MICROCOPY RESOLUTION TEST CHART

REPORT DOCUMENTATION PAGE

	1b. RESTRICTIVE MARKINGS
AD-A172 002	
	3 DISTRIBUTION/AVAILABILITY OF REPORT
	Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
	AFOSR-TR- 86-0679

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
State Univ. of New York/Buffalo		Air Force Office of Scientific Research/NL

6c. ADDRESS (City. State and ZIP Code)	7b. ADDRESS (City, State and ZIP Code)
Psychology Dept., 4230 Ridge Lea Road Amherst, New York 14226	Building 410 Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR	NL	F49620-83-C-0086

8c. ADDRESS (City. State and ZIP Code)	10 SOURCE OF FUNDING NOS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
Building 410 Bolling AFB DC 20332-6448	61102F	2313	A5	

11. TITLE (Include Security Classification)
Human Information Processing of Targets and Real-World Scenes

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 4/1/83 TO 8/31/85	85 July 30	80

16. SUPPLEMENTARY NOTATION

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

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

Substantial progress has been made on an empirical and theoretical analysis of human image understanding. The theory, termed Recognition-by-Components (RBC), holds that the perceptual recognition of objects is a process in which the image of the input is segmented at regions of deep concavity into simple volumetric components. These components can be derived from properties of the two dimensional image that are invariant over viewing position and image quality, such as collinearity and symmetry. Experimental results support the sufficiency of RBC in showing efficient speeded recognition of objects missing parts or lacking color and texture. Also confirmed was a prediction derived from RBC that selective contour deletion that bridged concavities and prevented retrieval of the components would render object identification impossible.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr. John F. Tangney	(202) 767-5021	NL

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE

FINAL TECHNICAL REPORT
F49620
30 JULY 1985

AFOSR·TR· 86-0679

HUMAN INFORMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES

By:

Irving Biederman
Professor of Psychology
State University of New Yorik at Buffalo
4230 Ridge Lea Road
Amherat, New York 14226

Contract F49620-83-C-0086

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
... ... MITTAL TO DTIC
... ... ical report has been reviewed and is
.... for public release IAW AFR 190-12.
Distribution is unlimited.
... THEW J. KERPER
Chief, Technical Information Division

QUALITY INSPECTED 1

The views and conclusions contained in this document are those of the
author and should not necessarily be interpreted as representing the
official policies or endorsements, either expressed or implied, of the
Air Force Office of Scientific Research of the United States
Government.

Controlling Office:

Air Force Office of Scientific Research/NL
Bolling AFB, D. C. 20332-6448

AUG 0 6 1985

86 9 15 061

Progress Report
HUMAN INFORMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES
AFOSR CONTRACT NO: F4962083C0086
I. BIEDERMAN

OVERVIEW

In the two years since the start date of the contract, we have
launched, to our knowledge, the most extensive investigation of human
image understanding ever undertaken. The major empirical results and
theoretical advances are summarized in the section of Recognition-by-
Components: A theory of Human Image Interpretation." A brief summary
of that research on object recognition follows:

   Theoretical Development: Our working hypothesis, termed
Recognition-by-Components (RBC), assumes that the perceptual
recognition of objects is a process in which the image of the input is
segmented at regions of deep concavity into simple volumetric
components, such as blocks, cylinders, wedges, and cones. This parsed
descriptions is then matched to a representation in memory.   As
initially proposed, subjective measures would have been employed to
discover the components, a methodology similar to the kind employed by
linguists determining the phoneme set for a given language.   This
approach is still useful and important for determining the
psychological reality of any proposed set of primitives.   However,
during the past year, I realized that dichotomous (or trichotomous)
contrasts in five ("nonaccidental") properties of edges in a two-
dimensional image--curvature, collinearity, degree of symmetry,
parallelism, and cotermination--if applied to generalized cones, could
generate a psychologically plausible set of primitive volumes [N
probably $\leq$ 36]. The nonaccidental properties allow strong 3-D
inferences to be made directly from properties of the image (cf. Marr,
1977; Lowe, 1984; Binford, 1981). If image edges (or segments or
points) are collinear, curved, symmetrical, parallel, or coterminate,
then the edges in the world giving rise to those images will always be
interpreted as collinear, curved, symmetrical, etc.

   A remarkable advantage accrues from this conceptualization:
Because the nonaccidental properties are invariant over (all but
extremely unlikely accidents of) viewpoint and noise, the components
generated from them will also be invariant with viewpoint and noise!
This may be why objects can be readily recognized from different
viewpoints or when degraded by noise. RBC thus provides a principled
account of the heretofore undecided relation between the classic
principles of perceptual organization and pattern recognition: The
constraints toward regularization (Pragnanz) characterize not the
complete object but the object's components.

   I calculated upper bound estimates of the number of readily
distinguishable object categories available to humans for quick
("primal access") classification of images ($\approx$ 30,000). The capacity

<center>OVERVIEW (Continued)</center>

of the 36 components to represent these categories was computed, assuming only readily detectable dichotomous (or trichotomous) relations between pairs of volumes. The assumed relations are themselves relatively invariant with viewpoint and noise (e.g., top vs. bottom; joined end-to-end vs. end to side). If only one percent of the possible combinations of components were actually used (i.e., 99 percent redundancy), and objects were distributed homogeneously among combinations of components, then only two or three volumes would be sufficient to unambiguously represent most objects! The problem of object recognition is thus reduced to one of determining a few components in their specified relations, all distinguishable through well-documented perceptual routines.

Empirical Developments:

A massive program of empirical research has been executed (and is ongoing). This is difficult research in that a large number of different, complex, pictorial stimuli have to be designed for each experiment. Because we are studying the recognition of stimuli, we cannot run hundreds or thousands of trials with just a few stimuli-- subjects would soon learn to respond to simple stimulus features and short circuit the memory access underlying recognition. Consequently, we have to run large numbers of subjects with just a modest number of trials for each subject. Under these conditions, we give ourselves a pat on the back for collecting usable data from over 750 subjects for a total of over 60,000 trials, most with reaction time measures. Appendix A presents a summary log of the experiments on object perception. We have concentrated on four kinds of problems:

a) **Partial Objects** (w. Ginny Ju). RBC leads to an expectation that two or three components should be sufficient, in most cases, for rapid though (not necessarily optimal) recognition. Four object-naming reaction-time experiments have documented that this is the case.

b) **Line Drawings vs Colored Photography** (w. Ginny Ju). RBC would hold that a sufficient (and often, necessary) representation for rapid recognition can be described by the line drawing of an object's components. Surface characteristics such as color, texture, or brightness are only secondary routes. Four experiments have documented that objects shown as simple line drawings can indeed be recognized about as quickly as high quality photography.

c) **Degraded Images** (w. Tom Blickle). In most natural cases of modest image degradation, as when an object is viewed behind foliage, the object still remains identifiable. RBC predicts a condition of contour deletion under which recognition should be impossible. If the concavities between components are deleted and the interrupted segments aligned through collinearity or constant curvature, then new components would be defined and the original ones lost. Object recognition should then be impossible. An equivalent amount of contour removed in midsegment or at a vertex where it could be

HUMAN INFORMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES
AFOSR CONTRACT NO: F49620-83-C-0086
I. BIEDERMAN

OVERVIEW (Continued)

restored through collinearity, curvature, or cotermination, should not
be nearly as disruptive. Six experiments provided strong support for
this prediction. Additional results from these experiments documented
a close dependence of object recognition on the amount of contour
deleted even when the contour could be restored through collinearity
or curvature.

d) Transfer across viewpoints (w. <Mary Lloyd, Ginny Ju, & Tom
Blickle). In these experiments, an object is viewed at one
orientation. On a subsequent trial it is presented at the same or at
a different orientation. We initially expected that the benefit on
recognition reaction time from a prior exposure would be a function of
the similarity, in terms of common minus distinctive components,
between the two views. In four studies we have not observed any
effect of orientation, so this prediction could not be tested. We now
believe that the access to a representation of a familiar object is so
fast on its very first experimental presentation that little effect of
a prior exposure in an experiment can be observed. In these
experiments all the objects were familiar. Novel objects, however,
might--and should--reveal a similarity effect. We are currently
designing studies to test this conjecture.

Other Empirical Projects.

Visual Search (w. Brian Fisher). We have started an experiment
on the reaction time for detecting or identifying stimuli as a
function of the size of the visual field that must be attended to and
the number of possible positions. The central question is whether
longer latencies result when the subject has to spread his or her
attention over a greater area, given that the identical stimulus event
occurs in the two cases. Subjects have to press a microswitch
whenever a signal occurs in a go no-go task. In condition A, the
signals are presented $3°$ to the left or right of fixation. In
condition B, the signals are $6°$ left or right of fixation; In
condition C, a signal can occur in any one of the four positions.
Consider the case when a signal occurs $3°$ from fixation in C. Will
RTs be longer then in A where the subject never has to consider the
possibility of signals $6°$ from fixation? Will RTs in B have shorter
latencies then in A? Affirmative answers to these questions would
support the spotlight metaphor of visual attention. If people have
to spread their attention over a broader visual field, there is less
capacity to respond to any one position. Our results indicate that
for simple detection (of a transient), there is absolutely no effect
of spread of attention. Identification may be another matter which we
are now testing.

An Analysis of Learning a Difficult Perceptual Activity: Sex-
Typing Day-Old Chicks (w. Margaret Shiffrar). A classic example of a
difficult perceptual learning activity has been learning how to sex
type day-old chicks. Presumably, it requires two to three years of
training before asymptotic performance levels are reached. In
October, 1984, I learned of an individual, Mr. Heimer Carlson, a

OVERVIEW (Continued)

resident of Petaluma California, who was going into semi-retirement after 50 years of sex typing day-old chicks. He had typed 55 million during his career. On the basis of a case study of Mr. Carlson and his serving as our informant, we were able to develop a testable hypothesis about the nature of the difficulty and a technique of training that would reduce learning time to under 2 min. In tests with six expert sexers and 32 naive undergraduates, we were able to train the naive subjects from chance to a level of performance that was identical to the experienced sexers on typing 18 pictures. The item correlations between the undergraduates and sexers was .88-- meaning that the undergraduates missed the same pictures as the sexers. The fundamental concept of training is a simple one and has been used in training Army personnel to distinguish NATO from Warsaw Pact tanks. The instructional materials must specify; a) where to look, and b) what to look for. If this is done competently, then what seemed to be difficult perceptual activities become relatively simple.

Adage Graphics System. We have spent an enormous amount of time and energy in developing the Adage so that it could be used as a computer based stimulus presentation and development system. This will allow considerable savings in time and effort in stimulus development. The system is now working and we are running the ADAGE's SOLIDS 3000 solids modeling package and PADDLE, a 3-D modeling package from the Production Automation Project (University of Rochester).

Recognition-by-Components: A Theory of Image Interpretation

Irving Biederman

State University of New York at Buffalo

ABSTRACT

The perceptual recognition of objects is conceptualized to be a process in which the image of the input is segmented at regions of deep concavity into simple volumetric components, such as blocks, cylinders, wedges, and cones. The fundamental assumption of the proposed theory, Recognition-by-Components (RBC), is that a modest set of components [N probably $\leq$ 36] can be derived from contrasts of five readily detectable properties of a two-dimensional image: curvature, collinearity, symmetry, parallelism, and cotermination. The detection of these properties is generally invariant over viewing position and image quality and consequently allows robust object perception when the image is projected from a novel viewpoint or degraded. RBC thus provides a principled account of the heretofore undecided relation between the classic principles of perceptual organization and pattern recognition: The constraints toward regularization (Pragnanz) characterize not the complete object but the object's components. Representational power derives from an allowance of free combinations of the components. A Principle of Componential Recovery can account for the major phenomena of object recognition: If an arrangement of two or three primitive components can be recovered from the input, objects can be quickly recognized even when they are occluded, rotated in depth, novel, or extensively degraded. The results from experiments on the perception of briefly presented pictures by human observers provide empirical support for the theory.

RECOGNITION-BY-COMPONENTS: A THEORY OF HUMAN IMAGE UNDERSTANDING

IRVING BIEDERMAN

STATE UNIVERSITY OF NEW YORK AT BUFFALO

Any single object can project an infinity of image configurations to the retina. The orientation of the object to the viewer can vary continuously, each giving rise to a different 2-D projection. The object can be occluded by other objects or texture fields, as when viewed behind foliage. The object can even be missing some of its parts or be a novel exemplar of its particular category. The object need not be presented as a full colored, textured image but instead can be a simplified line-drawing. But it is only with rare exceptions that an image fails to be rapidly and readily classified, either an instance of a familiar object category or an instance that cannot be so classified (itself a form of classification).

A DO-IT-YOURSELF EXAMPLE

Consider the object shown in figure 1. We readily recognize it as one of those objects that cannot be classified into a familiar category. Despite its overall unfamiliarity, there is near unanimity in its descriptions. We parse--or segment--its parts at regions of deep concavity and describe those parts with common, simple volumetric terms, such as "a block," "a cylinder," "a funnel or truncated cone." We can look at the zig-zag horizontal brace as a texture region or zoom in and interpret it as a series of connected blocks. The same is true of the mass at the lower left--we can see it as a texture area or zoom in and parse it into its various bumps.


Insert Figure 1 About Here

Although we know that it is not a familiar object, after a while we can say what it resembles: "A New York City hot dog cart, with the large block being the central food storage and cooking area, the rounded part underneath as a wheel, the large arc on the right as a handle, the funnel as an orange juice squeezer and the various vertical pipes as vents or umbrella supports." It is not a good cart, but we can see how it might be related to one. It is like a ten-letter word with four wrong letters.

We readily conduct the same process for any object, familiar or unfamiliar, in our foveal field of view. The manner of segmentation and analysis into components does not appear to depend on our familiarity with the particular object being identified.

The naive realism that emerges in descriptions of nonsense objects may be reflecting the workings of a representational system by which objects are identified.

Figure 1. A Do-it-Yourself Object. There is strong consensus in the segmentation of this configuration and in the description of its parts.
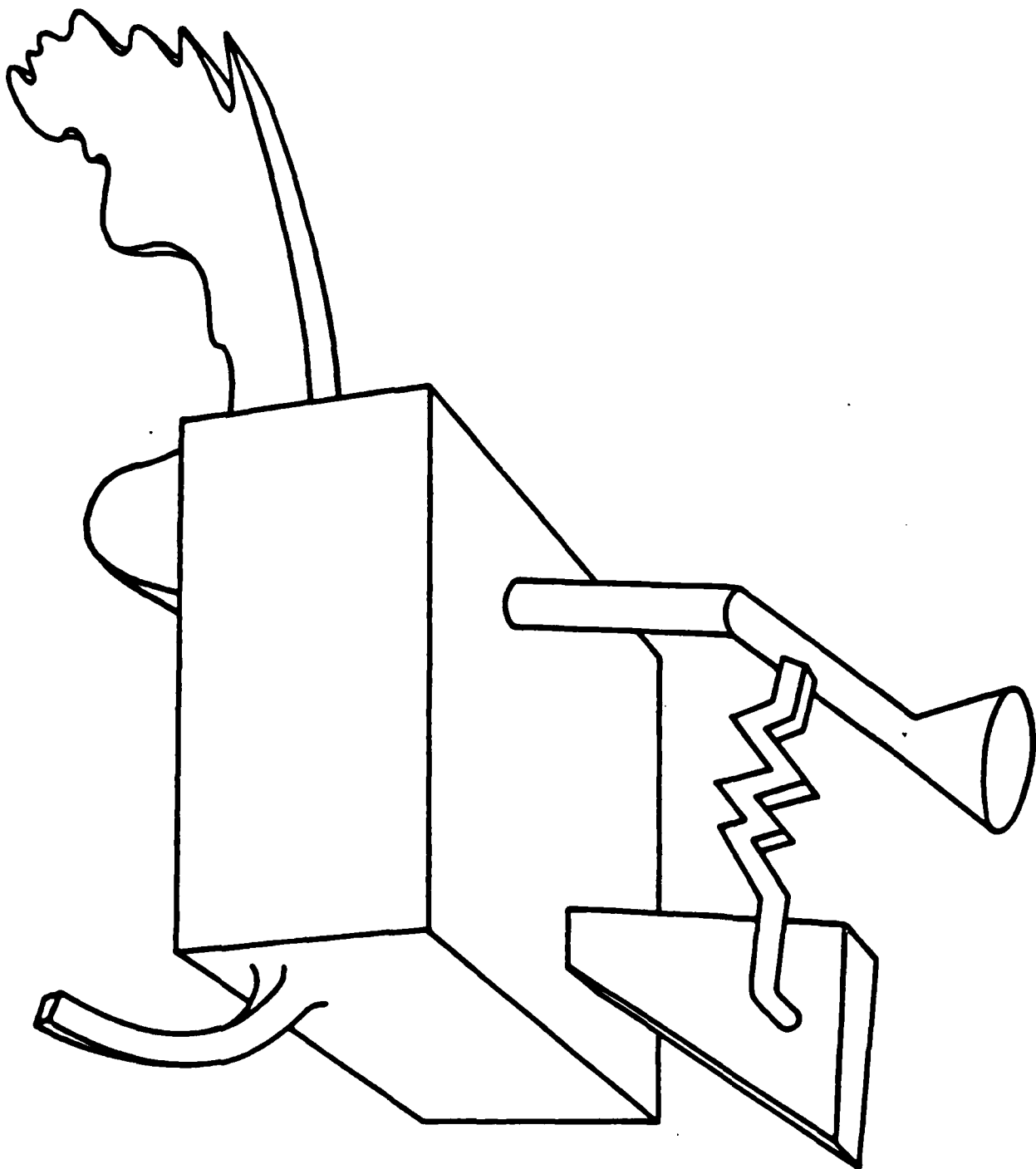
RECOGNITION: UNITS AND CATEGORIES

The number of categories into which we can classify objects would appear to rival the number of words that can be readily identified when listening to speech. Lexical access during speech perception can be successfully modeled as a process mediated by the identification of individual primitive elements, the phonemes, from a relatively small set of primitives (Marslen-Wilson, 1980). We only need about 38 phonemes to code all the words in English, 15 in Hawaiian, 55 to represent virtually all the words in all the languages spoken on Earth. Because the set of primitives is so small and each phoneme specifiable by dichotomous (or trichotomous) contrasts (e.g., voiced vs unvoiced, nasal vs oral) on a handful of attributes, one need not make particularly fine discriminations in the speech stream. The representational power of the system derives from its permissiveness in allowing relatively free combinations of its primitives.

The hypothesis explored here is that a roughly analogous system may account for our capacities for object recognition. In the visual domain, however, the primitive elements would not be phonemes but a modest number of simple volumes such as cylinders, blocks, wedges, and cones. Objects are segmented, typically at regions of sharp concavity and the resultant parts matched against the best fitting primitive. The set of primitives derives from combinations of contrastive characteristics of the edges in a 2-D image (e.g., straight vs curved, symmetrical vs asymmetrical) that define differences among a set of simple volumes (viz., those that tend to be symmetrical and lack sharp concavities). The particular properties of edges that are postulated to be relevant to the generation of the volumetric primitives have the desirable properties that they are invariant over changes in orientation and can be determined from just a few points on each edge. Consequently, they allow a primitive to be extracted with great tolerance for variations of viewpoint and noise.

Just as the relations among the phonemes are critical in lexical access--"fur" and "rough" have the same phonemes but are not the same words--the relations among the volumes are critical for object recognition: Two different arrangements of the same components could produce different objects. In both cases, the representational power derives from the enormous number of combinations that can arise from a modest number of primitives. The relations in speech are limited to left-to-right (sequential) orderings; in the visual domain a richer set of possible relations allows a far greater representational capacity from a comparable number of primitives. The matching of objects in recognition is hypothesized to be a process in which the perceptual input is matched against a representation that can be described by a few simple volumes in specified relations to each other.

THEORETICAL DOMAIN: PRIMAL ACCESS

Our theoretical goal is to account for the _initial_ _categorization_ of isolated objects. Often, but not always, this categorization will be at a _basic_ _level_, for example, when we know that a given object is a typewriter, banana, or giraffe (Rosch, Mervis, Gray, & Boyes-Braem 1976). Much of our knowledge about objects is organized at this level of categorization--the level at which there is typically some readily available name to describe that category (Rosch _et_ _al_, 1976). The hypothesis explored here predicts that in certain cases subordinate categorizations can be made initially, so that we might know that a given object is a floor lamp, sports car, or dachshund, more rapidly than we know that it is a lamp, car, or dog (e.g., Jolicour, Gluck, & Kosslyn, 1984).

The _role_ _of_ _surface_ _characteristics_. There is a restriction on the scope of this approach of volumetric modeling that should be noted. The modeling has been limited to concrete entities of the kind typically designated by English _count_ nouns.   These are concrete objects that have specified boundaries and to which we can apply the indefinite article and number. For example, for a count noun such as CHAIR we can say "a chair" or "three chairs." By contrast, _mass_ nouns are concrete entities to which the indefinite article or number cannot be applied, such as water, sand, or snow. So we cannot say "a water" or "three waters," unless we refer to a count noun shape as in "a drop of water," "a bucket of water," or a "grain of sand", each of which does have a simple volumetric description. We conjecture that mass nouns are identified primarily through surface characteristics such as texture and color, rather than through volumetric primitives.

Under restricted viewing conditions, as when an object is partially occluded, texture, color, and other cues (such as position in the scene and labels), may contribute to the identification of count nouns, as for example, when we identify a particular shirt in the laundry pile from just a bit of fabric. Such identifications are indirect, typically the result of inference over a limited set of possible objects. The goal of the present effort is to account for what can be called _primal_ _access_: the first contact of a perceptual input from an isolated, unanticipated object to a representation in memory.

BASIC PHENOMENA OF OBJECT RECOGNITION

Independent of laboratory research, the phenomena of every-day object identification provide strong constraints on possible models of recognition. In addition to the fundamental phenomenon that objects can be recognized at all (not an altogether obvious conclusion), at least five facts are evident. Typically, an object can be recognized:

1. Rapidly.
2. When viewed from novel orientations.
3. Under moderate levels of visual noise.
4. When partially occluded.

