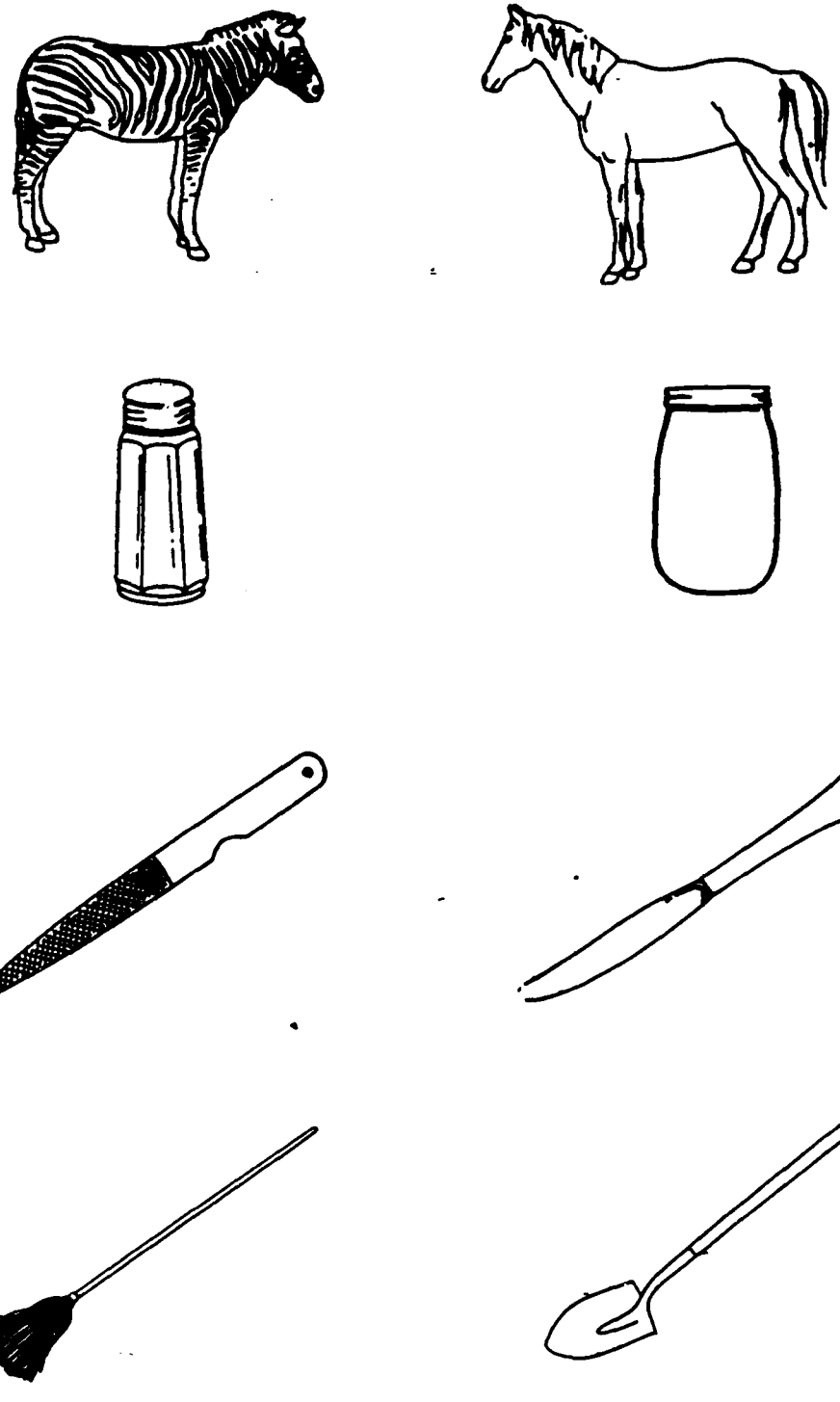
The advantage for the objects that could be recognized with a volumetric description alone supports our original contention that the earliest or most efficient access to memory for an image might be an edge-based description. By this account surface characteristics offer only secondary routes to object recognition.

V. The Perception of Partial and Degraded Objects

Background: Partial Objects and the Effects of Complexity. Complex objects, defined as those requiring six or more components to appear complete, as an airplane or a penguin, could be identified perfectly from only two or three of their geons, as long as subjects were not stressed to respond quickly (Biederman, 1987a). Under speed stress and with brief (100 msec) exposures, both naming reaction times (RTs) and errors increased with the removal of additional components from the complete versions. But even under these conditions, complex objects with less than half their components were accurately named on 75 percent of the trials.

TEXTURE

NOT REQUIRING TEXTURE



983 (18)

858 (8)

FIG. 6.

Importantly, for the complete versions of the objects, complex objects were identified more rapidly than simple objects (those requiring only two or three components to appear complete.