5. When it is a new exemplar of a category.

Implications

The preceding five phenomena constrain constrain theorizing about object interpretation in the following ways.

1. Access to the mental representation of an object should not be dependent on absolute judgments of quantitative detail, because such judgments are slow and error prone (Miller, 1956; Garner, 1966). For example, distinguishing among just several levels of the degree of curvature or length of an object typically requires more time than that required for the identification of the object itself. Consequently, such quantitative processing cannot be the controlling factor by which recognition is achieved.

2. The information that is the basis of recognition should be relatively invariant with respect to orientation and modest degradation.

3. Partial matches should be computable. A theory of object interpretation should have some principled means for computing a match for occluded, partial, or new exemplars of a given category. We should be able to account for the human's ability to identify, for example, a chair when it is partially occluded by other furniture, or when it is missing a leg, or when it is a new model.

RECOGNITION-BY-COMPONENTS

Our hypothesis, Recognition-by-Components (RBC), bears some relation to several prior conjectures for representing objects by parts or modules (e.g., Binford, 1971; Guzman, 1971; Marr & Nishihara, 1978; Tversky & Hemenway, 1984). RBC's contribution lies in its proposal for a particular vocabulary of components derived from perceptual mechanisms and its account of how an arrangement of these components can access a representation of an object in memory.

When an image of an object is painted across the retina, RBC assumes that a representation of the image is segmented--or parsed--into separate regions at points of deep concavity, particularly at cusps where there are discontinuities in curvature (Hoffman & Richards, 1984). Such segmentation conforms well with human intuitions about the boundaries of object parts, as was demonstrated with the nonsense object in figure 1. The resultant parsed regions are then approximated, whenever possible, by simple volumetric components that can be modeled by generalized cones (Binford, 1971). A generalized cone is the volume swept out by a cross section moving along an axis (see Fig 5). The cross section is typically hypothesized to be at right angles to the axis. Secondary segmentation criteria (and criteria for determining the axis of a component) are those that afford descriptions of volumes that maximize symmetry, length, and constancy of the size and curvature of the cross-section of the component. These secondary bases for segmentation and component identification are discussed below.

The primitive components are hypothesized to be simple, typically symmetrical volumes lacking sharp concavities, such as blocks, cylinders, spheres, and wedges. The fundamental perceptual assumption of RBC is that the components can be differentiated on the basis of perceptual properties in the 2-D image that are readily detectable and relatively independent of viewing position and degradation. These perceptual properties include several that have traditionally been thought of as principles of perceptual organization, such as good continuation, symmetry, and Pragnanz. RBC thus provides a principled account of the relation between the classic phenomena of perceptual organization and pattern recognition: although objects can be highly complex and irregular, the units by which objects are identified are simple and regular. The constraints toward regularization (Pragnanz) are thus assumed to characterize not the complete object but the object's components.

By the preceding account, surface characteristics such as color and texture typically have only secondary roles in primal access. This should not be interpreted as suggesting that the perception of surface characteristics per se is delayed relative to the perception of the components but merely that in most cases the surface characteristics are generally less efficient routes for accessing the classification of a count object. That is, we may know that a chair has a particular color and texture simultaneously with its volumetric description, but it is only the volumetric description that provides efficient access to the mental representation of CHAIR.[1]

Relations among the components. Although the components themselves are the focus of this article, as noted previously the arrangement of primitives is necessary for representing a particular object. Thus an arc side-connected to a cylinder can yield a cup as shown in fig 2. Different arrangements of the same components can readily lead to different objects, as when an arc is connected to the top of the cylinder to produce a pail in figure 2. Whether a

---

[1]There are, however, objects that would seem to require both a volumetric description and a texture region for an adequate representation, such as hairbrushes, typewriter keyboards, and corkscrews. It is unlikely that many of the individual bristles, keys, or coils are parsed and identified prior to the identification of the object. Instead those regions are represented through the statistical processing that characterizes their texture (e.g., Beck, Prazdny, & Rosenfeld, 1983; Juleaz, 1981), although we retain a capacity to zoom down and attended to the volumetric nature of the individual elements. The structural description that would serve as a representation of such objects would include a statistical specification of the texture field along with a specification of the larger volumetric components. These compound texture-componential objects have not been studied but it is possible that the characteristics of their identification would differ from objects that are readily defined solely by their arrangement of volumetric components.
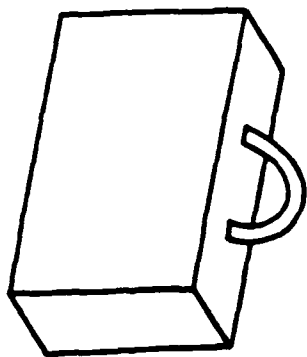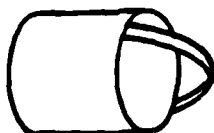
(a)
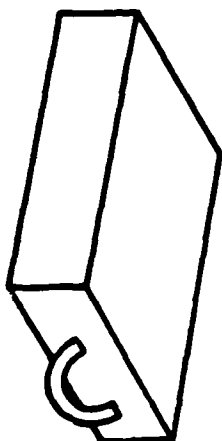
(b)

(c)

(d)

Figure 2. Different arrangements of the same components can produce different objects.

component is attached to a long or short surface can also affect
classification as with the arc producing either an attache case or a
trongbox in figure 2.

' Insert Figure 2 About Here

The identical situation between primitives and their arrangement
exists in the phonemic representation of words, where a given subset
of phonemes can be rearranged to produce different words.

The representation of an object would thus be a structural
description that expressed the relations among the components
(Winston, 1975; Nevatia, 1974; Ballard & Brown, 1982). A suggested
(minimal) set of relations is described in table 1, and would include
specification of the relative sizes of the components and their points
of attachment.

Stages of processing. Figure 3 presents a schematic of the
presumed subprocesses by which an object is recognized. An early edge
extraction stage provides a line drawing description of the object.
From this description, nonaccidental properties of the image,
described below, are detected. Parsing is performed at concave
regions simultaneously with a detection of nonaccidental properties.
The nonaccidental properties of the parsed regions provide critical
constraints on the identity of the components. Within the temporal
and contextual constraints of primal access, the stages up to and
including the identification of components are assumed to be bottom-
up. A delay in the determination of an object's components should
have a direct effect on the identification latency of the object. The

----
Insert figure 3 about here
----

arrangement of the components is then matched against a representation
in memory. It is assumed that the matching of the components occurs
in parallel, with unlimited capacity. Partial matches are possible
with the degree of match assumed to be proportional to the similarity

Figure 3. Stages in Object Perception
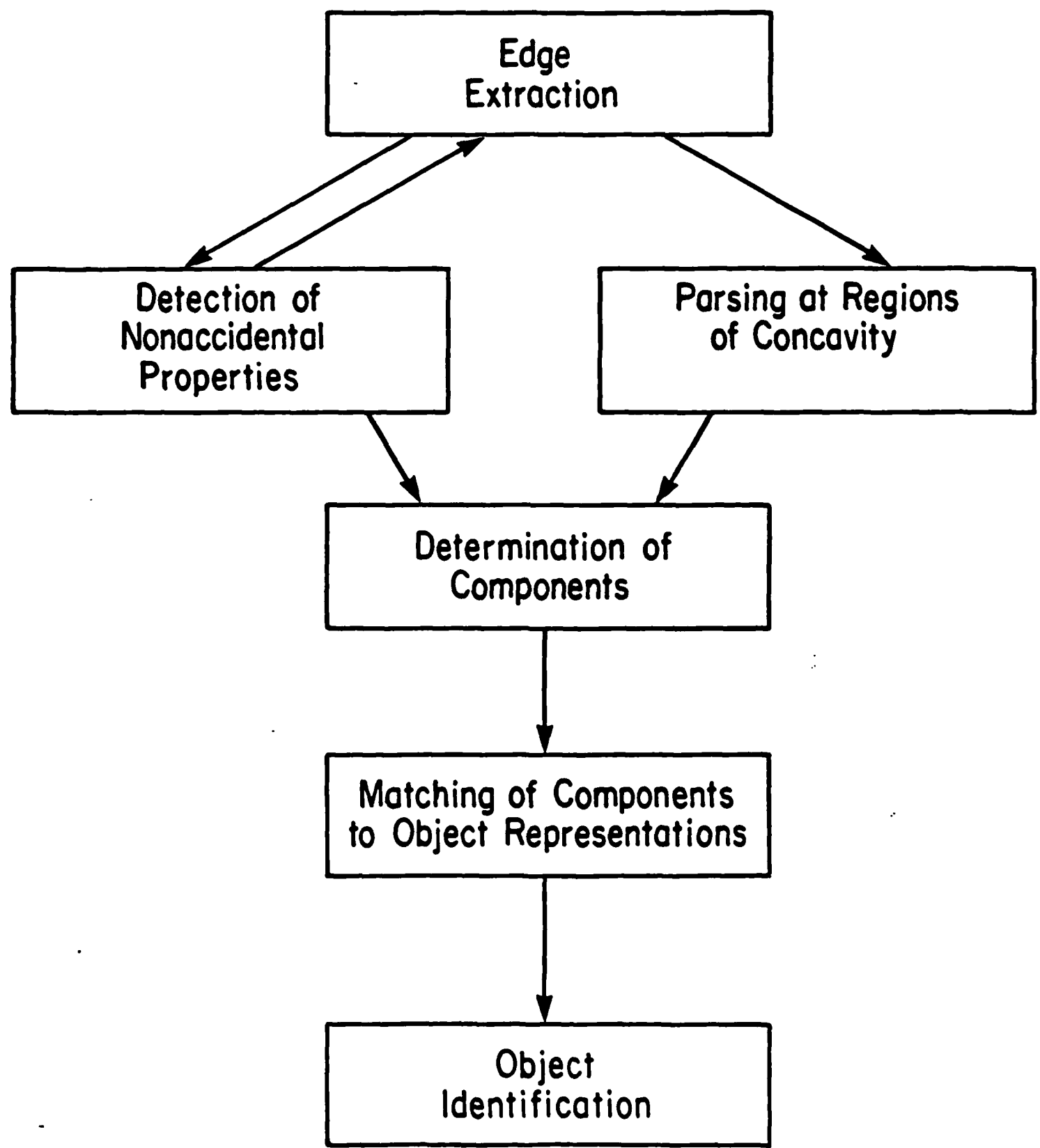
in the components between the image and the representation.[2] This stage model is presented to provide an overall theoretical context. The focus of this article is on the nature of the units of the representation.

A PERCEPTUAL BASIS FOR A COMPONENTIAL REPRESENTATION

Recent theoretical developments concerning perceptual organization (Binford, 1981; Lowe, 1984; Witkin & Tennenbaum, 1983) suggest a perceptual basis for RBC. The central organizational principle is that certain nonaccidental properties of the 2-D image are taken by the visual system as strong evidence that the 3-D object contains those same properties. For example, if there is a straight line in the image, the visual system infers that the edge producing that line in the 3-D world is also straight. The visual system ignores the possibility that the property in the image is merely a result of a (highly unlikely) accidental alignment of eye and a curved edge. Five of these properties and the associated inferences are described in Figure 4 (modified from Lowe, 1984). Witkin & Tanenbaum (see also Lowe, 1984) argue that the evidence for organizational

----
Insert figure 4 about here
----

constraints is so strong and the leverage provided for inferring a 3-D structure so powerful, that it poses a challenge to the effort in computer vision and perceptual psychology that ignored these constraints and assigned central importance to variation in local surface characteristics, such as luminance.

Psychological Evidence for The Rapid Use of Nonaccidental Relations

There is no doubt that images are interpreted in a manner consistent with the nonaccidental principles. But are these relations used quickly enough so as to provide a perceptual basis for the components that allow primal access? Although all the principles have

---

[2]Modeling the matching of an object image to a mental representation is a rich, relatively neglected problem area. Tversky's (1977) contrast model provides a useful framework with which to consider this similarity problem in that it readily allows distinctive features (i.e., components) of the image to be considered separately from the distinctive components of the representation. This allows principled assessments of similarity for partial objects (components in the representation but not in the image) and novel objects (containing components in the image that are not in the representation). It may be possible to construct a dynamic model as a modification of the kind proposed by McClelland & Rumelhart (1981) for word perception, with components playing the role of letters. One difficulty facing such an effort is that the dictionary for words of a given length is well specified and readily available; the set of neighboring objects is not.

Figure 4. Five nonaccidental relations [Adapted from Lowe, 1985.]

Principle of Non-Accidentalness: Critical information is unlikely to be a consequence of an accident of viewpoint.

Three Space Inference from Image Features

2-D Relation	3-D Inference	Examples
1. Collinearity of points or lines	Collinearity in 3-Space	
2. Curvilinearity of points of arcs	Curvilinearity in 3-Space	
3. Symmetry (Skew Symmetry ?)	Symmetry in 3-Space	
4. Parallel Curves (Over Small Visual Angles)	Curves are parallel in 3-Space	
5. Vertices -- two or more terminations at a common point	Curves terminate at a common point in 3-Space	

"L"   "Fork"   "Arrow"

not received experimental test, the available evidence does suggest
that the answer to the preceding question is "yes". There is strong
evidence that the visual system quickly assumes and uses collinearity,
curvature, symmetry and cotermination. This evidence is of two sorts:
(a) Demonstrations, often compelling, showing that when a given 2-D
relation is produced by an accidental alignment of object and image,
the visual system accepts the relation as existing in the 3-D world,
and (b) search tasks showing that when a target differs from
distractors in a nonaccidental property, the detection of that target
is facilitated compared to conditions where targets and background do
not differ in such properties.

Collinearity vs. Curvature. The demonstration of an assumption
of collinearity or curvature is too obvious to be performed as an
experiment. When looking at a straight segment, no observer would
assume that it is an accidental image of a curve. That the contrast
between straight and curved edges is readily available for perception
was shown by Neisser (1963). He found that a search for a letter
composed only of straight segments, such as a Z, could be performed
faster when in a field of curved distractors, such as C, G, O, and Q,
then when among other letters composed of straight segments such as N,
W, V, and M.

Symmetry and Parallelism. Many of the Ames demonstrations, such
as the trapezoidal window and Ames room derive from an assumption of
symmetry that includes parallelism (Ittelson, 1952). Palmer (1983)
demonstrated that the subjective directionality of arrangements of
equilateral triangles was based on the derivation of an axis of
symmetry for the arrangement. King, Tangney, Meyer, & Biederman
(1976) demonstrated that a perceptual bias towards symmetry accounted
for a number of shape constancy effects. Garner (1966), Checkosky &
Whitlock (1973), and Pomerantz (1978) provided ample evidence that not
only can symmetrical shapes be quickly discriminated from asymmetrical
stimuli, but the degree of symmetry was also a readily available
perceptual distinction. Thus stimuli that were invariant under both
reflection and 90° rotation could be rapidly discriminated from those
that were only invariant under reflection (Checkosky & Whitlock,
1973).

Cotermination. The "peephole perception" demonstrations, such as
the Ames chair (Ittelson, 1952) or the physical realization of the
impossible triangle (Penrose & Penrose, 1958), are produced by
accidental alignment of noncoterminous segments. The success of these
demonstrations document the immediate and compelling impact of this
relation.

The registration of cotermination is important for determining
vertices, which provide information that can serve to distinguish the
components. In fact, one theorist (Binford, 1979) has suggested that
the major function of eyemovements is to determine coterminous edges.
With polyhedra (volumes produced by planar surfaces), the Y, Arrow,
and L vertices allow inference as to the identity of the volume in the
image. For example, the silhouette of a brick contains a series of
six vertices, which alternate between Ls and Arrows, and an internal Y

vertex, as illustrated in any of the straight edged cross-sectioned volumes in figure 5. The Y vertex is produced by the cotermination of three segments, with none of the angles greater than 180°. (An arrow vertex contains an angle that exceeds 180°.) This vertex is not present in components that have curved cross sections, such as cylinders, and thus can provide a distinctive cue for the cross-section edge. Perkins (1983) has described a perceptual bias toward parallelism in the interpretation of this vertex.[3] [Chakraverty (1979) has discussed the vertices formed by curved regions.] Whether the presence of this particular internal vertex can affect primal access is not yet known but a recent study by Biederman & Blickle (1985, described below) demonstrated that deletion of vertices adversely affected object recognition.

An example of a non-coterminous vertex is the T. Such vertices are important for determining occlusion and thus segmentation (along with concavities), in that the edge forming the (normally) vertical segment of the T cannot be closer to the viewer than the (normally) top of the T. By this account, the T vertex should have a somewhat different status than the other three, the Y, Arrow, and L, in that the T's primary role would be in segmentation, rather than in establishing the identity of the volume.[4]

The high speed and accuracy of determining a given nonaccidental relation, e.g., whether some pattern is symmetrical, should be

---

[3]When such vertices formed the central angle in a polyhedron, Perkins (1983) reported that the surfaces would almost always be interpreted as meeting at right angles, as long as none of the three angles was less than 90°. Indeed, such vertices cannot be projections of acute angles (Kanade, 1981) but the human appears insensitive to the possibility that the vertices could have arisen from obtuse angles. If one of the angles in the central Y vertex was acute, then the polyhedra would be interpreted as irregular. Perkins found that subjects from rural areas of Botswana, where there was a lower incidence of exposure to carpentered (right-angled) environments, had an even stronger bias toward rectilinear interpretations than Westerners (Perkins & Deregowski, 1983).

[4]The arrangement of vertices, particularly for polyhedra, offers constraints on "possible" interpretations of lines as convex, concave, or occluding, e.g., Sugihara, 1984. In general, the constraints take the form that a segment cannot change its interpretation, e.g., from concave to convex, unless it passes through a vertex. "Impossible" objects can be constructed from violations of this constraint (Waltz, 1975) as well as from more general considerations (Sugihara, 1982; 1984). It is tempting to consider that the visual system captures these constraints in the way in which edges are grouped into objects, but the evidence would seem to argue against such an interpretation. The impossibility of most impossible is not immediately registered, but requires scrutiny and thought before the inconsistency is detected. What this means in the present context is that the visual system has a capacity to classify vertices locally, but no perceptual routines for determining the global consistency of a set of vertices.

contrasted with performance in making absolute judgments of variations
in a single, physical attribute, such as length or degree of tilt or
curvature.   Such judgments are notoriously slow and error prone
(Miller, 1956; Garner, 1962; Beck, et al. 1983; Virsu, 1971a,b; Fildes
& Triggs, 1985). Even these modest performance levels are challenged
when the judgments have to be executed over the brief 100 msec
intervals (Egeth & Pachella, 1969) that is sufficient for accurate
object identification.   Perhaps even more telling against a view of
object recognition that would postulate the making of absolute
judgments of fine quantitative detail is that the speed and accuracy
of such judgments decline dramatically when they have to be made for
multiple attributes (Miller, 1956; Garner, 1962; Egeth & Pachella,
1969).   In contrast, object recognition latencies are facilitated by
the opportunity for additional (redundant) components with complex
objects (Biederman, Clapper, & Ju, 1985, described below).

COMPONENTS GENERATED FROM DIFFERENCES IN NONACCIDENTAL PROPERTIES AMONG GENERALIZED CONES

     I have emphasized the particular set of nonaccidental properties
shown in Figure 4 because they may constitute a perceptual basis for
the generation of the set of components. Any primitive that is
hypothesized to be the basis of object recognition should be rapidly
identifiable and invariant over viewpoint and noise. These
characteristics would be attainable if differences among components
were based on differences in nonaccidental properties. Although
additional nonaccidental properties exist, there is empirical support
for rapid perceptual access to the five described in figure 4. In
addition, these five relations reflect intuitions about significant
perceptual and cognitive differences among volumes.

     From variation over only two or three levels in the nonaccidental
relations of four attributes of generalized cylinders, a set of 36
components can be generated. A subset is illustrated in figure 5.