A. **The effect of an inappropriate geon: Consequences of Three Geon Sufficiency** (w. Elizabeth Beiring, Ginny Ju, and Thomas Blickle). A consequence of the three-geon rule is that the addition of a fourth but inappropriate geon (middle column of figure 7) should not result in reduced recognition speed. The three appropriate geons (left column, Fig. 7) will be sufficient to activate the object's representation and unless the inappropriate geon results in the activation of a competing object, no interference should occur, even though that same geon would facilitate the recognition speed of an object when it was appropriate. This prediction assumes that there is no bottom-up inhibition from geons to objects. An experiment recently confirmed this expectation. Although the addition of a fourth component reduced RTs and error rates in the 100 msec identification of an object when that component was relevant, there was no effect of that geon when it was irrelevant.

B. **The Effects of Contour Deletion: A Function of Visual Angle or Proportion Removed?** (w. Tom Blickle).

Biederman & Blickle (described in Biederman, 1987a) found that the deletion of contour, even when it could be restored through collinearity or smooth curvature, resulted in considerable interference in the speed and accuracy of object identification. Moreover, the amount removed had large and consistent effects between a range of 25 to 65 percent deletion.

When we removed greater amounts of contour, the gap sizes also increased along with the greater proportions. We studied whether the effects were do to the larger gap sizes (in terms of visual angle) or proportion of the contour that was removed by expanding the images so that, for example, the 25 percent deletion condition matched, in gap size, the gaps of the original 45 percent condition. The results were clear: Only the proportion of an object's contour that was removed had any effect. There was no independent effect of the retinal gap that had to be bridged.

C. **Comparing Incidental vs. Viewpoint Invariant Image Features for Object Recognition.** (T. Blickle).

Tom Blickle has completed an extensive series of experiments examining the effects of contour deletion on object recognition. The major focus of this effort was on the comparative effects of deleting contour that would affect the nonaccidental characterization of the image versus metric and incidental aspects. Also under examination was the relative importance of contour that would be important for segmentation (viz., cusps) versus contour that would be instrumental for defining the geon. Blickle is currently writing up this research.

D. **Perceiving degraded vs. partial objects: Modeling Activation in Cascade** (Biederman, Gagnon, & Hilton, 1987). The model of RBC illustrated in figure 1d can be partitioned into two critical stages: a) those processes leading to and including the determination of the geons, and b) those processes involved in matching an arrangement of geons to memory.

Consider figure 8 which shows, for some sample objects, one version in which whole components are deleted so that only three (of six or nine) of the components remain and another version in which the same amount of contour is removed, but in midsegment distributed over all of the object's components. In objects missing components, the components cannot be added prior to recognition. Logically, one would have to know what object was being recognized to

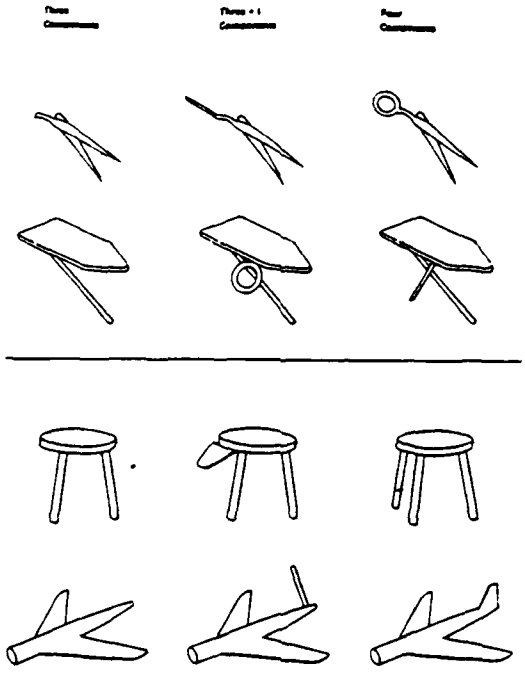


Figure 7. Left. Three-geon versions of complex objects requiring six or nine geons to look complete (left). Right. Four-geon versions of the same objects. Middle. Versions where a fourth--and inappropriate--geon from another object has been added to the three-geon version. Recognition error rates and naming reaction times were equivalent for the 3 and 3+I versions but error rates and RTs were lower for the 4 geon versions.

know what parts to add. With the midsegment deletion, components can be determined from processes employing collinearity or smooth curvature.

The two methods for removing contour may thus be affecting different stages. Deleting contour in midsegment affects processes prior to and including those involved in the activation of the geons (Fig.1d). The removal of whole components (the partial object procedure) is assumed to affect the matching stage, reducing the number of common components between the image and the representation and increasing the number of distinctive components in the representation.

The two stages can be regarded as being arranged in cascade, with an earlier geon determination stage relaying activation on the object matching stage. Figure 9 shows the expected activation functions from the two procedures for deleting contour. Deleting contour in midsegment results in an initial slow growth in activation as the relatively slow processes for smooth continuation are required to restore the deleted contours. Once the restoration is complete there is a rapid growth in activation at the object representation stage. By contrast, there is an initial rapid activation of the components from the partial objects which, however, asymptotes below the activation level of the midsegment deleted objects. The reason for this is that the missing components have activation levels of zero. Once the filling-in is completed for the objects with midsegment deletion, the complete complement of an object's components are available, providing a better match to the object's representation than is possible with a partial object that had only a few of its components. The net effect is to produce a crossover interaction over exposure duration which produces a similar effect on the next stage, activation of the representation of the object.

This prediction was supported from the results of an experiment (described in Biederman, 1987a) which studied the naming speed and accuracy of six- and nine-component objects undergoing these two types of contour deletion. At brief exposure durations (e.g., 65 msec) performance with partial objects was better than objects with the same amount of contour removed in midsegment both for errors and RTs. At longer exposure durations (200 msec), the RTs reversed, with the midsegment deletion now faster than the partial objects.

VI. Variation within Object Classes (w. John Hilton)

One hindrance to the development of a mature science of image understanding is the current absence of clear criteria--or even consciousness--by which presumed image processing operations or theoretical claims are to be evaluated. Perhaps the most common "method" by which operations or claims have traditionally been offered for evaluation is the appeal to the *Method of Casual Viewing* (Biederman, 1988). (This has also been termed the "beauty pageant method.") With this method, a processed image is presented for viewing on a page and the reader is invited to identify it. Typically the image would have information that is irrelevant to the theory or not passed by the filter deleted from it. An accurate identification of the image is supposed to be interpreted as support for the author's theory of the effectiveness of the detector.

Perhaps the only task that such a method is relevant to is that of "ultimate identifiability" (Biederman, 1988a). That is, one can conclude that there is sufficient information in the image to allow a classification but virtually nothing can be concluded about the nature and efficiency of the processing that produced this classification. In particular, one cannot conclude anything about the processes by which an original image of an object is initially recognized--what Biederman (1987a) termed "primal access." The reason for this is that the human has been characterized as possessing a number of routes ("bag of tricks") through which he or she can achieve recognition. *These routes can differ greatly in the amount of time and attentional effort*

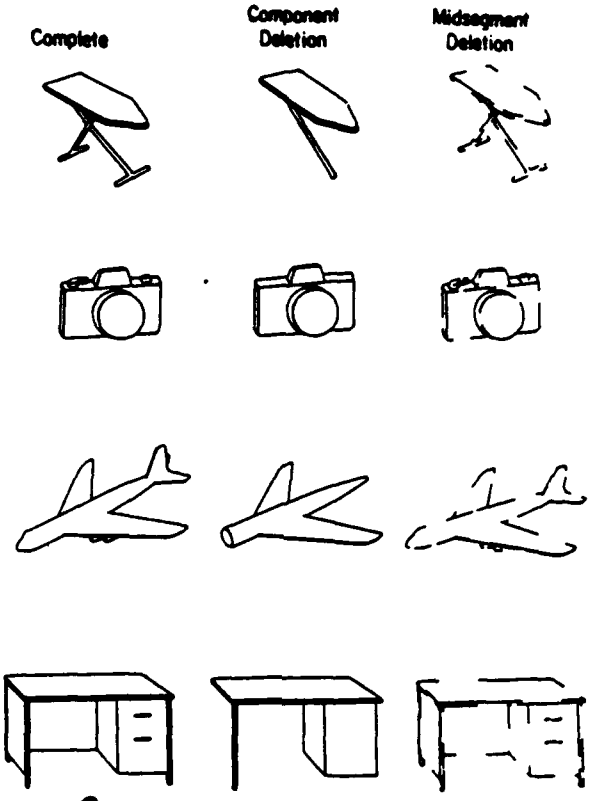


Figure 8. Sample versions of objects with midsegment *deletion and component (geon) deletion.*

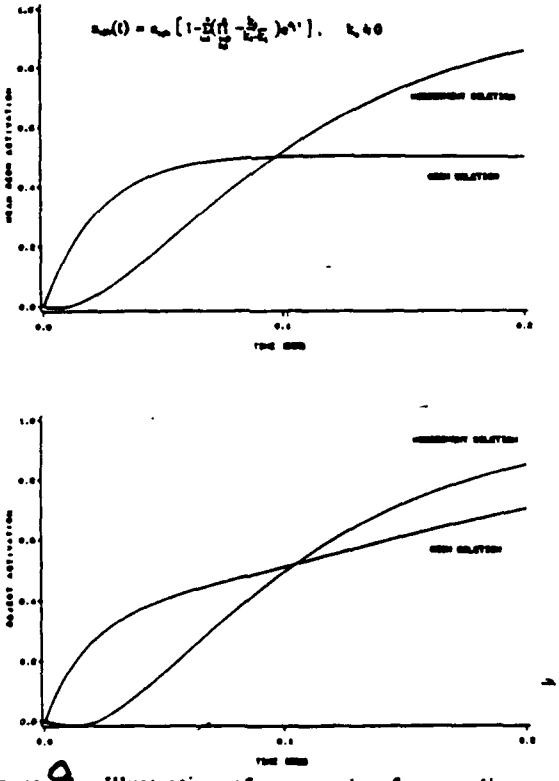


Figure 9. Illustration of a cascade of an earlier geon activation stage and a subsequent object representation stage. The activation of the geons causes activation of the object representations. Component deletion results in a lower asymptote at the geon activation stage because missing components never get activated *prior to object activation.*

they require to achieve recognition. What makes the method of casual viewing especially inappropriate is that *dramatic differences in processing time and effort are not readily available to casual introspection.*