                    Insert Figure 5 About Here

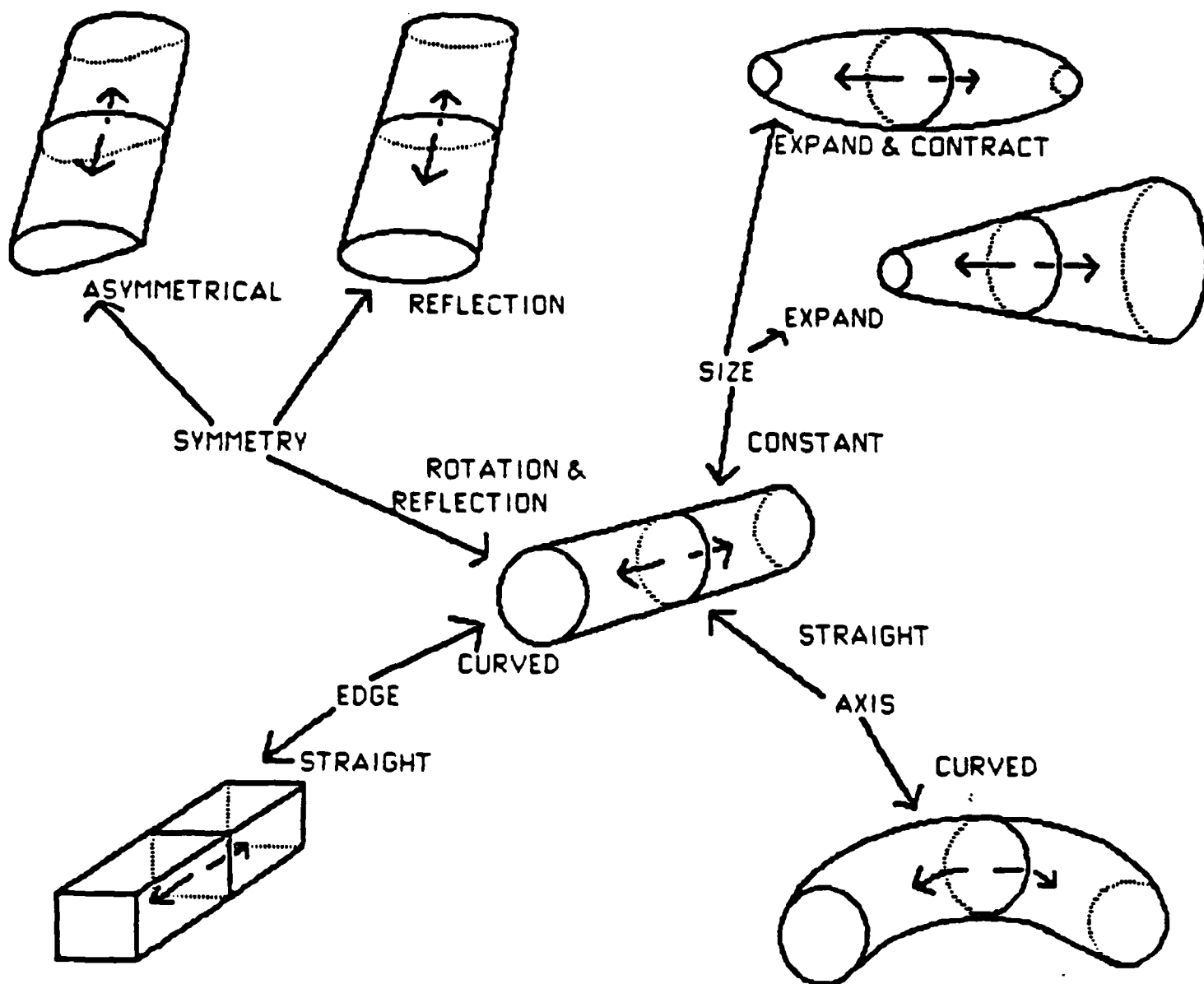Some of the generated volumes and their organization are shown in

Figure 5. Variations in generalized cones that can be detected
through nonaccidental properties. Constant-sized cross sections have
parallel sides; expanded or expanded & contracted cross sections have
sides that are not parallel. Curved vs Straight cross sections and
axes are detectable through Collinearity or Curvature. The three
values of cross-section Symmetry (Symmetrical under Reflection & 90°
Rotation; Reflection only; or Asymmetrical) are detectable through the
symmetry relation.

Figure 6.   Three of the attributes describe characteristics of the
cross section; its shape, symmetry, and constancy of size as it is
swept along the axis. The fourth attribute describes the shape of the
axis.

1.   Cross Section
     A.   Edges
          O   Straight
          O   Curved
     B.   Symmetry
          ++ Symmetrical: Invariant under Rotation & Reflection
          +   Symmetrical: Invariant under Reflection
          -   Asymmetrical
     C.   Constancy of size of cross section as it is swept along axis
          +   Constant
          -   Expanded
          -- Expanded and Contracted
2.   Axis
     D.   Curvature
          +   Straight
          -   Curved

     The values of these four attributes are presented as contrastive

----
Insert Figure 6 About Here
----

differences in   nonaccidental properties: straight vs.   curved,
symmetrical vs asymmetrical, parallel vs nonparallel. Cross section
edges and curvature of the axis are distinguishable by collinearity or
curvilinearity.   The constant vs expanded size of the cross section
would be detectable through parallelism; a constant cross section
would produce a generalized cone with parallel sides (as with a
cylinder or brick); an expanded cross section would produce edges that
were not parallel (as with a cone or wedge), and a cross section that
expanded and   then contracted   would produce an ellipsoid with
nonparallel sides and an extrema of positive curvature (as with a
lemon).   As Hoffman & Richards (1985) have noted, such extrema are
invariant with viewpoint. The three levels of cross-section symmetry
are equivalent to Garner's (1966) distinction of the number of
different stimuli produced by 90° rotations and reflections of a
stimulus.   Thus a square or circle would be invariant under 90°
rotation and reflection; but a rectangle or ellipse would be invariant
only under reflection, as 90° rotations would produce a second figure.
Asymmetrical figures would produce eight different figures under 90°
rotation and reflection.

Negative Values
     The plus values are those favored by perceptual biases and memory
errors. No bias is assumed for straight and curved edges of the cross
section.   For symmetry, clear biases have been documented.   For
example, if an image could have arisen from a symmetrical object, then
it is interpreted as asymmetrical (King, et al., 1976). The same is

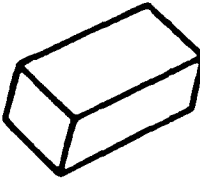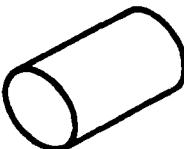# Partial Tentative Geon Set Based on Nonaccidentalness Relations

CROSS SECTION
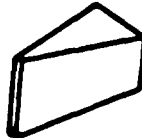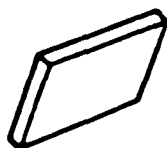
Geon	Edge Straight Curved	Symmetry Rot & Ref ++ Ref + Asymm −	Size Constant ++ Expanded − Exp & Cont −−	Axis Straight + Curved −
		++	++	+
		++	++	+
		+	−	+
		++	+	−
		++	−	+
		+	+	+

Figure 6.   Proposed partial set of volumetric primitives (Geons)
derived from differences in nonaccidental properties.

apparently true of parallelism. If edges could be parallel, then they are typically interpreted as such, as with the trapezoidal room or window.

Curved axes. Figure 7 shows three of the most negatively marked primitives with curved crossed sections. Such volumes often resemble biological entities. An expansion and contraction of a rounded cross section with a straight axis produces an ellipsoid (lemon) (figure 7a); an expanded cross section with a curved axis produces a horn (figure 7b), and an expanded and contracted cross section with a rounded cross section produces a banana slug or gourd (figure 7c).

Insert Figure 7 About Here
----

In contrast to the natural forms generated when both cross section and axis are curved, a smoothly curved axis with a straight edged cross section appears unfamiliar, as illustrated for the components on the first, third, and fifth rows of figure 8. Given the presence in the image of curves and straight edges, attention may be required to determine, for some combinations of volumes, which kind of edge is a property of the axis and which is a property of the cross section (Treisman & Gelade, 1980). In the present case there also appears to be a bias in that the presence of the straight edged cross section is not immediately apparent when the axis is curved although curved cross sections are readily identifiable when run along straight axis (to produce a cylinder or cone). Fortunately, this issue as to the role of attention in identifying volumes is empirically tractable using the paradigms created by Treisman and her colleagues (Treisman & Gelade, 1980; Treisman, 1982; Treisman & Schmidt, 1983).

Insert Figure 8 About Here

Asymmetrical cross sections. There are an infinity of possible cross sections that could be asymmetrical. How does RBC represent this variation? RBC assumes that the differences in the departures from symmetry are not readily available and thus do not affect primal access. For example, the difference in the shape of the cross section for the two straight edged volumes in Fig. 9 might not be apparent sufficiently quickly to affect object recognition. This does not mean that an individual could not store the details of the volume produced by an asymmetrical cross section. But if such detail required

----
Insert Figure 9 About Here
----

additional time for its access (and for its storage as well), then the expectation is that it could not mediate rapid object perception. A second way in which asymmetrical cross sections need not be individually represented is that they often produce volumes that

Cross Section:
    Edge: Curved ( )
    Symmetry: Yes (+)
    Size: Expanded & Contracted: (--)
Axis: Straight (+)
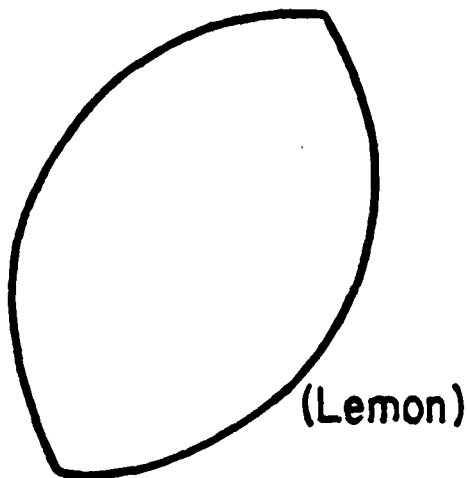
(Lemon)

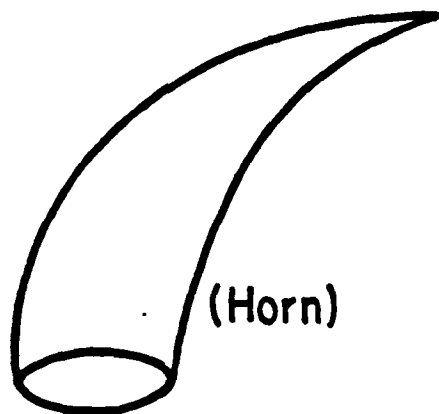**Cross Section:**
    Edge: Curved ( )
    Symmetry: Yes (+)
    Size: Expanded (−)
Axis: Curved (−)

(Horn)
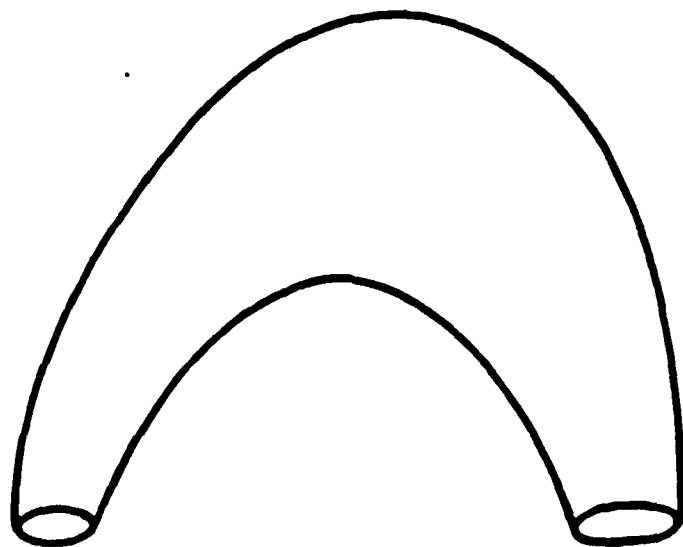
Cross Section:
    Edge: Curved ( )
    Symmetry: Yes (+)
    Size: Expanded & Contracted (--)
Axis: Curved (−)

(Gourd)

Figure 7. Three curved components with curved axes or expanded and/or contracted cross sections. These tend to resemble biological forms.
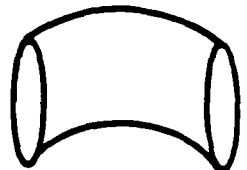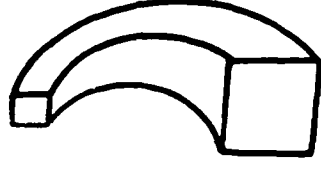
Figure 12 b

CROSS SECTION

Geon	Edge Straight Curved	Symmetry Rot & Ref ++ Ref + Asymm −	Size Constant ++ Expanded − Exp & Cont −−	Axis Straight + Curved −
		+	++	−
		+	++	−
		++	−	−
		++	−	−
		+	−	−
		+	−	−

Figure 8. Components with curved axis and straight or curved cross sections. Determining the shape of the cross section, particularly if straght, might require attention.
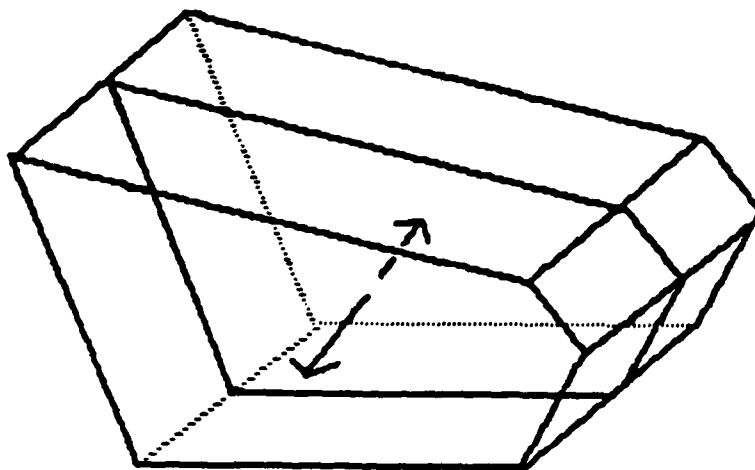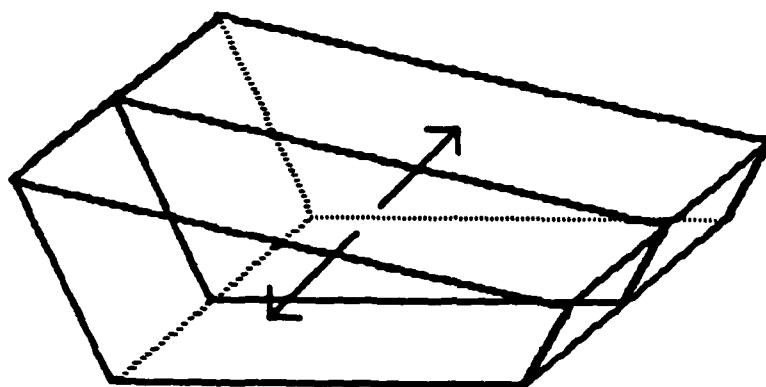
Figure 9. Volumes generated with different asymmetrical, straight edged cross sections. Detection of differences between such volumes might require attention.

resemble symmetrical, but truncated, wedges. This latter form of representing asymmetrical cross sections would be analogous to the schema-plus-correction phenomenon noted by Bartlett (1938). The implication of a schema-plus-correction representation would be that a single primitive category for asymmetrical cross sections and wedges might be sufficient. For both kinds of volumes, their similarity may be a function of the detection of a lack of parallelism in the volume. One would have to exert scrutiny to determine if a lack of parallelism had originated in the cross section or in a size change of a symmetrical cross section. In this case, as with the components with curved axes described in the preceding section, a single primitive category for both wedges and asymmetrical straight edged volumes could be postulated that would allow a reduction in the number of primitive components.

Attentional effects. The extent to which Treisman and Gelade's (1980) demonstration of conjunction-attention effects may be applicable to the perception of volumes and objects has yet to be evaluated. In the extreme, in a given moment of attention, it may be the case that the values of the four attributes of the components are detected as independent features. In cases where the attributes, taken independently, can define different volumes, an act of attention might be required to determine the specific component generating those attributes. At the other extreme, it may be that an object recognition system has evolved to allow automatic determination of the components.

The more general issue is whether relational structures for the primitive components are defined automatically or whether a limited attentional capacity is required to build them from their individual edge attributes. It could be the case that some of the most positively marked volumes are detected automatically, but that the negative volumes might require attention. That some limited capacity is involved in the perception of objects (but not necessarily their components) is documented by an effect of the number of irrelevant objects on perceptual search (Biederman, 1981). Reaction times and errors for detecting an object, e.g., a chair, increased linearly as a function of the number of nontarget objects in a 100 msec presentation of a clockface display (Biederman, 1981). Whether this effect arises from the necessity to use a limited capacity to construct a component from its attributes or whether the effect arises from the construction of the object from the components or both remains to be investigated.

Variation in aspect ratio. If the perceptual input is organized into components for recognition, then one problem is how to conceptualize quantitative variations in the dimensions of a given component type, either because of differences in viewpoint of the component or for different instances of the component type. One way to consider such variation is in terms of each component's aspect ratio, the width-to-height ratio of a bounding rectangle that would just enclose the component. [It is somewhat unclear as to how to handle components with curved axis. The bounding rectangle could simply enclose the component, whatever its shape. Alternatively, two

rectangles could be constructed.]   Aspect ratios are not invariant with viewpoint.

One possibility is to include specification of a range of aspect ratios in the structural description of the object.   It seems plausible to assume that recognition can be indexed, in part, by aspect ratio in addition to a componential description. An object's aspect ratio would thus play a role similar to that played by word length in the tachistoscopic identification of words, where long words are rarely proffered when a short word is flashed.   Consider an elongated object, such as a baseball bat with a (real) aspect ratio of 15:1.   When the orientation of the object is orthogonal to the viewpoint, so that its aspect ratio is 15:1, recognition might be faster then when it is shown at an orientation where its length is only slightly larger than its diameter, so that the aspect ratio of its image is only 2:1.   One need not have a particularly fine tuned function for aspect ratio as large differences in aspect ratio between two components would, like parallelism, be preserved over a large proportion of arbitrary viewing angles.

Another way to incorporate variations in the aspect ratio of an object's image is to represent only qualitative differences, so that variations in aspect ratios exert an effect only when the relative size of the longest dimensions undergo reversal. Specifically, for each component and the complete object, three variations could be defined depending on whether the axis was much smaller, approximately equal to, or much longer than the longest dimension of the cross section.   For example, in a component whose axis was longer than the diameter of the cross section (which would be true in most cases), only when the projection of the cross section became longer than the axis would there be an effect of the object's orientation, as when the bat was viewed almost from on end so that the diameter of the handle was greater than the projection of its length.

A close dependence of object recognition performance on the preservation of the aspect ratio of a component in the image would be inconsistent with the emphasis by RBC on dichotomous contrasts of nonaccidental relations.   Fortunately, these issues on the role of aspect ratio are readily testable.   Bartram's (1976) experiments, described in the section on Orientation Variability, suggest that sensitivity to variations in aspect ratio need not be given heavy weight: Recognition speed is unaffected by variation in aspect ratio across different views of the same object.

Planar Components.   A special case of aspect ratio needs to be considered: When the axis for a constant cross section is much smaller than the greatest extent of the cross section, a component may lose its volumetric character and appear planar, as the flipper of the penguin in fig. 13, or the eye of the elephant in figure 12. Such shapes can be conceptualized in two ways.   The first (and less favored) is to assume that these are just quantitative variations of the volumetric components, but with an axis length of zero.   They would then have default values of a straight axis (+) and a constant

cross-section (+).   Only the edge of the cross section and its
symmetry could vary.

        Alternatively,   it might   be that   a flat   shape is not related
perceptually to the foreshortened projection of the volume that could
have produced it. Using the same variation in cross-section edge and
symmetry as with the volumetric components, seven <u>planar</u> components
could be defined. For ++symmetry there would be the square and circle
(with straight and curved edges, respectively), for +symmetry the
rectangle, triangle, and ellipse.   Asymmetrical(-) planar components
would include trapezoids (straight edges), and drop shapes (curved
edges).   The addition of these seven planar components to the 36
volumetric components yields 43 components (a number close to the 55
phonemes required to represent all languages). [The triangle is here
assumed to define a separate component, although a triangular cross
section was not assumed to define a separate volume under the
intuition that a prism (produced by a triangular cross section) is not
quickly distinguishable from wedges).]   My preference for assuming
that planar   components are   not perceptually related to their
foreshortened volumes is based on the extraordinary difficulty of
recognizing objects from views that are parallel to the axis of the
major components, as shown in figures 26 and 27.

        <u>Selection of axis</u>.   Given that a volume is segmented from the
object, how is an axis selected? Subjectively, it appears that an
axis is selected that would maximize its length, the symmetry of the
cross section, and the constancy of the size of the cross section. It
may be that by having the axis correspond to the longest extent of the
component, bilateral symmetry can be more readily detected as the
sides would be closer.   Typically, a single axis satisfies all three
criteria, but sometimes these criteria are in opposition and two (or
more) axes (and component types) are plausible (Brady, 1983). Under
these conditions, axis will often be aligned to an external frame,
such as the vertical (Humphreys, 1983).

RELATIONS OF RBC TO PRINCIPLES OF PERCEPTUAL ORGANIZATION

        Textbook presentations of perception typically include a section
of Gestalt organizational principles.   This section is almost never
linked to any other function of perception.   RBC posits a specific
role for these organizational phenomena in pattern recognition.
Specifically, as suggested by the section on generating components
through nonaccidental properties, the Gestalt principles (or better,
nonaccidental principles)   serve   to   determine   the individual
components, rather than the complete object. A complete object, such
as a chair, can be highly complex and asymmetrical, but the components
will be simple volumes. A consequence of this interpretation is that
it is the components that will be stable under noise or perturbation.
If the components can be recovered and object perception is based on
the components, then the object will be recognizable.

        This may be the reason why it is difficult to camouflage objects
by moderate doses of random occluding noise, as when a car is viewed
behind foliage.   According to RBC, the components accessing the

representation of an object can readily be recovered through routines of collinearity or curvature that restore contours (Lowe, 1984). These mechanisms for contour restoration will not bridge cusps. For visual noise to be effective, by these considerations, it must obliterate the concavity and interrupt the contours from one component at the precise point where they can be joined, through collinearity or constant curvature, with the contours of another component. The liklihood of this occurring by moderate random noise is, of course, extraordinarily low and it is a major reason why, according to RBC, objects are rarely rendered unidentifiable by noise. Experiments subjecting these conjectures to test are described in a later section.

A LIMITED NUMBER OF COMPONENTS?

The motivation behind the conjecture that there may be a limit to the number of primitive component derives from both empirical and computational considerations, in addition to the limited number of components that can be discriminated from differences in nonaccidental properties among generalized cones. Empirically, there is evidence documenting severe limitations of the capacity for making absolute and rapid judgments of shapes and the nature of errors in memory for shapes. Computationally, a limit is suggested by estimates of the number of objects we might know and the capacity for RBC to readily represent a far greater number with a limited number of primitives.

Empirical support for a limit. Although the visual system is capable of representing extremely fine detail, I have been assuming that the number of volumetric primitives sufficient to model rapid human object recognition may be limited. It should be noted that the number of proposed primitives is greater than the three--cylinder, sphere, and cone--advocated by many how-to-draw books. Although these three may be sufficient for determining relative proportions of the parts of a figure and can furnish aid for perspective, they are not sufficient for the rapid identification of objects.[5] Similarly, Marr & Nishihara's (1978) pipe-cleaner (viz., cylinder) representations of animals (1978) would also appear to posit an insufficient number of primitives. On the page, in the context of other labeled pipe-cleaner animals, it is certainly possible to arrive at an identification of a particular (labeled) animal, e.g., a Giraffe. But the thesis proposed here would hold that the identifications of objects that were distinguished only by the aspect ratios of a single component type, would require more time than if the representation of the object preserved its componential identity. In modeling only animals, it is likely that Marr & Nishihara capitalized on the possibility that appendages, e.g., legs and neck, can often be modeled by the cylindrical forms of a pipe cleaner. By contrast, it is unlikely that

---

[5]Paul Cezanne is often incorrectly cited on this point. "Treat nature by the cylinder, the sphere, the cone, everything in proper perspective so that each side of an object or plane is directed towards a central point." (Italics mine, Cezanne, 1904/1941.) Cezanne was referring to perspective, not the veridical representation of objects.

a pipe-cleaner representation of a desk would have had any success. The lesson from Marr & Nishihara's demonstration, even limited for animals, may well be that a single component, varying only in aspect ratio (and arrangement with other components), is insufficient for primal access.

As noted earlier, one reason not to posit a representation system based on fine quantitative detail, e.g., many variations in degree of curvature, is that such absolute judgments are notoriously slow and error prone unless limited to the 7±2 values argued by Miller (1956). Even this modest limit is challenged when the judgments have to be executed over a brief 100 msec interval (Egeth & Pachella, 1969) that is sufficient for accurate object identification. A further reduction in the capacity for absolute judgments of quantitative variations of a simple shape would derive from the necessity, for most objects, to make simultaneous absolute judgments for the several shapes that constitute the object's parts (Miller, 1956; Egeth & Pachella, 1969). This limitation on our capacities for making absolute judgments of physical variation, when combined with the dependence of such variation on orientation and noise, makes quantitative shape judgments a most implausible basis for object recognition. RBC's alternative is that the perceptual discriminations required to determine the primitive components can be made qualitatively, requiring the discrimination from only two or three viewpoint-independent levels of variation[6].

Our memory for irregular shapes shows clear biases toward "regularization" (e.g., Woodworth, 1938). Amply documented in the classical shape memory literature was the tendency for errors in the reproduction and recognition of irregular shapes to be in a direction of "regularization," in which slight deviations from symmetrical or regular figures were omitted in attempts at reproduction. Alternatively, some irregularities were emphasized ("accentuation"), typically by the addition of a regular subpart. What is the significance of these memory biases? By the RBC hypothesis, these errors may have their origin in the mapping of the perceptual input into a representational system based on regular primitives. The memory of a slight irregular form would be coded as the closest regularized neighbor of that form. If the irregularity was to be represented as well, an act that would presumably require additional time and capacity, then an additional code (sometimes a component) would be added. The latter was referred to as "Schema with Correction" (Bartlett, 1932).

Computational Considerations

[5]Paul Cezanne is often incorrectly cited on this point. "Treat nature by the cylinder, the sphere, the cone, everything in proper perspective so that each side of an object or plane is directed towards a central point." (Italics mine, Cezanne, 1904/1941.) Cezanne was referring to perspective, not the veridical representation of objects.

Are 36 Components sufficient?   Is there sufficient coding
capacity in a set of 36 components to represent the basic level
categorizations that we can make? Two estimates are needed to provide
a response to this question: (a) the number of readily available
perceptual categories, and (b) the number of possible objects that
could be represented by 36 components. Obviously, the value for (b)
would have to be greater than the value for (a) if 36 components are
to prove sufficient.

How many readily distinguishable objects do people know? How
might one arrive at a liberal estimate for this value? One estimate
can be obtained from the lexicon.   There are approximately 1,000
relatively common basic level object categories, such as chairs and
elephants.[7]   Assume that this estimate is too small by a factor of
three, so we can discriminate approximately 3,000 basic level
categories. As is discussed below, RBC holds that perception is based
on the particular, subordinate level object rather than the basic
level category so we need to estimate the number of instances, within
a basic level category, that would have different structural
descriptions.   Almost all natural categories, appear to have one or
only a few instances, such as elephants or giraffes, in that we know
of few (one?) componential description(s).   Only a few natural
categories, such as dogs, have considerable variation in their
descriptions.   Person-made categories vary in the number of allowable
types, but this number often tends to be greater than the natural
categories.   Cups, typewriters and lamps have just a few (in the case
of cups) to perhaps 15 or more (in the case of lamps) readily
discernible exemplars.   Let's assume (liberally) that the mean number
of types is 10.   This would yield an estimate of 30,000 readily
discriminable objects   (3,000 categories X 10 types/category).   The
second source for the estimate is the rate of learning new objects.

---

[6]This limitation on our capacities for absolute judgments also occurs
in the auditory domain (Miller, 1956).   It is possible that the
limited number of phonemes derives more from this limitation for
accessing memory for fine quantitative variation than it does from
limits on the fineness of the commands to the speech musculature.

[7]This estimate was obtained from three sources: (a)Several linguists
and cognitive psychologists provided guesses of 300 to 1,000 concrete
noun object categories.   (b) The six year old child can name most of
the objects that he or she sees on television and has a vocabulary
that is under 10,000 words.   Perhaps ten percent, at most, are
concrete nouns.   (c) Perhaps the most defensible estimate was obtained
from a sample of Webster's Seventh New Collegiate Dictionary. The
author sampled 30 pages and counted the number of readily
identifiable, unique concrete nouns that would not be subordinate
subsumed under another nouns.   Thus "Wood thrush" was not counted
because it could not be readily discriminated from a "Sparrow".
"Penguin" and "Ostrich" and any doubtful entries were counted as
separate noun categories. The mean number of nouns per page was 1.4,
with a 1,200 word dictionary this is equivalent to 1,600 noun
categories.

Thirty thousand objects would require learning an average of 4.5 objects per day, every day for 18 years, the median age of the subjects in these experiments.

Although the value of 4.5 objects per day approximates the maximum rates of word acquisition during the ages of 2-6 years (Carey, 1976; Templin, 1957; Miller, 1977), it certainly overestimates the number of new objects learned by adults. In fact, a child of six shows enormous visual competence, easily understanding the basic level categories of almost everything that appears on television. If the six year old child knew 30,000 visual categories, then that number would require learning 13.5 objects per day, or about one per waking hour.

How many objects could be represented by 36 components? calculations of this estimate are presented in Table 1. If we consider the number of possible objects that could be represented by just two components, with a conservative estimate of the number of

----

Insert Table 1 About Here

----

readily discriminably different ways in which those components might combine, then 55,987 objects can be generated. Five relations among pairs of components are considered: a) whether Component A is above or below Component B, a relation, by the author's estimate, that is defined for at least 80% of the objects. Thus giraffes, chairs, and typewriters have a top-down specified organization of their components but forks and knives do not. b) whether the connection between any pair of joined components is end-to-end or end-to-side, producing one or two concavities, respectively (Marr & Nishihara, 1978); c) whether Component A is much greater than, smaller than, or approximately equal to Component B; d) whether each component is connected at its longer or shorter side. The difference between the attache case in figure 3a and the strongbox in figure 3b are produced by differences in relative lengths of the surfaces of a brick that is connected to the arch (handle). The handle on the shortest surface produces the strongbox; on a longer surface, the attache case. Similarly, among other differences, the cup and the pail in figures 3c and 3d, respectively, differ as to whether the handle is connected to the long surface of the cylinder (to produce a cup) or the short surface (to produce a pail). [Other than a sphere and a cube, all primitives will have at least a long and a short surface, ignoring the orientation of the surface. Other than a brick and a cylinder, which have two, most, such as a wedge, will have at least five distinguishably different surfaces, if we ignore left-right differences. That is, there will be a front, back, top, bottom, and side. Now, a second volume can be joined to the first at its top, bottom, front, back, or side. There are four degrees of freedom if the second volume is joined to the bottom of the first, then it cannot be joined at its top.

TABLE 1

GENERATIVE POWER OF 36 COMPONENTS

36   First Component, C_1

X

36   Second Component, C_2

X

3   Size [$C_1 \gg C_2$, $C_2 \gg C_1$, $C_1 = C_2$]

X

1.8 C_1 top or bottom (represented for 80% of the objects)

X

2   Nature of Join [End-to-End or End-to-Side]

X

2   Join at long or short surface of C_1

X

2   Join at long or short surface of C_2

= 55,987 possible two Component objects

With 3 Components, ignoring relations:

55,987 X 36 = 2 million possible objects.

Equivalent to learning 304 new objects every day (approx. 20/waking
hour) for 18 years.

Consequently, there are 20 possible combinations (joins) made between two five surfaced primitives. The tabled estimate considers only two levels of this variation.]

If a third Component is added to the two, then 2 million possible objects can be represented, even if we completely ignore the relations among this third volume and the other two! This would be equivalent to learning 304 objects/day every day for 18 years or 20 objects per hour of the 16 waking hours of every day for those 18 years.

The representational capacity is, of course, a multiplicative function of the number of primitives or relations. Slight increases in either have a dramatic effect on the representational capacity. For example, with 50 components, a value close to the number of phonemes, there are 108,000 two-component objects and 5.4 million possible three-component objects, again ignoring the relation between the third component and the other two. This would be equivalent to learning 960 objects/day every day for 18 years or an object a minute of the 16 waking hours of every day for those 18 years.[8]

How many components would be required for the unambiguous identification of most objects? If only one percent of the possible combinations of components were actually used (i.e., 99% redundancy), and objects were distributed homogeneously among combinations of components, then only two or three components would be sufficient to unambiguously represent most objects.

We do not yet know if there is a real limit to the number of components but the task to determine if one exists may ultimately prove similar to the task faced by the phonetician as he or she attempts to determine the set of phonemes that characterizes the linguistic corpus for a given community. The phonemes required to represent a large sample of words from the corpus are noted. At some point, an asymptote is reached and additional words can be represented according to the already existing phoneme set. The issues in vision are: (a) whether an asymptote will be reached as observers generate components from a large corpus of objects (Figure 10), (b) whether there would be a strong consensus as to the members of this set, and

----
Insert Figure 10 About Here

---

[8]Fifty primitives does seem like a considerable number, given most psychological theories. But it would be approximately equivalent to the number of phonemes and well within the capacity of current recent chip technology. A recently announced VLSI chip (Rosenthal, 1984), the PF474, can perform several thousand string comparisons per second with a ranked list of the 16 best matches that might have the potential to code tests for the discrimination among the components and perform the matching of component arrangements for object perception. (Each component can be represented by a single string.) It has already been applied in speech perception systems.
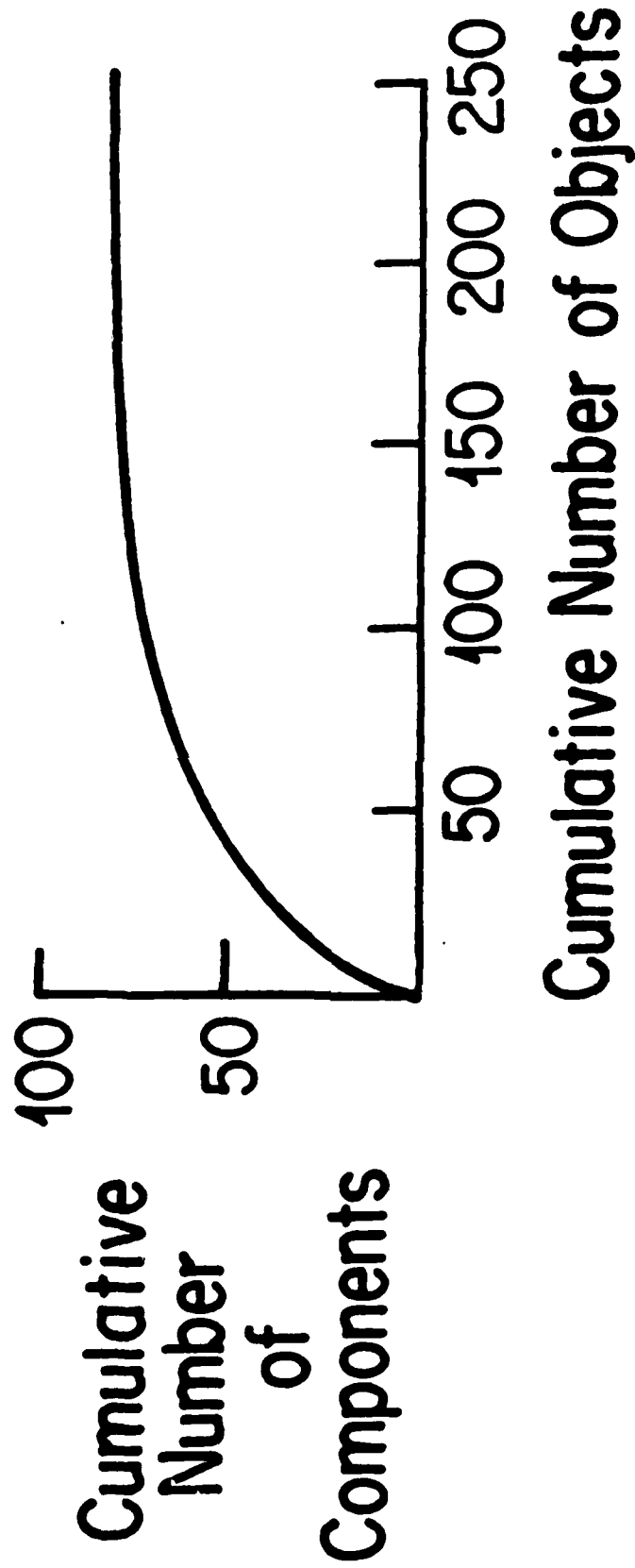
Figure 10. Predicted asymptote in the number of generated components from the segmentation of a large number of objects.

(c) whether objects generated from these components would be identified as readily as their natural counterparts. A limit to the number of components would imply categorical effects such that variations in the contours of an object that did not alter a component's identity would have less of an effect on the identification of the object itself, compared to contour variations that did alter a component's identity.

EXPERIMENTAL SUPPORT FOR A COMPONENTIAL REPRESENTATION

According to the RBC hypothesis, the preferred input for accessing object recognition is that of the volumetric components. In most cases, only a few appropriately arranged volumes would be all that is required to uniquely specify an object. Rapid object recognition should then be possible. Neither the full complement of an object's components, nor its texture, nor its color, nor the full bounding contour (or envelope or outline) of the object need be present for rapid identification. The task of recognizing tens of thousands of possible objects becomes, in each case, just a simple task of identifying a few components, from a limited set, in a particular arrangement.

Overview of Experiments

Several object naming reaction time experiments have provided support for various aspect of the RBC hypothesis. In all experiments, subjects named briefly presented pictures of common objects. That RBC may provide a sufficient account of object recognition was supported by experiments indicating that objects drawn with only two or three of their components, could be accurately identified from a single, 100 msec exposure. When shown with a complete complement of components, these simple line drawings were identified almost as rapidly as full colored, textured slides of the same objects. That RBC may provide a necessary account of object recognition was supported by a demonstration that degradation (contour deletion), if applied at the regions that are critical according to RBC, rendered an object unidentifiable. All the original experimental results reported here have received at least one, and often several, replications.

PERCEIVING INCOMPLETE OBJECTS

Biederman, Ju, & Clapper (1985) studied the perception of briefly presented partial objects lacking some of their components. A prediction of RBC was that only two or three components would be sufficient for rapid identification of most objects. If there was enough time to determine the components and their relations, then object identification should be possible. Complete objects would be maximally similar to their representation and should enjoy an identification speed advantage over their partial versions.

Stimuli. The experimental objects were line drawings of 36 common objects, half of which are illustrated in figures 11 and 12.

The depiction of the objects and their partition into components was done subjectively, according to generally easy agreement among at least three judges. The artists were unaware of the set of components described in this article. For the most part, the components corresponded to the parts of the object. Seventeen component types (ignoring aspect ratios), were sufficient to represent the 180 components comprising the complete versions of the 36 objects.

Insert Figures 11 and 12 About Here

The objects were shown either with their full complement of components, or partially, but never with less than two components. The first two components that were selected were the largest and most diagnostic components from the complete object and additional components were added in decreasing order of size or diagnosticity, as illustrated in figures 13 and 14. Additional components were added in decreasing order of size and/or diagnosticity, subject to the constraint that the additional component be connected to the existing components. For example, the airplane which required nine components to look complete, would have the fuselage and two wings when shown with three of the nine components. The objects were displayed in black line on a white background and averaged 4.5° in greatest extent.

Insert Figures 13 and 14 About Here

The purpose of this experiment was to determine whether the first few components that would be available from an unoccluded view of a complete object would be sufficient for rapid identification of the object. In normal viewing, the largest and most diagnostic components are available for perception. We ordered the components by size and diagnosticity because our interest, as just noted, was on primal access in recognizing a complete object. Assuming that the largest and most diagnostic components would control this access, we studied the contribution of the nth largest and most diagnostic component, when added to the n-1 already existing components, because this would more closely mimic the contribution of that component when looking at the complete object. (Another kind of experiment might explore the contribution of an "average" component by balancing the order of addition of the components. Such an experiment would be relevant to the recognition of an object that was occluded in such a way that only the displayed components would be available for viewing.)

Complexity.--The objects shown in figures 11 and 12 illustrate the second major variable in the experiment. Objects differ in complexity; by RBC's definition, in the number of components that they require to look complete. As noted previously, it would seem plausible that partial objects would require more time for their identification than complete objects, so that a complete airplane of nine components, for example, might be more rapidly recognized than only a partial version of that airplane, with only three of its
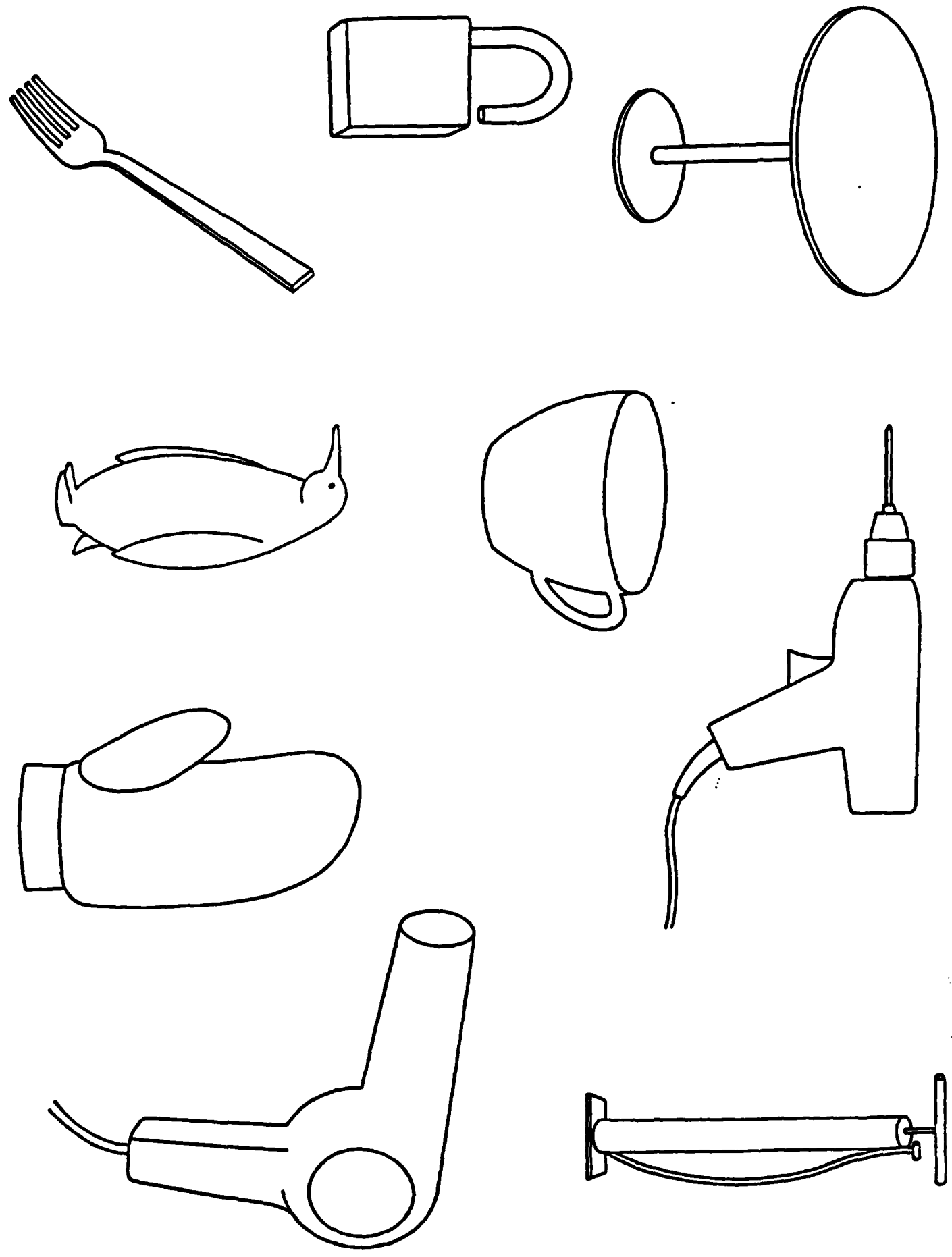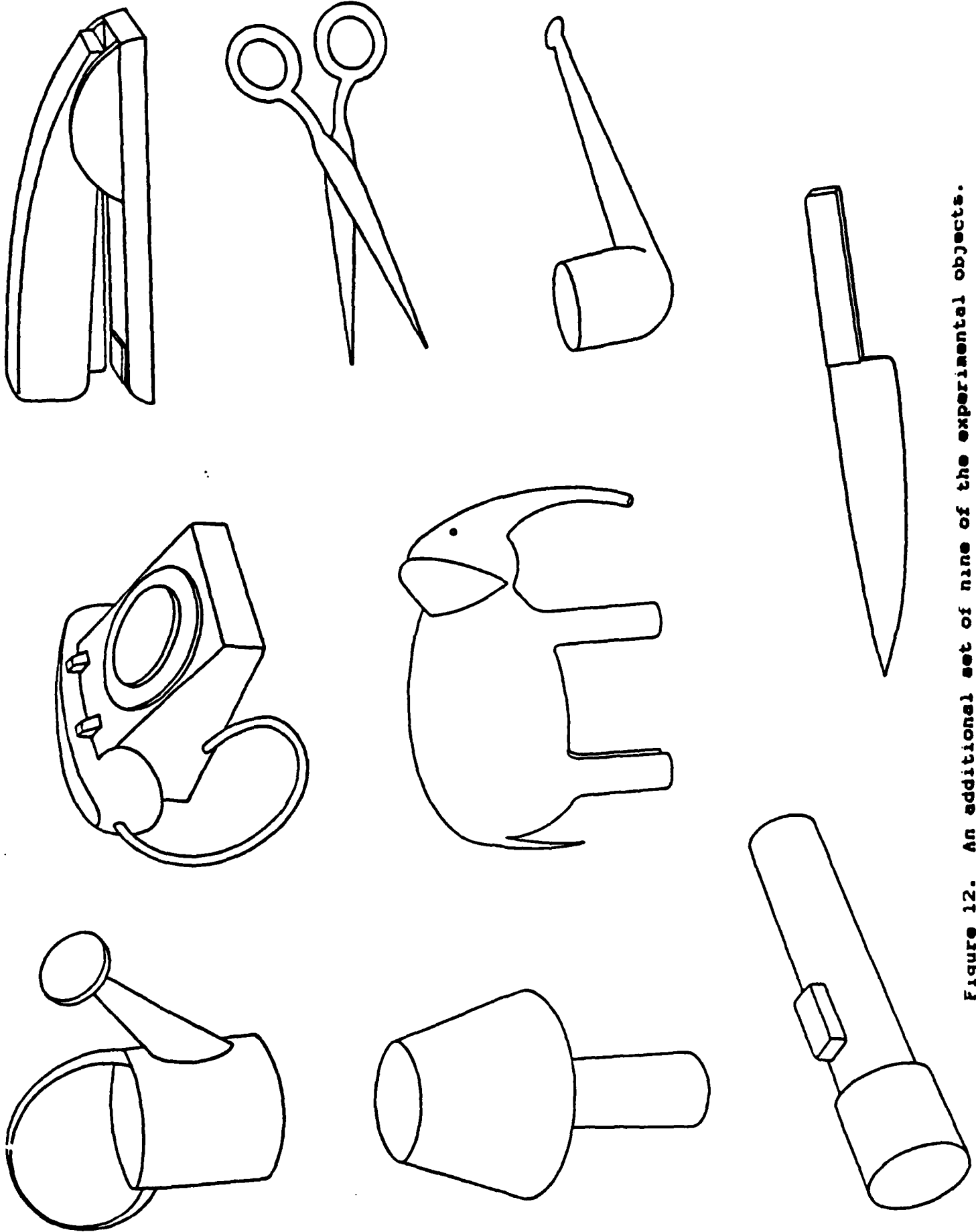
Figure 11. Nine of the experimental objects.