Consider, as an example, Attneave's (1954) oft reproduced image of a cat, as shown in the upper left panel of Figure 10, which was drawn by connecting points at extrema of curvature. (This is generally equivalent to the replacement of curved lines with straight lines.) The subjective ease by which most readers can identify this image has lead to a conclusion, by many, that such points are sufficient for primal--not just ultimate--access, in general.

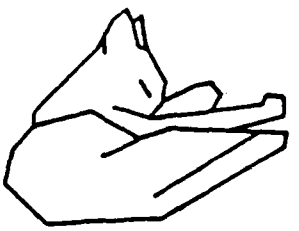
Biederman and Hilton (described in Biederman 1988a) measured the speed and accuracy of naming Attneave's image (as "CAT") and a number of other pictures from a 100 msec exposure immediately followed by a mask of random appearing lines. The mean correct naming reaction times (RTs) and errors (in parenthesis) are shown under the particular images. Attneave's cat averaged 1078 msec with a 17% error rate. Removing the eye slit increased error rates to 42% (the RTs at such a high error rate are unstable). David Lowe's cat, drawn by connecting the midpoints of Attneave's cat, had an error rate of 39%. Shown at the right are images of three cats in which curved edges are present. (There never was an original image of Attneave's cat so three pictures were selected from general sources.) These latter images are recognized perfectly with RTS that averaged about 300 msec less than those required for Attneave's cat. (This result--a large detrimental effect on object recognition performance from the straightening of image curves--can be shown to hold, in general, for any curved object.)

Similar results were found for a set of images of various chairs, including several published by Kolers (Figure 11). Here the major variable was not the elimination of curved edges but the prototypicality of the exemplar. An office chair, made up only of simple volumetric parts, could be named in 684 msec, with 0 percent errors. Kolers' rocking chair required 1129 msec with 53 percent errors! Similar variation was round for a set of lamps (fig. 8).

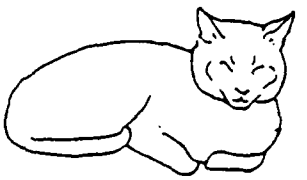
To recapitulate, an image processing operation of which we are subjectively unaware can actually require several times the additional perceptual processing time than that required for the original image. A correction for the higher error rates would produce an even greater value of the additional processing time. This phenomenon--large perceptual processing consequences from subjectively innocuous image processing operations--can be shown to hold for a large number of image processing operations.

The approximately 300 msec increase in naming RTs for the modified or stylized images, though representing an increase of 43 percent over the RTs for the original images, actually represents a considerably greater increase when only the central time for recognition is considered. At least half of the 700 msec mean naming RTs for the standard or original versions of the objects is used for initial sensory registration and the selection and execution of an overt naming response. The 300 msec increase in RTs then represents an 86 percent increase in the time required for recognition. The problem here is that the increased time for the stylized or modified images may allow them to be recognized as symbols or through inference, rather than through a direct mapping of a representation of image contours to a stored representation of an object.

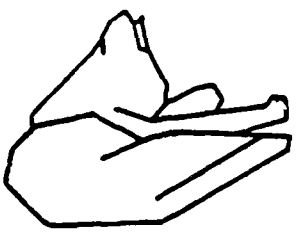
The classic issue addressed by this experiment concerns how an infinitely variable class--such as chairs or cats--can be recognized. Neisser referred to such classes as "ill defined" and it was certainly what Kolers had in mind in presenting his page of chairs. The answer offered by this study--and RBC--is that the mental representation of an object does not include all the



1078 (17)



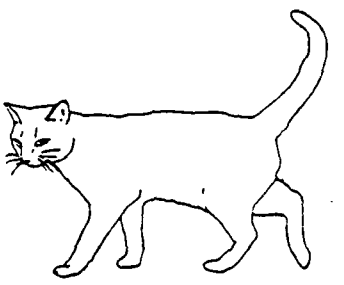
689 (0)



845 (42)

697 (0)

939 (39)



714 (0)

Figure 10. Six cats with their naming reaction times and error rates (in parenthesis) from a 100 msec. masked exposure. Upper left. The original version of Attneave's (1954) cat constructed by connecting points at extrema of curvature. *Middle left:* Removal of the eye slit resulted in markedly higher error rates. *Lower left:* Lowe's (1984) cat constructed from Attneave's by connecting points midway between points at extrema of curvature. *Right side:* Three cats with curved regions retained. These could be identified perfectly with markedly lower reaction times. The middle cat on right is from Snodgrass and Vanderwart (1980).

possible details and contour variations. Recognition is achieved despite such image variations rather than because of it. There may be representations for one or several prototypical exemplars and images are matched to those exemplars. To the extent that the images lack detail specified by the mental representation or to the extent that the images include detail that are not in the mental representation, recognition will be slow and, under brief exposure durations, likely to be in error.