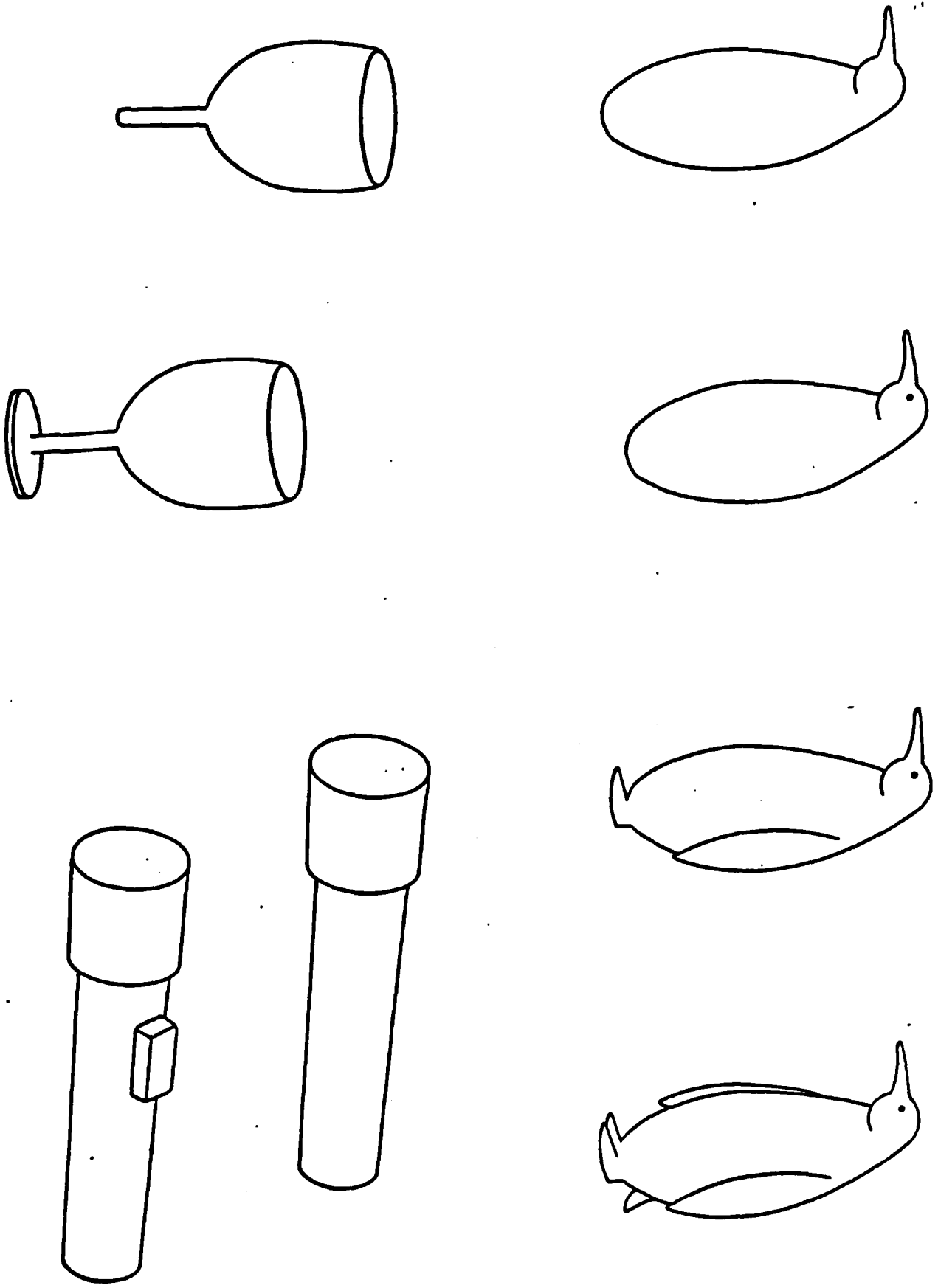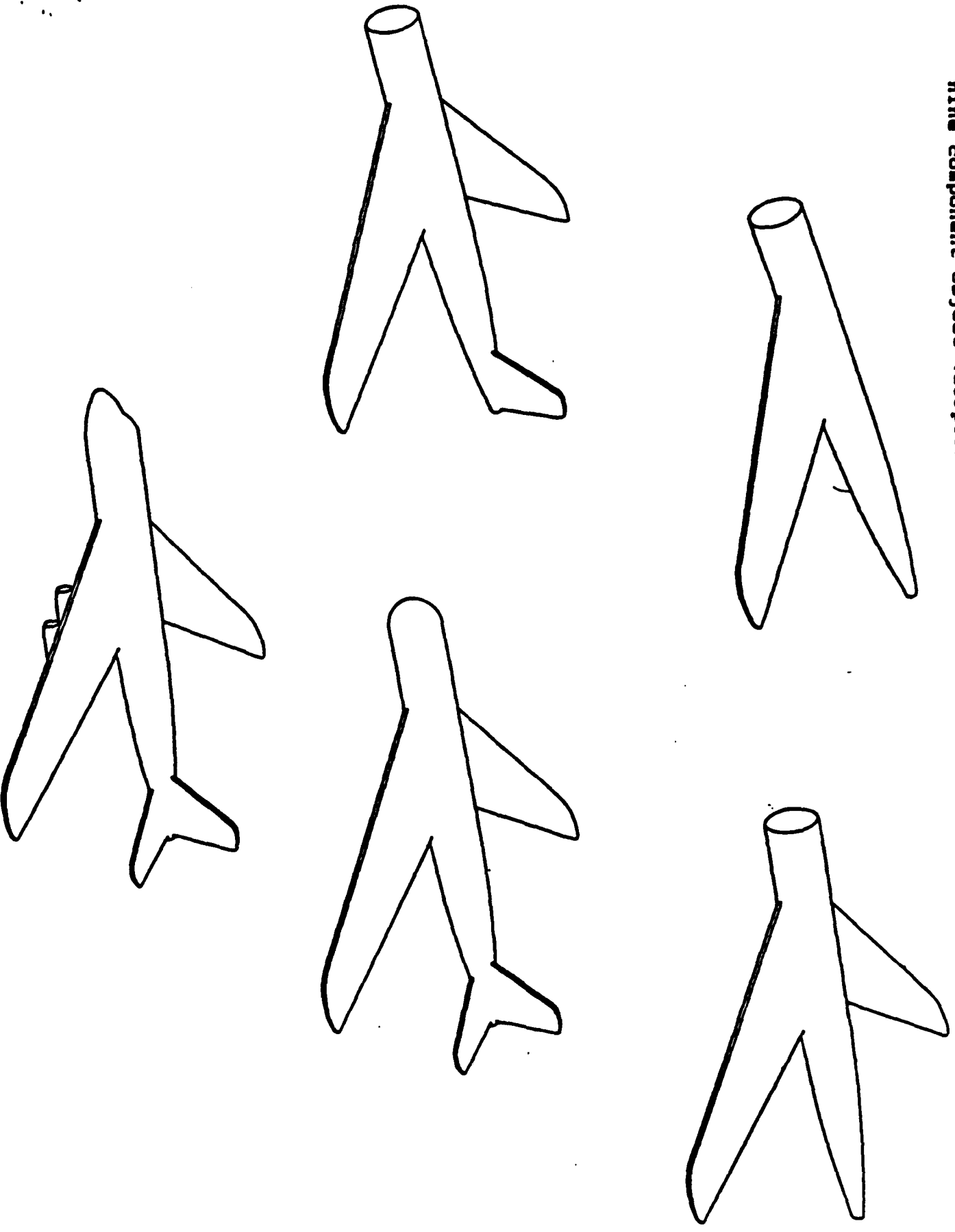Figure 12. An additional set of nine of the experimental objects.

Figure 13. Illustration of the partial and complete versions of two three component objects (the wine glass and flashlight) and a nine component object (the penguin).

Figure 14. Illustration of partial and complete versions of a nine component object (airplane).

components. The prediction from RBC was that complex objects, by furnishing more diagnostic combinations of components, would be more rapidly identified than simple objects. This prediction is contrary to those models that postulate that objects are recognized through a aerial contour tracing process (e.g., Hochberg, 1978; Ullman, 1983).

General procedure. Trials were self paced. The depression of a key on the subject's terminal initiated a sequence of exposures from three projectors. First, the corners of a 500 msec fixation rectangle (6° wide) which corresponded to the corners of the object slide was shown. The fixation slide was immediately followed by a 100 msec exposure of a slide of an object that had varying numbers of its components present. The presentation of the object was immediately followed by a 500 msec pattern mask consisting of a random appearing arrangement of lines. The subject's task was to name the object as fast as possible into a microphone which triggered a voice key. The experimenter recorded errors. Prior to the experiment, the subjects read a list of the object names to be used in the experiment. [Subsequent experiments revealed that this procedure for name familiarization produced no effect. When subjects were not familiarized with the names of the experimental objects, results were virtually identical to when such familiarization was provided. This result indicates that the results of these experiments are not a function of inference over a small set of objects.] Even with the name familiarization, all responses that indicated that the object was identified were considered correct. Thus "pistol," "revolver," "gun," and "handgun" were all acceptable as correct responses for the same object. RTs were recorded by a microcomputer which also controlled the projectors and provided speed and accuracy feedback on the subject's terminal after each trial.

Design. Objects were selected that required 2, 3, 6, or 9 components to look complete. There were nine objects for each of these complexity levels yielding a total set of 36 objects. The various combinations of the partial versions of these objects brought the total number of experimental trials (slides) to 99. Each of 48 subjects viewed all the experimental slides. in addition, two slides of other objects preceded and followed each block of experimental trials as buffer slides for warm up. These were not included in the data analyses.

The various conditions are notated as follows: the digit in parenthesis indicates the number of displayed components and the digit preceding the parenthesis indicates the number of components required for the object to look complete. Thus the airplane shown with three of its nine components would be designated as 9(3). The combinations used were: 2(2), 3(2), 3(3), 6(3), 6(4), 6(5), 6(6), 9(3), 9(4), 9(6), and 9(9). The 11 conditions with nine objects each yielded 99 experimental trials that were organized into 2 blocks of 53 or 54 trails each (the 44 or 45 experimental slides plus two buffer slides at the beginning and end of each block). The blocks were balanced by Latin square and run forward and backward, so that each slide had the same mean aerial position.

Results.   Figure 15 shows the mean error rates as a function of
the number of components actually displayed on a given trial for the
conditions in which no familiarization was provided. Each function is
the mean for the nine objects at a given complexity level.

Insert Figure 15 About Here

Each subject saw all 99 slides but only the data for the first time
that a subject viewed a particular object will be discussed here.
These responses were unaffected by prior trials in which the subject
might have viewed that object in partial or complete form.   (The
primary effect of including prior trials of an object was to improve
the performance on those trials where the subjects viewed a partial
object that had previously been experienced in a complete or more
complete version.)   For a given level of complexity, increasing
numbers of components resulted in better performance but error rates
were modest.   When only three or four components for the complex
objects (those with six or nine components to look complete) were
present, subjects were almost 90 percent accurate (10 percent error
rate).   In general, the complete objects were named without error so
it is necessary to look at the RTs to see if differences emerge for
the complexity variable.

Mean correct RTs, shown in figure 16, provide the same general
outcome as the errors, except that there was a slight tendency for the
more complex objects, when complete, to have shorter RTs than the

----
Insert Figure 16 About Here
----

simple objects.   This advantage for the complex objects was actually
underestimated in that the complex objects had longer names (three and
four syllables) and were less familiar than the simple objects.
Oldfield (1959) showed that object-naming RTs were longer for names
that have more syllables or are infrequent. This effect of slightly
shorter RTs for naming complex objects has been replicated and it
seems safe to conclude, conservatively, that complex objects do not
require more time for their identification than simple objects. This
result is contrary to aerial-contour tracing models of shape
perception (e.g., Hochberg, 1978; Ullman, 1984; Noton & Stark, 1966).
Such models would predict that complex objects would require more time
to be seen as complete compared to simple objects, which have less
contour to trace.   The slight RT advantage enjoyed by the complex
objects is an effect that would be expected if their additional
components were affording a redundancy gain from more possible
diagnostic matches to their representations in memory.

LINE DRAWINGS VS COLORED PHOTOGRAPHY

The components that are postulated to be the critical units for
recognition can be depicted by a line drawing.   Color and texture

Figure 15. Mean per cent error as a function of the number of components in the displayed object (abscissa) and the number of components required for the object to appear complete (parameter). Each point is the mean for nine objects on the first occasion when a subject saw that particular object.
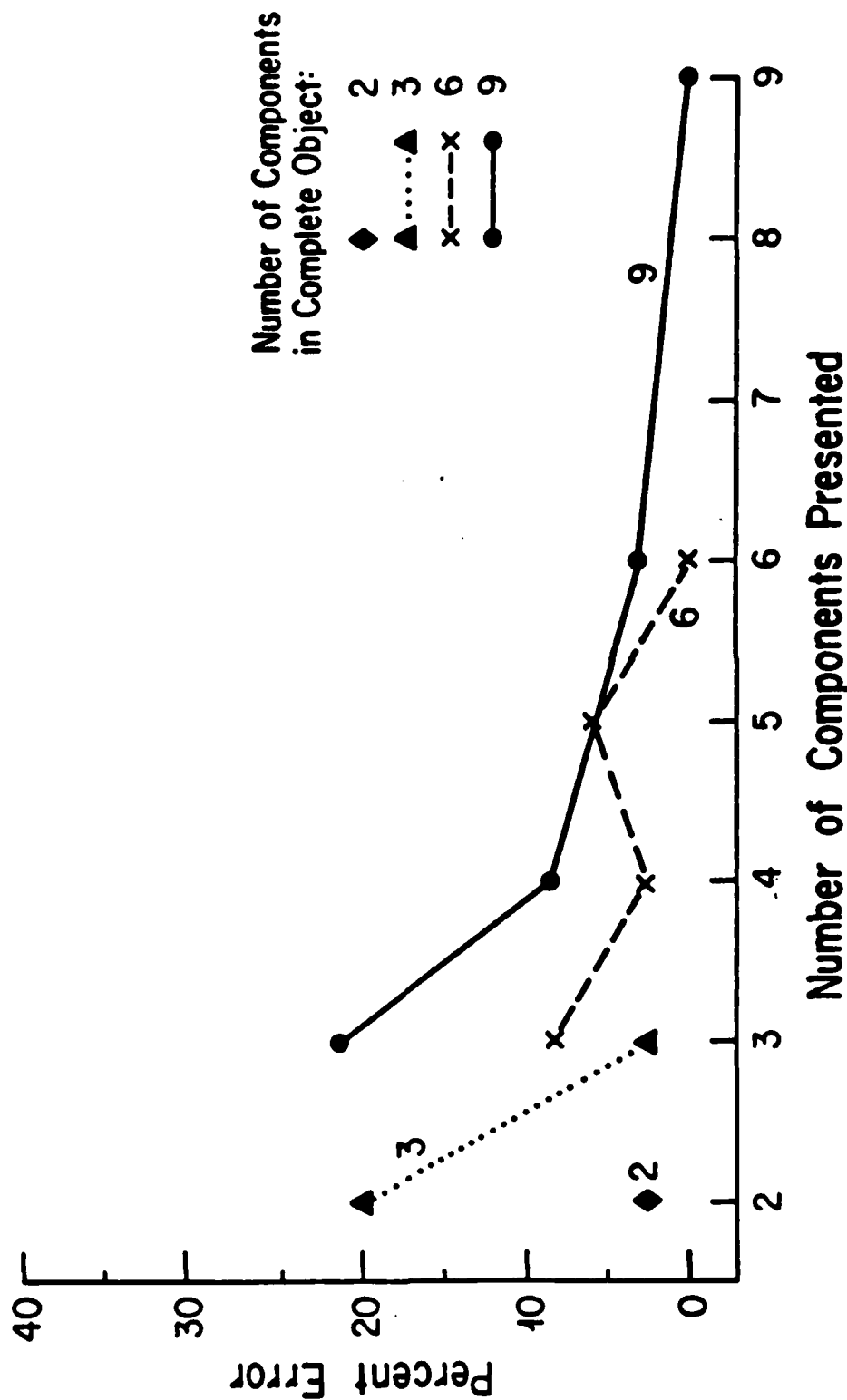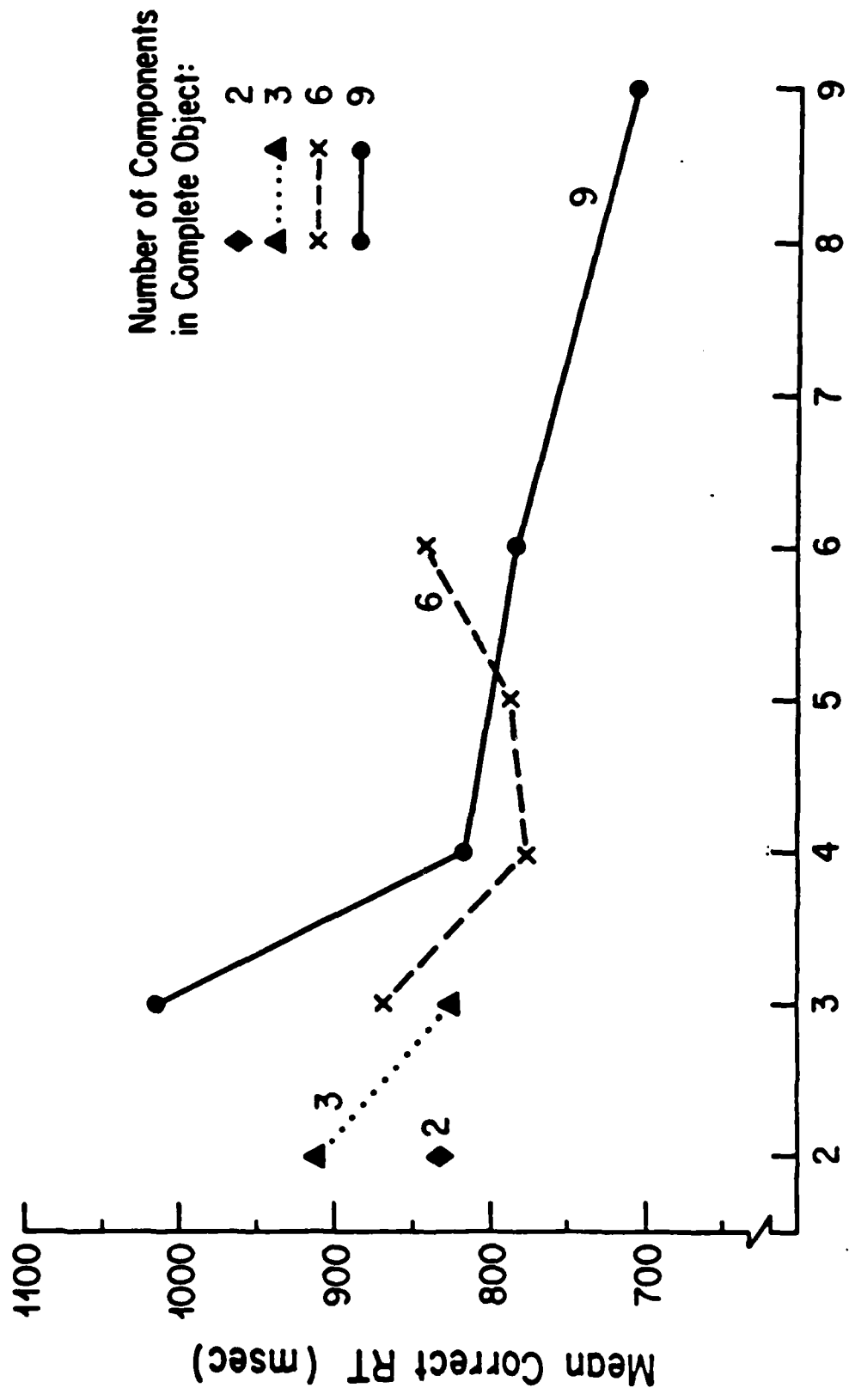
Figure 16. Mean correct reaction time as a function of the number of components in the displayed object (abscissa) and the number of components required for the object to appear complete (parameter). Each point is the mean for nine objects on the first occasion when a subject saw that particular object.

would be secondary routes for recognition. From this perspective, Biederman and Ju (1985) reasoned that naming RTs for objects shown as line drawings should closely approximate naming RTs for those objects when shown as colored photographic slides with complete detail, color, and texture. To our knowledge, no previous experiment had compared these different forms of representing objects on the speed and accuracy of basic-level object classification.[9]

General method. The general procedure and design closely followed that described for the previously described experiment. Thirty subjects viewed brief presentations of slides of line drawings and professionally photographed full colored slides of the same objects in the same orientation.

A line drawing and colored photography version of each of 29 objects yielded 58 experimental slides. Conditions of exposure, luminance, and masking were selected which would favor the colored slides, so RT correlates of this advantage could be explored. An earlier experiment had shown that the colored slides were more adversely affected by a mask (a colored slide of a complex collage of many colored shapes and textures), so the mask was omitted. [The effects of a number of variables on the difference between line drawings and colored slides are described in another report (Biederman & Ju, 1985].

Results. Mean correct naming times were 804 msec for the line drawings and 784 msec for the colored slides. Error rates averaged 2% for both conditions.

An analysis of the individual stimuli indicated that the 20 sec naming RT advantage for the colored slides was not due to a contribution of color or lightness (and often texture) of these stimuli. This was determined by partitioning the slides into two sets: those whose color was diagnostic as to the objects' identity (e.g., mushroom, fork, camera, fish) and those objects whose color was not diagnostic to their identity (e.g., chair, hair dryer, pen, mitten). If color was responsible for the 20 msec advantage, those

---

[9]An oft cited study, Ryan & Schwartz (1956), did compare photography (black & white) against line and shaded drawings and cartoons. Subjects had to determine not the basic level categorization of an object but which one of four configurations of three objects (the positions of five double-throw electrical knife switches, the cycles of a steam valve, and the fingers of a hand) was being depicted. For two of the three objects, the cartoons had lower thresholds than the other modes. But stimulus sampling and drawings and procedural specifications make it difficult to interpret this experiment. For example, the determination of the switch positions was facilitated in the cartoons by filling in the handles so they contrasted with the background contacts. The cartoons did not have lower thresholds than the photographs for the hands, the stimulus example that is most frequently shown in secondary sources (e.g., Neisser, 1966; Rock, 1984). Even without a mask, threshold presentation durations were an order of magnitude longer than was required in the present study.

objects for which it was diagnostic should have had a greater
advantage for the color slides over the line drawings. But the
opposite was true. Objects (N=12) whose color was not diagnostic
enjoyed a 33 msec color advantage compared to an 8 msec color
advantage for the color-diagnostic objects (N=17). Thus, the slight
advantage in naming speed for the colored slides was not a consequence
of the diagnostic use of color and brightness but, in our opinion,
likely derived from more accurate rendition of the components. For
example, a number of the objects or parts, such as the hairdryer or
front leg of the elephant, were drawn in silhouette so they appear
planar.

The conclusion from these studies is that simple line drawings,
when depicting the complete object, can be identified almost as
quickly (within 20 msec) as a full-colored slide of that same object.
That simple line drawings can be identified so rapidly as to approach
the naming speed of fully detailed, textured, colored photographic
slides supports the premise that the earliest access to a mental
representation of an object can be modeled as a matching of a line-
drawing representation of a few simple components. Such componential
descriptions are thus sufficient for primal access.

THE PERCEPTION OF DEGRADED OBJECTS

Evidence that a componential description may be necessary for
object recognition (under conditions where contextual inference is not
possible) derives from experiments on the perception of objects which
have been degraded by deletion of their contour (Biederman & Blickle,
1985).

RBC holds that parsing of an object into components is performed
at regions of concavity. The nonaccidental relations of collinearity
and curvilinearity allow filling-in: They extend broken contours that
are collinear or smoothly curvilinear. In concert, the two
assumptions of: (a) parsing at concavities, and (b) filling-in through
collinearity or smooth curvature lead to a prediction as to what
should be a particularly disruptive form of degradation: If contours
were deleted at regions of concavity in such a manner that their
endpoints, when extended through collinearity or curvilinearity,
bridge the concavity, then the components would be lost and
recognition should be impossible. The cup in the right column of the
top row of figure 17 provides an example. The curve of the handle of
the cup is drawn so that it is continuous with the curve of the
cylinder forming the back rim of the cup. This form of degradation,
in which the components cannot be recovered from the input through the
nonaccidental properties, is referred to as nonrecoverable degradation
and is illustrated for the objects in the right column of figure 17.

Insert Figure 17 About Here

An equivalent amount of deleted contour in a midsection of a
curve or line should prove to be less disruptive as the components

Figure 17. Example of five stimulus objects in the experiment on the
perception of degraded objects. The left column shows the original
intact versions. The middle column shows the recoverable versions.
The contours have been deleted in regions where they can be replaced
through collinearity or smooth curvature. The right column shows
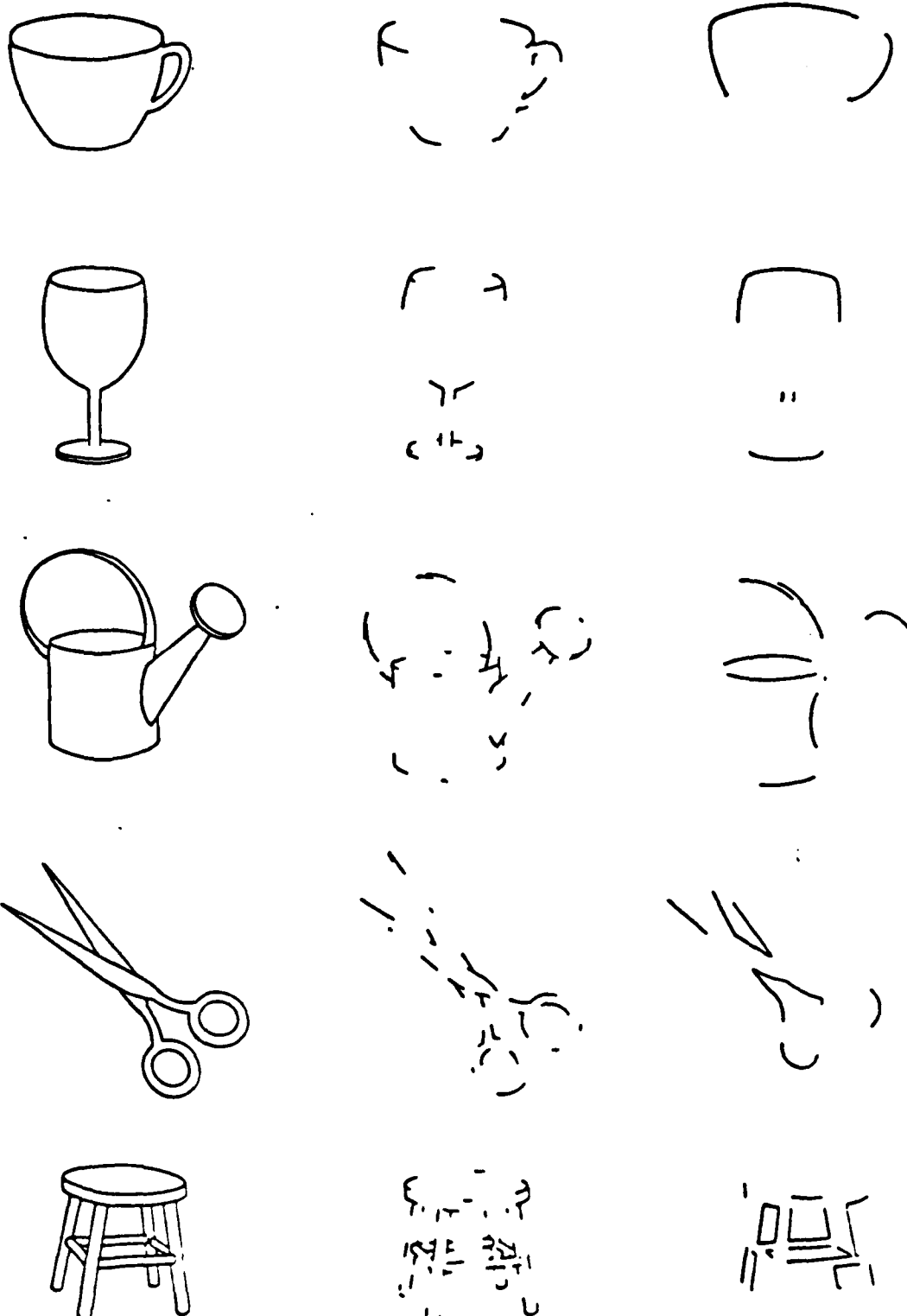
the nonrecoverable versions. The contours have been deleted at
regions of concavity so that collinearity or smooth curvature of the
segments bridges the concavity. In addition, vertices have been
altered, e.g., from Ys to Ls, and misleading symmetry and
parallelism introduced.

could then be restored through collinearity or curvature. In this case the components should be <u>recoverable</u>. Example of recoverable forms of degradation are shown in the middle column of figure 17.

<u>General method</u>. Recoverable and nonrecoverable versions of 35 objects were prepared, yielding 70 experimental slides. In addition to the procedures for producing nonrecoverable versions described above, components were also camouflaged by contour deletion that produced symmetry, parallelism, and vertices that were not characteristic of the original object. For example, in figure 17, the watering can has false vertices suggested in the region of its spout and the stool has a number of T vertices transformed to L vertices. Symmetrical regions of the stool also suggest components where they would not be parsed in the original intact version. Even with these techniques, it was difficult to remove all the components and some remained in nominally nonrecoverable versions, as with the handle of the scissors.

The slides were arranged in two blocks, each with all 35 objects. Approximately half (17 or 18) of the slides in each version were recoverable and the other half were unrecoverable versions. Slides were displayed for 100, 200, or 750 msec. Four sequences were used in which the order of the blocks was balanced and half the subjects viewed each block in forward order; the other half in reverse order. These orders were balanced over slide durations so that each slide (a) had the same mean serial position, and (b) was presented with equal frequency at the three presentation durations. A separate group of six subjects viewed the slides at a 5 sec exposure duration.

Prior to the experiment, all subjects were shown several examples of the various forms of degradation for several objects that were not used in the experiment. In addition, familiarization with the experimental objects was manipulated between subjects. Prior to the start of the experimental trials, different groups of six subjects: (a) viewed a three second slide of the intact version of the objects, e.g., the objects in the left column of Fig. 17, which they named, (b) were provided with the names of the objects on their terminal, or (c) were given no familiarization. As in the prior experiments, the subjects task was to name the objects.

A glance at the second and third columns in figure 15 is sufficient to reveal that one doesn't need an experiment to show that the nonrecoverable objects would be more difficult to identify than the recoverable versions. But we wanted to determine if the nonrecoverable versions would be identifiable at extremely long exposure durations (5 sec.) and whether the prior exposure to the intact version of the object would overcome the effects of the contour deletion. The effects of contour deletion in the recoverable condition was also of considerable interest when compared to the comparable conditions from the partial object experiments.

<u>Results</u>. The error data are shown in figure 18. Identifiability of the nonrecoverable stimuli was virtually impossible: The median

error rate for those slides was 100 per cent. Subjects rarely guessed
wrong objects in this condition. Almost always they merely said that

----
Insert Figure 18 About Here
----

they "didn't know." In those few cases where a nonrecoverable object
was identified, it was for those instances where some of the
components were not removed, as with the circular rings of the handles
of the scissors. Even at 5 sec, error rates for the nonrecoverable
stimuli, especially in the Name and No Familiarization conditions, was
extraordinarily high. Objects in the Recoverable condition were named
at high accuracy at the longer exposure durations,

       There was no effect of familiarizing the subjects with the names
of the objects compared to the condition in which the subjects were
provided with no information about the objects.   There was some
benefit, however, of providing intact versions of the pictures of the
objects.   Even with this familiarity,   performance in the
Nonrecoverable condition was extraordinarily poor, with error rates
exceeding 60 per cent when subjects had a full five seconds for
deciphering the stimulus. As noted previously, even this value
underestimated the difficulty of identifying objects in the
Nonrecoverable condition, in that identification was possible only
when the contour deletion allowed some of the components to remain
recoverable.

       The emphasis on the poor performance in the Nonrecoverable
condition should not obscure the extensive interference that was
evident at the brief exposure durations in the Recoverable condition.
The previous experiments had established that intact objects, without
picture familiarization, could be identified at near perfect accuracy
at 100 msec. At this exposure duration, error rates for the
recoverable stimuli in the present experiment, whose contours could be
restored through collinearity and curvature, were approximately 65
percent. The high error rates at 100 msec exposure duration suggests
that these filling-in processes require both time (on the order of 200
msec) and an image--not merely a memory representation--to be
successfully executed.

       The dependence of componential recovery on the availability of
contour and time was explored parametrically by Biederman, Ju, &
Beiring (1985). To produce the nonrecoverable versions of the objects
it was necessary to delete or modify the vertices. The recoverable
versions of the objects tended to have their contours deleted in
midsegment. It is possible that some of the interference in the
nonrecoverable condition was a consequence of the removal of vertices,
rather than the production of inappropriate components. The
experiment also compared these two loci (vertex or midsegment) as
sites of contour deletion. Contour deletion was performed either at
the vertices or at midsegments for 18 objects, but without the
accidental bridging of components through collinearity or curvature
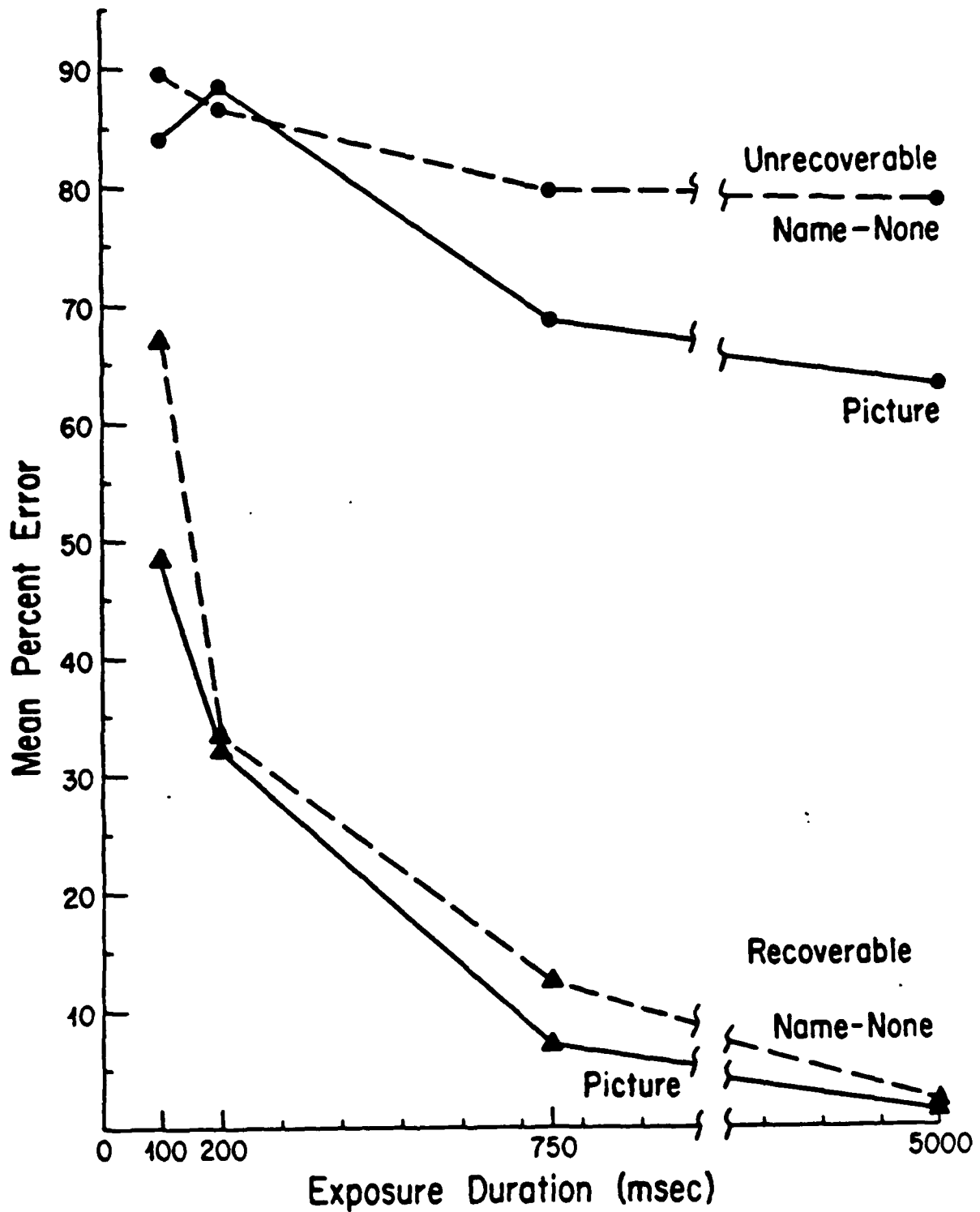that was characteristic of the nonrecoverable condition. The percent

Figure 18. Mean per cent errors in object naming as a function of
exposure duration, nature of contour deletion (Recoverable vs.
Nonrecoverable components), and prefamiliarization (None, Name, or
Picture). No differences were apparent between the None and Name
pretraining conditions so they have been combined into one function.

contour removed was also varied with values of 25, 45, and 65 percent
removal and the objects were shown for 100, 200, or 750 msec. Other
aspects of the procedure were identical to the previous experiments
with only name familiarization provided. Figure 19 shows an example
for a single object.

```
        --------------------------
        Insert Figure 19 About Here
        --------------------------
```

The mean percent errors are shown in Figure 20. At the briefest
exposure duration and the most contour deletion (100 msec exposure
duration and 65 percent contour deletion), removal of the vertices
resulted in considerably higher error rates then the midsegment
removal, 54 and 31 percent errors, respectively. With less contour
deletion or longer exposures, the locus of the contour deletion had

```
        --------------------------
        Insert Figure 20 About Here
        --------------------------
```

only a slight effect on naming accuracy. Both types of loci showed a
consistent improvement with longer exposure durations, with error
rates below 10 percent at the 750 msec duration. By contrast, the
error rates in the nonrecoverable condition in the prior experiment
exceeded 75 percent, even after 5 sec. the filling-in of contours,
whether at midsegment or vertex, is a process that can be completed
within 1 sec.   But the suggestion of a misleading component through
collinearity or curvature produces an image that cannot index the
original object, no matter how much time there is to view the image.
Although accuracy was less affected by the locus of the contour
deletion at the longer exposure durations and the lower deletion
proportions, there was a consistent advantage on naming latencies of
the midsegment removal, as shown in figure 21. (The lack of an effect
at the 100 msec exposure duration with 65 percent deletion is likely a
consequence of the high error rates for the vertex deletion stimuli.)
This result shows that if contours are deleted at a vertex they can be
restored, as long as there is no accidental filling-in, but the

```
        --------------------------
        Insert Figure 21 About Here
        --------------------------
```

restoration will require more time than when the deletion is at
midsegment.   Overall, both the error and RT data document a striking
dependence of object identification on what RBC assumes to be a prior
stage of componential determination.

Perceiving degraded vs. partial objects. In the experiments with
partial objects and contour deletion, objects were shown with less
than their full amount of contour. With the partial objects, the
missing contours were in the form of complete components that were
missing; the components that were present were present in intact form.
With the degraded objects, the deleted contour was distributed across

Figure 19. Illustration for a single object of 25, 45, and 65 percent contour removal centered at either midsegment or vertex.