VII. ATTENTIONAL DEMANDS OF OBJECT RECOGNITION

A. **Visual Search for Geons** (Ju & Biederman).

Ginny Ju is currently collecting data for her dissertation exploring the attentional demands of object perception using the Treisman search paradigm. In an already completed experiment, the subject had to attempt to detect a given geon, e.g., a curved cylinder. In the disjunctive condition, the distractors for that target might be bricks and cylinders, where the cross section of one distractor could not recombine with the axis to form the target by illusory correlation. In the conjunctive condition, such an illusory correlation was possible in that the distractors (for the curved cylinder) might be cylinders and curved bricks. Display size was varied from 4 to 16 objects. The results showed an large increase in RTs and error rates as a function of display size in the conjunction condition. RTs and error rates were hardly affected by display size in the disjunctive condition. This would suggest that attention is required to detect the geons. (A cautionary note on these results: The stimuli were not antialiased. The volumes, particularly those with a curved cross section, had a jagged appearance. The current experiments are being run on the Mac IIs which have higher resolution and better quality images.) Ju's experiments will be a major examination of the problems of shape recognition is multielement displays. She will be exploring conjunctive costs across geons and relations.

The preliminary results suggest that at least part of the increase in object detection RTs and error rates in multiobject, nonscene displays (Biederman, Blickle, Teitelbaum, Klatsky, & Mezzanotte, 1988) is a function of the attention required to determine the geons.

B. **Visual Search for Objects.**

Biederman, I., Blickle, T. W., Teitelbaum, R. C., Klatsky, G. J., & Mezzanotte, R. J. (1988). Object identification in multi-object, nonscene displays.

When we look at a chair or a giraffe we cannot suppress a semantic interpretation of that image, although we need not name it (e.g., Smith & McGee, 1980). Given that classification of object images is mandatory, is it capacity free? A picture analog to the Egeth, Jonides, and Wall (1972) letter-digit classification experiment was run in which subjects attempted to detect the presence or absence of a target object, specified by basic-level name, in a nonscene (clockface) arrangement of pictures of common objects. The number of objects varied from one to six. Presentation duration was 100 msec. There was a sharp monotonic decrease in detectability as a function of the number of objects in the display, indicating that object detection under these conditions is an attention-demanding process. This result was unconfounded with similarity because larger displays were constructed by adding objects in order of decreasing similarity to the target. The target object was either consistent with the other objects in the field in that it would be relatively likely to appear in a setting which contained those objects, e.g., a target of TEA KETTLE among kitchen objects such as a stove, toaster, frying pan, and spice rack; or inconsistent, such as TRACTOR among the same objects. Although consistent targets have been found to be more readily detected than inconsistent targets in real-world scenes (e.g., Biederman et al, 1982), inconsistent targets were slightly more detectable in the nonscene displays used in the present investigation. This latter result is evidence against an account of the perceptual

interference found for inconsistent (or improbable) objects in real-world scenes which holds that the interference derives from an inventory listing of the objects without regard to their spatial relations. A geon cluster hypothesis is proposed to account for the rapid activation of a scene's semantic representation without an attentional cost from the number of objects.

VIII. Expert Visual Identifications.

Biederman, I., & Shiffrar, M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual learning task.

The sexing of day old chicks has been regarded as an extraordinarily difficult perceptual task requiring years of extensive practice for its mastery. Experts can sex chicks at over 98 percent accuracy at a rate of 1,000 chicks per hour spending less than a half second viewing the cloacal region. A group of naive subjects were shown 18 pictures of cloacal regions of male and female chicks (in random appearing arrangement) and asked to judge the sex of each chick. The pictures included a number of rare and difficult configurations. The subjects were then briefly instructed as to the location of a critical cloacal structure for which a simple contrast in shape (convex vs concave or flat) could serve as an indicant of sex. When the subjects judged the pictures again (in a different order), accuracy increased from 60.5 to 84.0 percent, a value that matched the performance level of a group of professional sexers with these pictures. The correlation (over items) between the naive subjects and the professionals before instruction, was .21; after instruction, .82. The instructions were based on an interview and observation of an expert (HC) who had spent 50 years sexing 55 million chicks. Much of the reported difficulty in developing perceptual expertise in this task may stem from the need to classify extremely rare configurations in which the convexity of the structure is not apparent. It is possible that the rate of learning of these instances could be greatly increased through the use of simple instructions, such as those used in the present investigation, that specified the location of diagnostic contour contrasts. A parallel is drawn between learning to sex chicks and learning to classify tanks as friend or foe.

IX. An Extension to Scene Perception (Biederman, 1988a).

The mystery about the perception of scenes is that the exposure duration required to have an accurate perception of an integrated real-world scene is not much longer then what is typically required to perceive individual objects. The recognition of a visual array as a scene requires not only the identification of the various entities but also a semantic specification of the interactions among the object and an overall semantic specific of the arrangement.

However, the perception of a scene is not, in general, derived from an initial identification of the individual objects comprising that scene. That is, in general we do <u>not</u> first identify a stove, refrigerator, and coffee cup, in specified physical relations and <u>then</u> come to a conclusion that we are looking at a kitchen.

Some demonstrations and experiments suggest a possible basis for understanding rapid scene recognition. Mezzanotte showed that a readily interpretable scene could be constructed from arrangements of single geons that just preserved the overall aspect ratio of the object, such as those shown in Figure 12. In these kinds of scenes, none of the entities, when shown in isolation, could be identified as anything other than a simple volumetric body, e.g., a brick. Most important, Mezzanotte found that such settings were sufficient to cause interference effects on the identification speed of intact objects that were inappropriate to the setting.

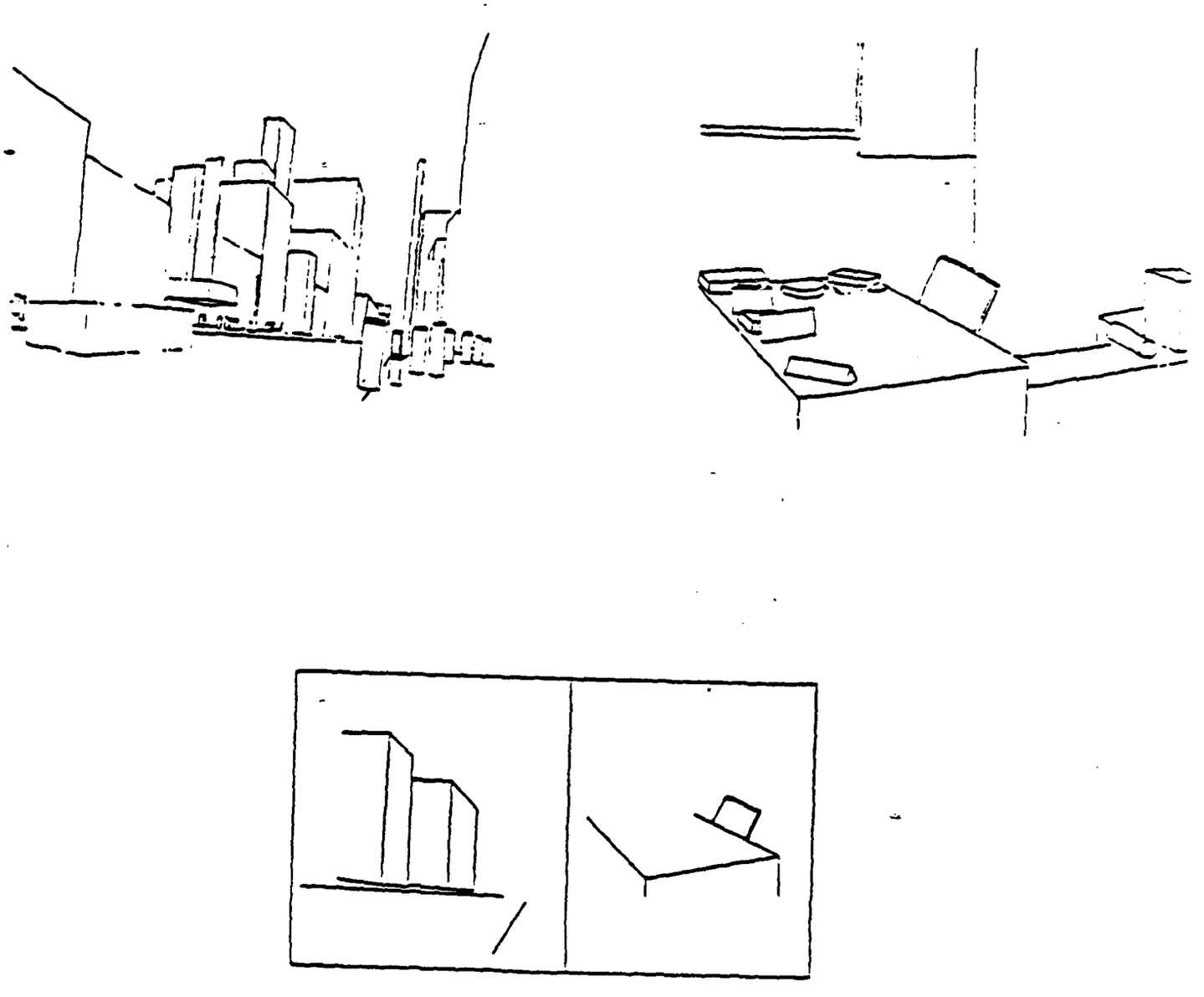


Figure 11-12. Upper portion. Two of Mezzanotte's scenes. "City Street" and "Office." Lower portion. Possible geon clusters for the scenes in above.