Figure 20.   Mean percent object naming errors as a function of locus
    of contour removal (midsegment or vertex), percent removal, and
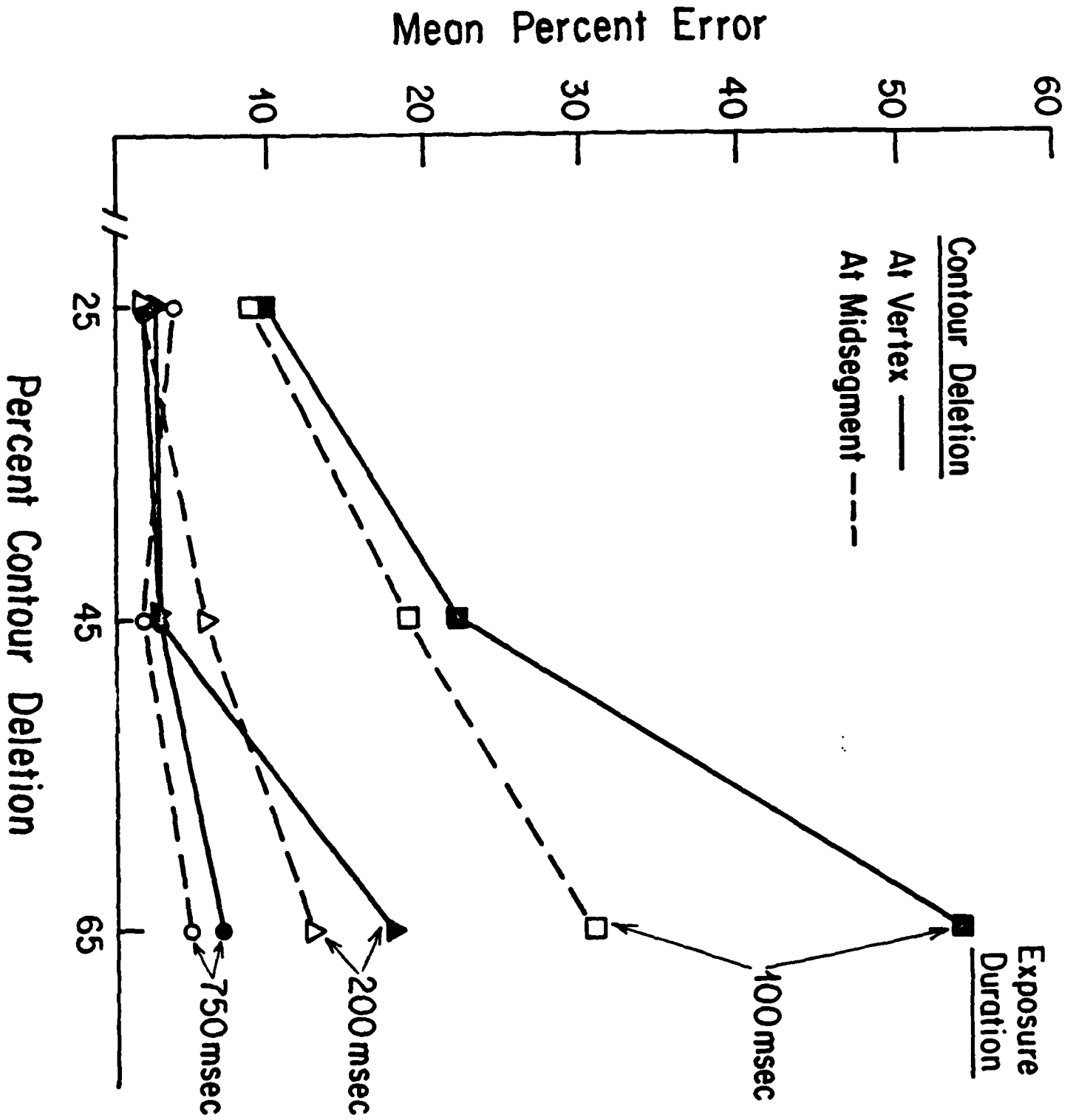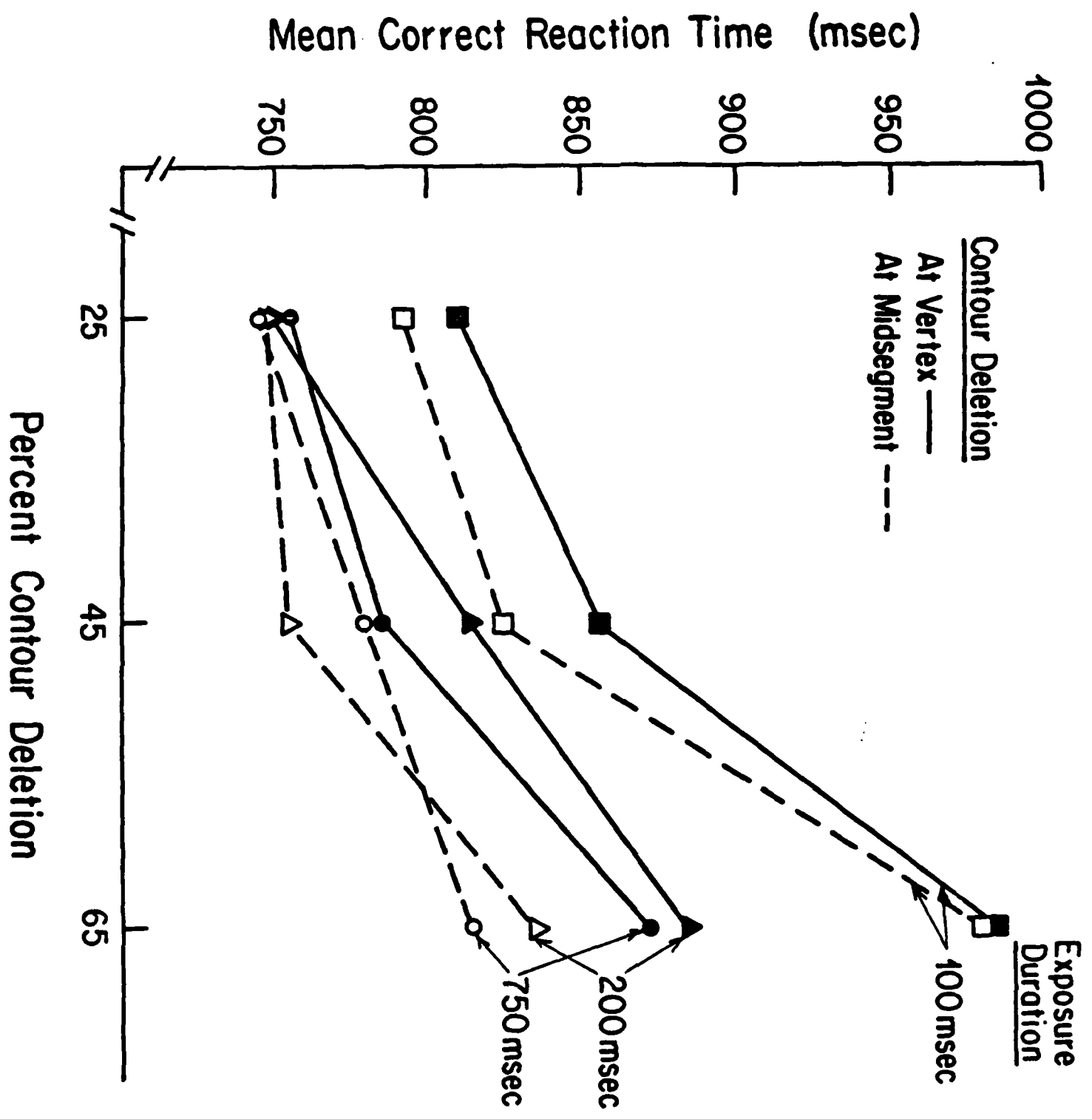   exposure duration.

Figure 21. Mean correct object naming reaction time (msec) as a function of locus of contour removal (midsegment or vertex), percent removal, and exposure duration.

all of the object's components. Biederman, Beiring, & Ju (1985)
compared the effects of midsegment contour deletion, where the
contours could be restored through collinearity or curvature, with the
removal of whole components, when an equivalent amount of contour was
deleted for each object. With partial objects, it is unlikely that
the missing components are added imaginally, prior to recognition.
Logically, one would have to know what object was being recognized to
know what parts to add. Instead, indexing (addressing) a
representation most likely proceeds in the absence of the parts. The
two methods for removing contour are thus seen as affecting different
stages. Deleting contour in midsegment affects affects processes
prior to and including those involved in the determination of the
components (fig. 3). Removing 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 relative degree of disruption from the two
methods is not, as yet, a prediction that can be made by RBC. If it
is assumed that contour filling-in is a fast, low level process then a
demonstration that partial objects (with only three of their six or
nine components present) can be recognized more readily than objects
whose contours can be restored through filling-in, documents the high
efficiency of those three components in accessing a representation.

     The procedure for this experiment closely followed the previous
experiments. The stimuli were the 18 objects requiring six or nine
components to look complete in the partial object experiment. The
three component versions of these objects were selected as the partial
object stimuli. For each of these objects, contour was deleted in
midsegment to produce a version that had the same amount of contour
removed. For example, removing six of the nine components of the
stool removed 45 percent of its contour. The degraded version of the
stool also had 45 percent of its contour removed, except that the
removal was distributed in midsegment throughout the object. The mean
deletion was 33 percent (S.D. 15.6; range 11.7 to 68.2 percent).
Objects were presented for 65, 100, and 200 msec.

     At the shortest (65 msec) exposure duration, removing components
was less disruptive than deleting contours in midsegment, 27 to 42
percent errors, respectively (figure 22). This difference was reduced

```
----------------------------
Insert Figure 22 About Here
----------------------------
```

and even reversed at the longer exposure durations. The RTs (figure
23) show the interaction even more strongly.

```
----------------------------
Insert Figure 23 About Here
----------------------------
```

The result of this comparison provides additional support for the
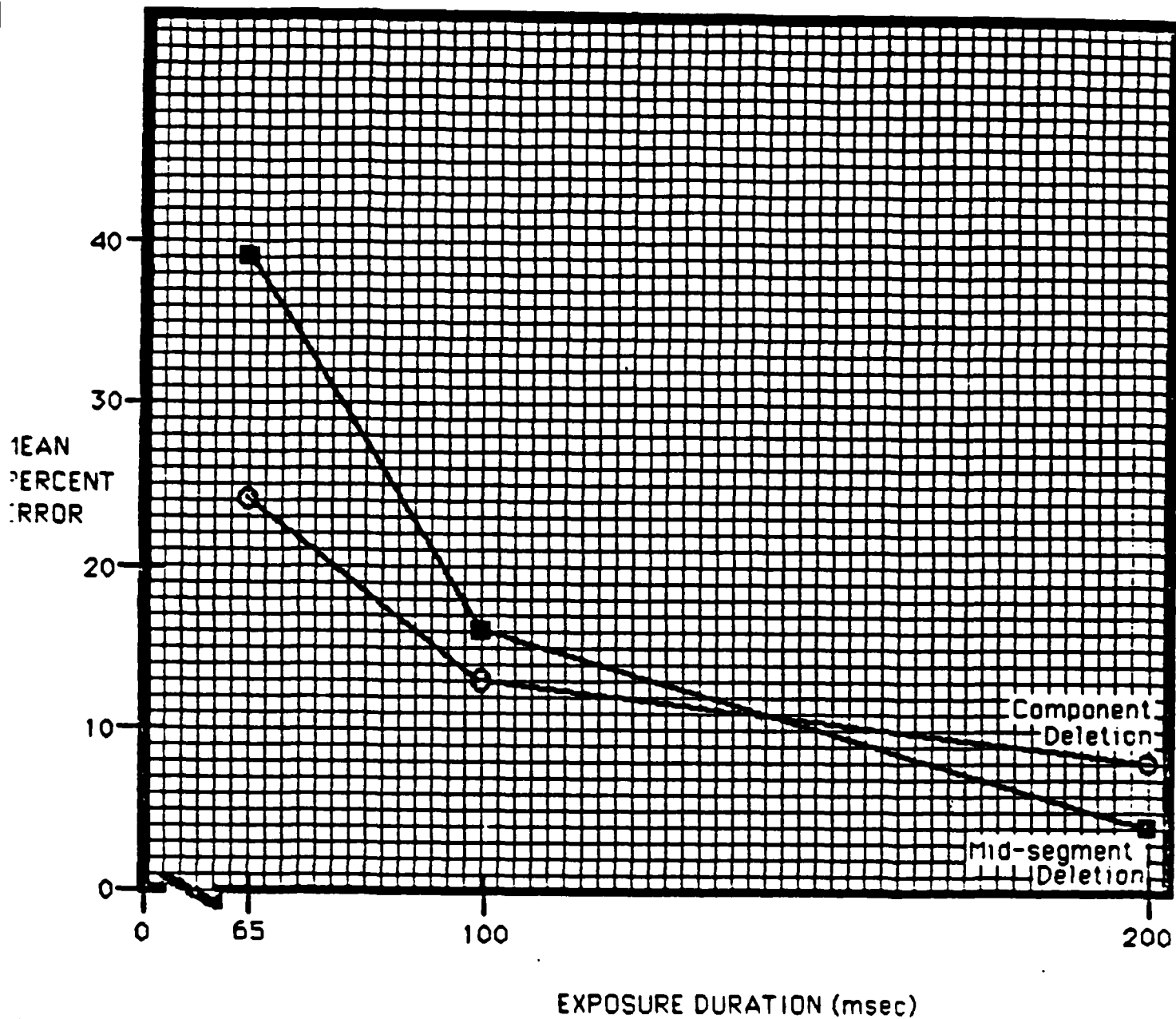dependence of object recognition on componential identification. RBC

Figure 22. Mean percent errors of object naming as a function of the nature of contour removal (deletion of midsegment or component) and exposure duration.
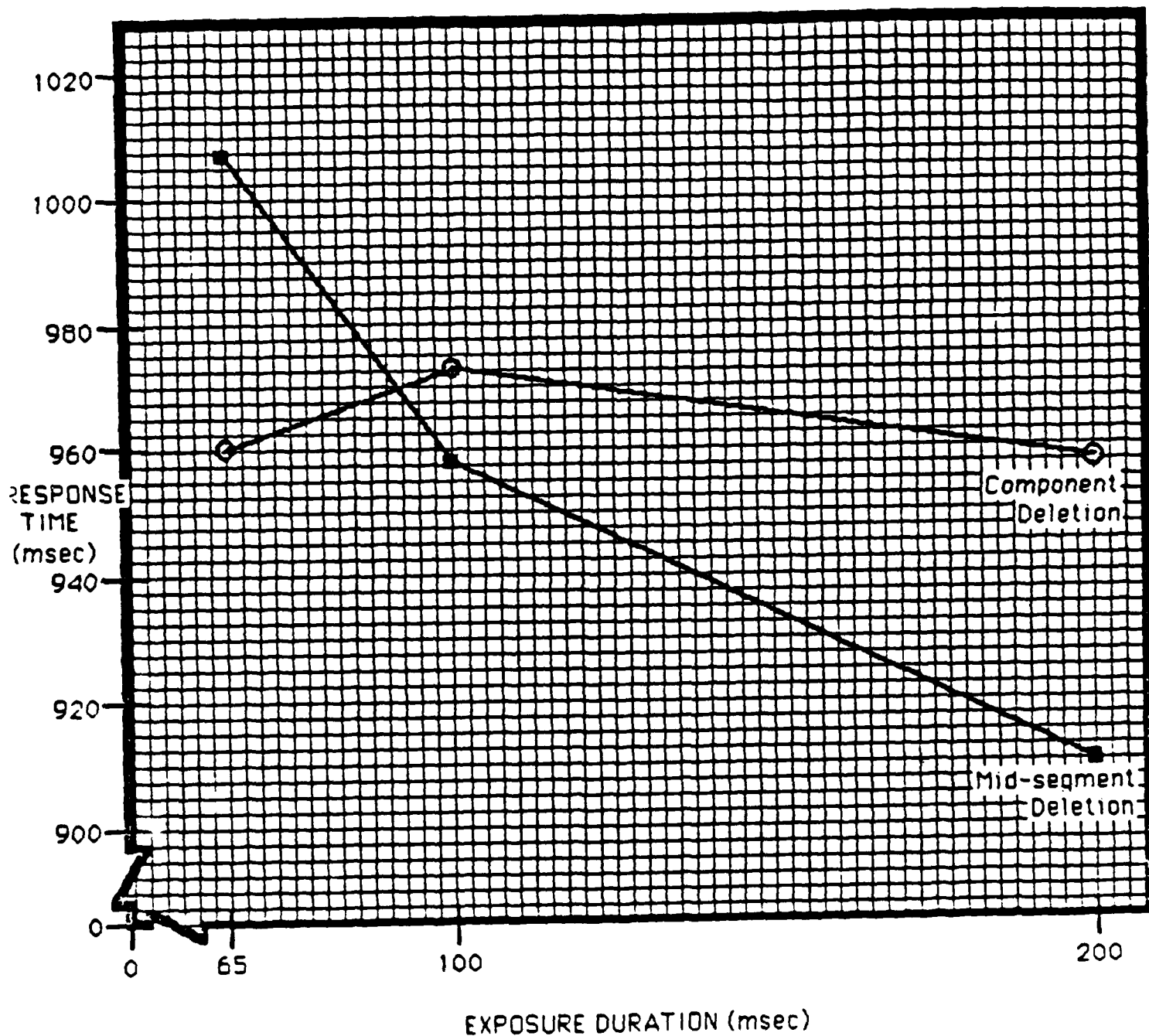
Figure 23. Mean correct reaction time (msec.) in object naming as a function of the nature of contour removal (deletion of midsegment or component) and exposure duration.

posits that a sufficient input for recognition is a diagnostic subset of a few components (a partial object). If all of an object's components were degraded (but recoverable), recognition would be delayed while the contours were restored through the filling-in routines. Once the filling-in was completed, a better match to the object's representation would be possible than with a partial object that had only a few of its components. Longer exposure durations increase the likelihood that filling-in would be completed. The results indicate that the costs on identification speed and accuracy for contour deletion were greater than the costs from removing some of an object's components at the briefest exposure durations. A subjective demonstration of the processing time required for contour restoration is presented in the next section.

Contour deletion by occlusion. The degraded recoverable objects in the right columns of figure 17 have the appearance of flat drawings of objects with interrupted contours. Biederman & Blickle (1985) designed a demonstration of the dependence of object recognition on componential identification by aligning an occluding surface so that it appeared to produce the deletions. If the components were responsible for an identifiable volumetric representation of the object, we would expect that with the recoverable stimuli, the object would complete itself under the occluding surface and assume a three dimensional character. This effect should not occur in the nonrecoverable condition. This expectation was met as shown in figures 24 and 25. These stimuli also provide a demonstration for the time (and effort?) requirements for contour restoration through collinearity or curvature. We have not yet obtained objective data on this effect, which may be complicated by masking effects from the presence of the occluding surface, but we invite the reader to share our subjective impressions. When looking at a nonrecoverable version of an object in figure 24, no object becomes apparent. In the recoverable version in 25, an object does pop into a 3-D appearance, but most observers report a delay (our own estimate is approximately 500 msec) from the moment the stimulus is first fixated to when it appears as an identifiable 3-D entity.

Insert Figures 24 & 25 About Here

This demonstration of the effects of an occluding surface to produce contour interruption also provides a control for the possibility that the difficulty in the nonrecoverable condition was a consequence of inappropriate figure-ground groupings, as with the stool in Fig. 17. With the stool, the ground that was apparent through the rungs of the stool became figure in the nonrecoverable condition. (In general, however, only a few of the objects had holes in them where this could have been a factor.) This would not necessarily invalidate the RBC hypothesis but merely would complicate the interpretation of the effects of the nonrecoverable noise, in that some of the effect would derive from inappropriate grouping of contours into components and some of the effect would derive from inappropriate figure-group grouping. That the objects in the
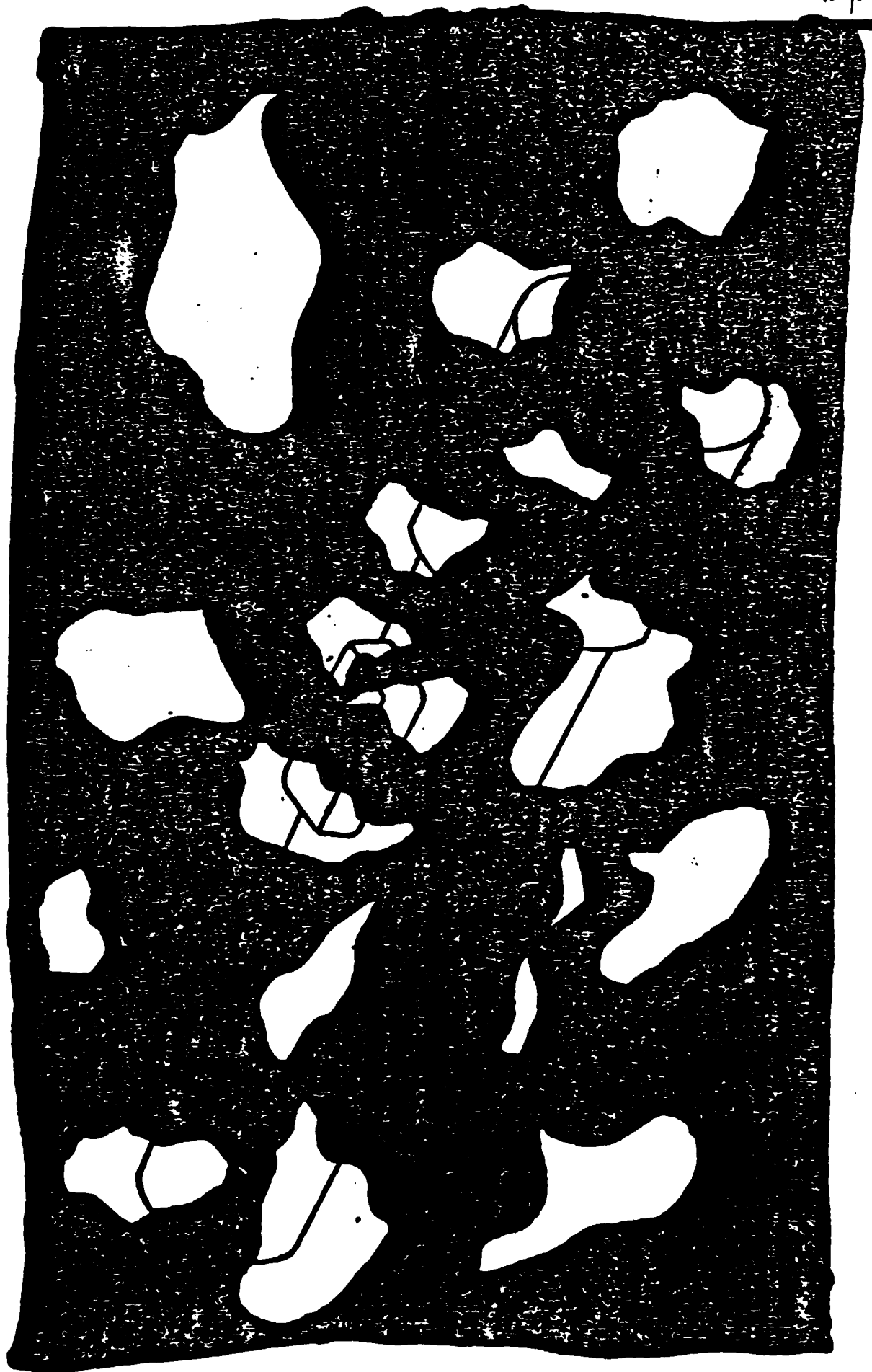
Figure 24. Nonrecoverable version of an object where the contour deletion is produced by an occluding surface.