We have been exploring the possibility that quick understanding of a scene is often mediated by the perception of geon clusters. A geon cluster is an arrangement of geons from different objects that preserve the relative size and aspect ratio and relations of the largest visible geon of each object. In such cases, the individual geon will be insufficient to allow identification of the object. However, just as an arrangement of two or three geons almost always allows identification of an object, an arrangement of two or more geons from different objects may produce a recognizable combination. The cluster acts very much as a large object. Figure 12 shows two examples. If this account is true, fast scene perception should only be possible in scenes where such familiar object clusters are present. This account awaits empirical test.

X. PUBLICATIONS AND PRESENTATIONS OF GRANT RESEARCH

PUBLICATIONS

Biederman, I. (1986). Recognition-by-Components: A Theory of Human Pattern Recognition. In G. H. Bower (Ed.) The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 21, Pp. 1-54. New York: Academic Press.

Blickle, T. W., & Biederman, I. (1986). Perceiving degraded objects. Proceedings of the Human Factors Society, 30, 301-305.

Ju, G., & Biederman, I. (1986). The perception of objects depicted by color or line drawings. Proceedings of the Human Factors Society, 30, 297-300.

Biederman, I. (1986). Computational theories of vision research: An Overview. In P. Ebert-Flatteau (Ed.) New Approaches to Vision (Pp. 61-62). Washington, D.C.: National Academy Press.

Biederman, I. (1987a). Recognition-by-Components: A Theory of Human Image Understanding. Psychological Review, 94, 115-147.

Biederman, I. (1987b). Matching Image Edges to Object Memory. In Proceedings of the First International Conference on Computer Vision, pp. 384-392, IEEE Computer Society, London, England, June, 1987. [Fourth Workshop on Human and Machine Vision.]

Biederman, I. (1987c). Review of Visual Cognition by S. Pinker (Ed.). American Journal of Psychology, 101, 146-148.

Biederman, I. (1987d). Human image understanding: Recent Research and a theory. In A. Rosenfeld (Ed.) (1987). Human and Machine Vision II. (pp. 13-57). Orlando, Fl.: Academic Press. [Reprinted from Computer Vision, Graphics, and Image Processing, 1985, 32, 29-73.]

Biederman, I., & Shiffrar, M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual learning task. Journal of Experimental Psychology: Human Learning, Memory, and Cognition, 13, 640-645.

Biederman, I., & Ju, G. (1987). Surface vs. Edge-Based Determinants of Visual Recognition. Cognitive Psychology, 20, 38-64.

Biederman, I., Blickle, T. W., Teitelbaum, R. C., Klatsky, G. J., & Mezzanotte, R. J. (1987). Object identification in nonscene displays. Journal of Experimental Psychology: Human Learning, Memory, and Cognition, 14, 456-467.

Biederman, I. (1988a). Aspects and Extensions of a Theory of Human Image Understanding. In Z. Pylyshyn, (Ed.) Computational Processes in Human Vision: An Interdisciplinary Perspective. Pp. 370-428. New York: Ablex.

Biederman, I. (1988b). Factors affecting target identification in pilot performance modeling. In J. Elkind, S. Card, J. Hochberg, & B. Huey (Eds.) Human Performance Models *for Computer-Aided Engineering*. Washington: National Academy Press.

Biederman, I., Blickle, T. W., Ju, G., Hilton, H. J., & Hummel, J. E. (1988). Empirical analyses and connectionist modeling of real-time human image understanding. Proceedings of the Tenth Annual Conference of the Cognitive Science Society.

Biederman, I., Rosenfeld, A., & Flattau, P. E. (1988). Visual processing issues in computer-aided target recognition. National Research Council, Committee on Vision Report.

Hummel, J. E., Biederman, I., Gerhardstein, P. C., & Hilton, H. J. (1988). From image edges to geons: A connectionist approach. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.) *Proceedings of the 1988 Connectionist Models Summer School*. Pp. 462-471. San Mateo, CA: Morgan Kaufmann.

Biederman, I. (1989a). Higher level vision. In D. N. Osherson, H. Lasnik, S. Kosslyn, J. Hollerbach, E. Smith, & N. Bloch (Eds). *An invitation to cognitive science*. Cambridge, MA: MIT.

Biederman, I. (1989b). Algorithms for Vision. Review of W. Richards & S. Ullman (Eds.) Image Understanding, 1985-86. Contemporary Psychology, In press.

Biederman, I. (1989c). The uncertain case for cross cultural effects in pictorial object recognition. *Behavioral and Brain Sciences*, in press.

PRESENTATIONS AT SCIENTIFIC MEETINGS

Biederman, I. Human Image Understanding. Invited address to the LOVE Conference (Lake Ontario Visionary Establishment), Niagara Falls, Ontario, Canada, February, 1986.

Biederman, I. Cognitive Science. Presentation to the Air Force Office of Scientific Research, Research Initiatives Review Board, March, 1986.

Biederman, I. Human Image Interpretation. Workshop conducted at the Advanced Institute for Artificial Intelligence, Western Ontario, Canada: April 1986.

York, J. L., & Biederman, I. Hand movement speed and accuracy as a function of gender, alcohol use, and aging. Paper presented at the meetings of the American Medical Society on Alcoholism and the Research Society on Alcoholism, San Francisco: April, 1986.

Biederman, I. Computational Approaches to Vision. Symposium on New Frontiers in Vision, National Academy of Sciences, Washington, D.C.: April, 1986.

Biederman, I. Image Interpretation. Invited address to the Human Perception and Performance Workshop for System Designers, Research Institute, University of Dayton, June, 1986.

Biederman, I. Visual perception and performance in space: Some methodological and empirical problems. Invited paper at the Exploratory Meeting on Changes in Vision During Long-Term Space Flight. NASA Ames (California): June, 1986.

Blickle, T. W., & Biederman, I. Perceiving degraded objects. Paper presented at the Meetings of the Human Factors Society, Dayton, Ohio: September, 1986.

Ju, G., & Biederman, I. The perception of objects depicted by color or line drawings. Paper presented at the Meetings of the Human Factors Society, Dayton, Ohio: September, 1986.

Biederman, I. Aspects and extensions of a theory of human image understanding. Paper presented at the First Queens University Vision Conference, Kingston, Ontario: April, 1987.

Biederman, I. The role of scene perception in the analysis of aquatic injuries. Invited address to the Aquatic Injury Safety Association, Ft. Lauderdale, Florida: May, 1987.

Biederman, I. Matching Image Edges to Object Memory. Paper presented at the First International Conference on Computer Vision, IEEE Computer Society [Fourth Workshop on Human and Machine Vision.]. London, England: June, 1987.

Biederman, I. (1987). Perceptual and attentional factors in the processing of multisensor displays. Invited address to the National Academy of Sciences-National Research Council Committee on Vision and U. S. Army Workshop on human Processing of Computer Aided Target Images. Aberdeen Proving Grounds, Maryland: July.

Biederman, I. (1987). Real-time human image understanding and pilot performance models. Invited presentation to the NASA-NRC Working Group on Pilot Performance Models. Washington, D. C.: Dec.

Biederman, I. (1987). Human image understanding. Invited paper to the Air Force Review of Vision Research. Annapolis, MD.: December.

Biederman, I. (1988). Computational and empirical analyses of human image understanding. Invited address to the Twin Cities Special Interest Group in Artificial Intelligence. Minneapolis, MN.: January.

Biederman, I. (1988). A psychologist looks at belief in the paranormal. Invited address to the Minnesota Skeptics. Minneapolis, MN.: January.

Biederman, I. (1988). Invariant Visual Primitives for Object Recognition. Invited address to the AAAS Symposium on High-Level Vision; Interdisciplinary Approaches to Object Recognition. Boston, Massachusetts: February.

Biederman, I. (1988). Geons: Visual Primitives for Human Image Understanding. Invited paper presented at the Meeting for Visual Form and Motion Perception: Psychophysics, Computation, and Neuro Networks. Boston, Massachusetts: March.

Biederman, I. (1988). Real-time human image understanding. Paper presented at the Spring Symposium on Physical and Biological Approaches to Computational Vision, American Association for Artificial Intelligence. Stanford, CA.: March.

Biederman, I. (1988). Human image understanding. Invited address to the Meetings of the Midwestern Psychological Association. Chicago, IL.: April.

Biederman, I. (1988). Visual object recognition. Invited presentation at the James S. McDonnell Foundation Summer Institute in Cognitive Neuroscience. Harvard University: June.

INVITED COLLOQUIA: 1986-88

Columbia University

Duke University

University of Rochester

University of Minnesota [Two colloquia: Engineering and Computer Science, and Psychology].

University of Maryland

Georgia Technical Institute

Brandeis University

University of Pittsburgh

MacMaster University

University of Toronto

Cambridge University (APU)

University of London

Harvard University

University of South Florida

Columbia University (Animal Cognition Group)

Honeywell

MIT (Psychology Department and the Center for Biological Information Processing)

US Army Research Institute, Ft. Benning, Ga

Stanford University

McGill University (Departments of Electrical Engineering; Cognitive Science Program)

University of Auckland, New Zealand.