UP ↑

Figure 25. Recoverable version of an object where the contour deletion is produced by an occluding surface. The object is the same as that shown in figure 24. The reader may note that the 3-D percept in this figure does not occur instantaneously.

nonrecoverable condition remain unidentifiable when the contour interruption is attributable to an occluding surface suggests that figure-ground grouping cannot be the primary cause of the interference from the nonrecoverable deletions.

SUMMARY AND IMPLICATIONS OF THE EXPERIMENTAL RESULTS

The sufficiency of a component representation for primal access to the mental representation of an object was supported by two results: a) that partial objects with two or three components could be readily identified under brief exposures, and b) comparable identification performance between the line drawings and the colored photography. The experiments with degraded stimuli established that the components are necessary for object perception. These results suggest an underlying principle by which objects are identified.

COMPONENTIAL RECOVERY PRINCIPLE

The results and phenomena associated with the effects of degradation and partial objects can be understood as the workings of a single Principle of Componential Recovery: If the components, in their specified relations, can be readily identified, object identification will be fast and accurate. The principle of componential recovery can be readily extended to four additional phenomena in object perception: a) that objects can be more readily recognized from some orientations than other orientations (orientation variability), b) objects can be recognized from orientations not previously experienced (object transfer), c) articulated (or deformable) objects, whose componential relations can be altered, can be recognized even when the specific configuration might not have been experienced previously (deformable object invariance), and d) the perceptual basis of basic level categories.

ORIENTATION VARIABILITY

Objects can be more readily identified from some orientations compared to other orientations (Palmer, Rosch, & Chase, 1981). According to the RBC hypothesis, difficult views will be those in which the components extracted from the image are not the components (and their relations) in the representation of the object. Often such mismatches will arise from an "accident" of viewpoint where an image property is not correlated with the property in the 3-D world. For example, when the viewpoint in the image is parallel to the major components of the object, the resultant foreshortening converts one or some of the components into surface components, such as disks and rectangles in Figure 26, which are not included in the componential description of the object. In addition, as illustrated in Fig. 26,

----
Insert Figure 26 About Here
----

the surfaces may occlude otherwise diagnostic components. Consequently, the components extracted from the image will not readily
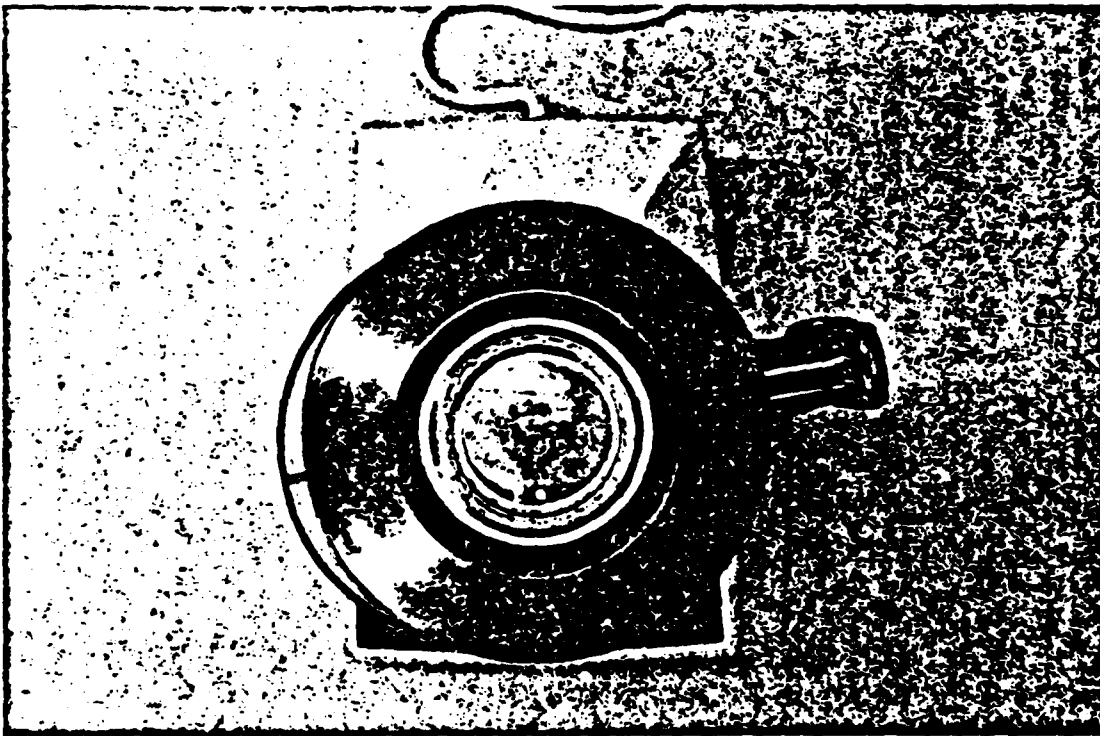
Figure 26.   A viewpoint parallel to the axes of the major components
   of a common object.

match the mental representation of the object and identification will
be much more difficult compared to an orientation, such as that shown
in figure 27, which does convey the components. A second condition
under which viewpoint affects identifiability of a specific object
arises when the orientation is simply unfamiliar, as when a sofa is
viewed from below or when the top-bottom relations among the
components are perturbed as when a normally upright object is
inverted.

Insert Figure 27 About Here

     Palmer, Rosch, & Chase (1981) conducted an extensive study of the
perceptibility of various objects when presented at a number of
different orientations. Generally, a three-quarters front view was
most effective for recognition. Their subjects showed a clear
preference for such views. Palmer et al. termed this effective and
preferred orientation of the object its canonical orientation. The
canonical orientation would be, from the perspective of RBC, a special
case of the orientation that would maximize the match of the
components in the image to the representation of the object.

     An apparent exception to the preference for three-quarters
frontal view preference was Palmer et al.'s (1981) finding that
frontal (facial) views enjoyed some favor in viewing animals. But
there is evidence that routines for processing faces have evolved to
differentially respond to cuteness (Hildebrandt, 1982; Hildebrandt &
Fitzgerald, 1983), age (e.g., Mark & Todd, 1985), and emotion and
threats (e.g., Coss, 1983; Trivers, 1985). Faces may thus constitute
a special stimulus case in that specific mechanisms have evolved to
respond to biologically relevant quantitative variations and caution
may be in order before results with face stimuli are considered as
characteristic of the perception of objects in general.

TRANSFER BETWEEN DIFFERENT VIEWPOINTS

     When an object is seen at one viewpoint or orientation it can
often be recognized as the same object when subsequently seen at some
other viewpoint, even though there can be extensive differences in the
retinal projections of the two views. The componential recovery
principle would hold that transfer between two viewpoints would be a
function of the componential similarity between the views. This could
be experimentally tested through priming studies with the degree of
priming predicted to be a function of the similarity (viz., common
minus distinctive components) of the two views. If two different
views of an object contained the same components RBC would predict
that, aside from effects attributable to variations in aspect ratio,
there should be as much priming as when the object was presented at an
identical view. An alternative possibility to componential recovery
is that a presented object would be mentally rotated (Shepard &
Metzler, 1971) to correspond to the original representation. But
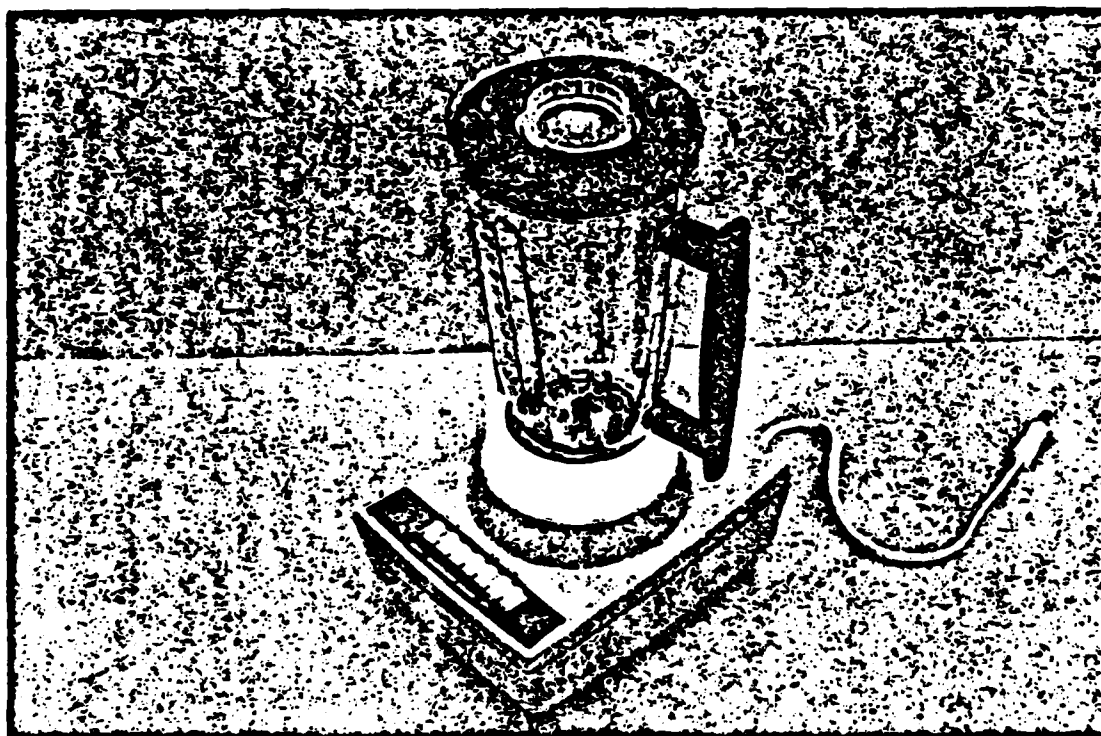mental rotation rates appear to be too slow and effortful to account

Figure 27.   The same object as in figure 26, but with a viewpoint not
   parallel to the major components.

for the ease and speed in which transfer occurs between different orientations.

There may be a restriction on whether a similarity function for priming effects will be observed. Although unfamiliar objects (or nonsense objects) should reveal a componential similarity effect, the recognition of a familiar object, whatever its orientation, may be too rapid to allow an appreciable experimental priming effect. Such objects may have a representation for each orientation that provides a different componential description. Bartram's (1974) results support this expectation that priming effects might not be found across different views of familiar objects. Bartram performed a series of studies in which subjects named 20 pictures of objects over eight blocks of trials. [In another experiment, Bartram (1976) reported essentially the same results from a Same-Different name matching task in which pairs of pictures were presented.] In the _Identical_ condition, the pictures were identical across the trial blocks. In the _Different View_ condition, the same objects were depicted from one block to the next but in different orientations. In the _Different Exemplar_ condition, different exemplars, e.g., different instances of a chair, were presented, all of which required the same response. Bartram found that the naming RTs for the Identical and Different View conditions were equivalent and both were shorter than control conditions, described below, for concept and response priming effects. Bartram theorized that observers automatically compute and access all possible 3-D viewpoints when viewing a given object. Alternatively, it is possible that there was high componential similarity across the different views and the experiment was insufficiently sensitive to detect slight differences from one viewpoint to another. However, in four experiments with colored slides, we (Biederman & Lloyd, 1985) failed to obtain any effect of variation in viewing angle and have thus replicated Bartram's basic effect (or lack of an effect). At this point, our inclination is to agree with Bartram's interpretation, with somewhat different language, but restrict its scope to familiar objects. It should be noted that from Bartram's and our results are inconsistent with a model that assigned heavy weight to the aspect ratio of the image of the object or postulated an underlying mental rotation function.

DIFFERENT EXEMPLARS WITHIN AN OBJECT CLASS

Just as we might be able to gauge the transfer between two different views of the same object based on a componential based similarity metric, we might be able to predict transfer between different exemplars of a common object, such as two different instances of a lamp or chair.

Bartram (1974) also included a _Different Exemplar_ condition, in which different objects with the same name, e.g., different cars, were depicted from block to block. Under the assumption that different exemplars would be less likely to have common components, RBC would predict that this condition would be slower than the Identical and Different View conditions but faster than a _Different Object_ control

condition with a new set of objects that required different names for every trial block. This was confirmed by Bartram.

For both different views of the same object, as well as different exemplars (subordinates) within a basic level category, RBC predicts that transfer would be based on the overlap in the components between the two views. The strong prediction would be that the same similarity function that predicted transfer between different orientations of the same object would also predict the transfer between different exemplars with the same name.

THE PERCEPTUAL BASIS OF BASIC LEVEL CATEGORIES

Consideration of the similarity relations among different exemplars with the same name raises the issue as to whether objects are most readily identified at a basic, as opposed to a subordinate or superordinate, level of description. The componential representations described here are representations of _specific_, subordinate objects, though their identification was always measured with a basic level name. Much of the research suggesting that objects are recognized at a basic level have used stimuli, often natural, in which the subordinate level had the same componential description as the basic level objects. Only small componential differences, or color or texture, distinguished the subordinate level objects. Thus distinguishing Asian elephants from African Elephants or Buicks from Oldsmobiles require fine discriminations for their verification. It is not at all surprising that with these cases basic level identification would be most rapid. On the other hand, many human-made categories, such as lamps, or some natural categories, such as dogs (which have been bred by humans), have members that have componential descriptions that differ considerably from one exemplar to another, as with a pole lamp vs a ginger jar table lamp, for example. The same is true of objects that are different from a prototype, as penguins or sport cars. With such instances, which unconfound the similarity between basic level and subordinate level objects, perceptual access should be at the subordinate (or instance) level, a result supported by a recent report by Jolicour, Gluck, & Kosslyn (1984).

It takes but a modest extension of the Componential Recovery Principle to problems of the similarity of objects. Simply put, similar objects will be those that have a high degree of overlap in their components and in the relations among these components. A similarity measure reflecting common and distinctive components (Tversky, 1977) may be adequate for describing the similarity among a pair of objects or between a given instance and its stored or expected representation, whatever their basic or subordinate level designation.

THE PERCEPTION OF NONRIGID OBJECTS

Many objects and creatures, such as people and telephones, have articulated joints that allow extension, rotation, and even separation of their components. There are two ways in which such objects can be accommodated by RBC. One possibility is that independent structural

descriptions are necessary for each sizable alteration in the arrangement of an object's components. For example, it may be necessary to establish a different structural description for figure 28a than 28d. If this was the case, then a priming paradigm might not reveal any priming between the two stimuli. Another possibility is that the relations among the components can include a range of

----
Insert Figure 28 About Here
----

possible values (Marr & Nishihara, 1979). In the limit, with a relation that allowed complete freedom for movement, the relation might simply be JOINED. Even that might be relaxed in the case of objects with separable parts, as with the handset and base of a telephone. In that case, it might be either that the relation is NEARBY or else different structural descriptions are necessary for attached and separable configurations. Empirical research needs to be done to determine if less restrictive relations, such as JOIN or NEARBY, have measurable perceptual consequences. It may be the case that the less restrictive the relation, the more difficult the identifiability of the object. Just as there appear to be canonical views of rigid objects (Palmer et al., 1981), there may be a canonical "configuration" for a nonrigid object. Thus, figure 28d might be identified as a woman more slowly than figure 28a.

CONCLUSION

To return to the analogy with speech perception made in the introduction of this article, the characterization of object perception that RBC provides bears close resemblance to many modern views of speech perception. In both cases, one has a modest set of primitives: In speech, the 55 or so phonemes that are sufficient to represent almost all words of all the languages on earth; in object perception, perhaps, a limited number of simple components. The ease by which we are able to code tens of thousands of words or objects may derive less from a capacity for making exceedingly fine physical discriminations than it does from allowing free combination of a modest number of categorized primitives.
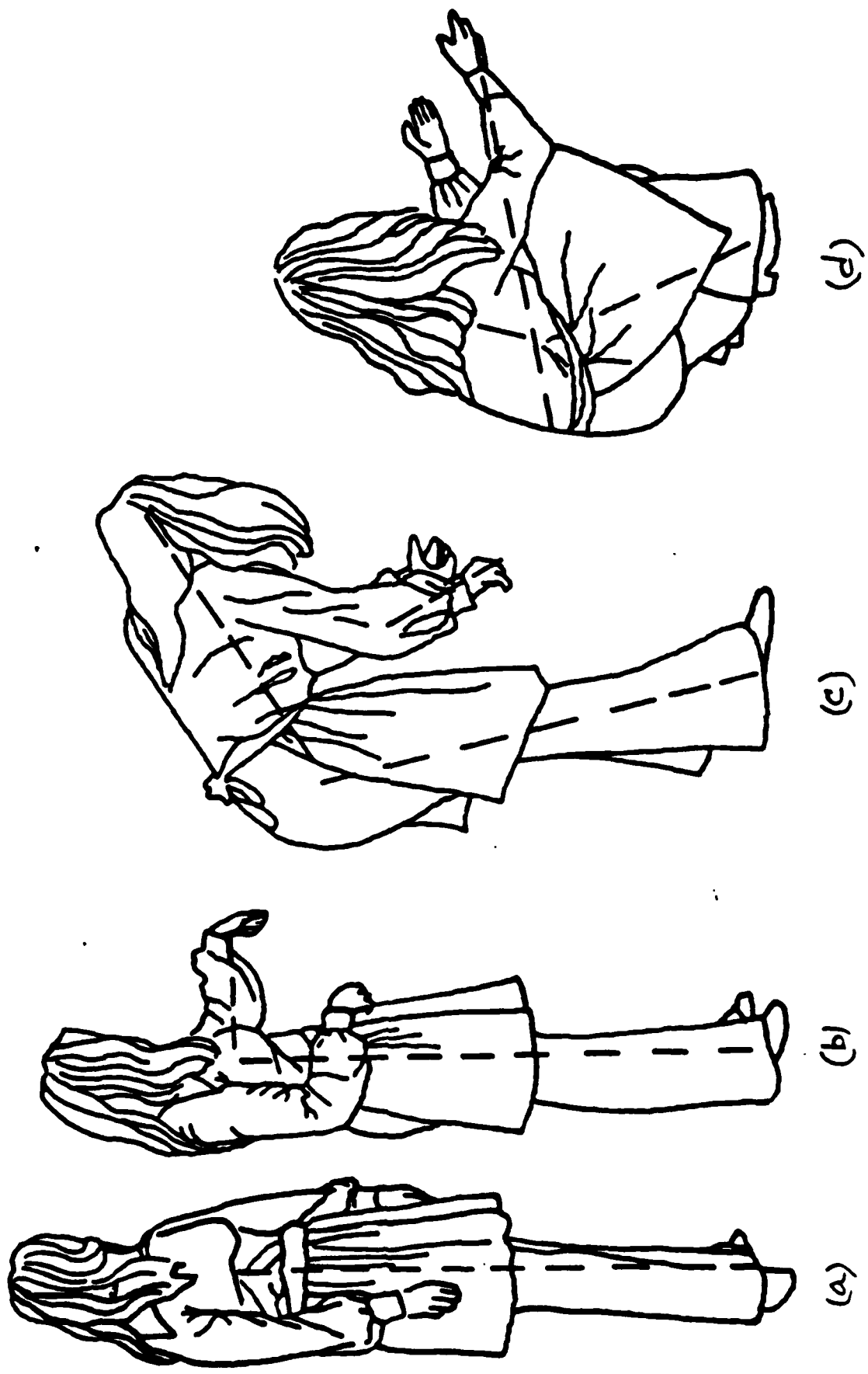
(a)   (b)   (c)   (d)

Figure 28. Four configurations of a nonrigid object.

References

Ballard, D., & Brown, C. M. (1982). Computer Vision. Englewood Cliffs, NJ: Prentice-Hall.

Barrow, H. G., & Tenenbaum, J. M. (1981). Interpreting line-drawings as three-dimensional surfaces. Artificial Intelligence, 17, 75-116.
Bartram, D. (1974). The role of visual and semantic codes in object naming. Cognitive Psychology, 6, 325-356.

Bartram, SD. (1976). Levels of coding in picture-picture comparison tasks. Memory & Cognition, 4, 593-602.

Beck, J., Prazdny, K., & Rosenfeld, A. (1983). A theory of textural segmentation. In Beck, J., Hope, B., & Rosenfeld, A. (Eds.) Human and Machine Vision. New York: Academic Press.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy, & J. R. Pomerantz (Eds.) Perceptual Organization. Hillsdale, N.J.: Erlbaum.

Biederman, I. & Blickle, T. (1985). The perception of degraded objects. Unpublished manuscript. State University of New York at Buffalo.

Biederman, I., Ju, G., & Clapper, J. (1985). The perception of partial objects. Unpublished manuscript. State Univesity of New York at Buffalo.

Biederman, I., & Ju, G., (1985). A comparison of the perception of line drawings and colored photography. Unpublished manuscript. State University of New York at Buffalo.

Biederman, I., Ju., J., & Beiring, E. (1985). A comparison of the perception of partial vs degraded objects. Unpublished manuscript. State University of New York at Buffalo.

Biederman, I., & Lloyd, M. Experimental studies of transfer across different object views and exemplars. Unpublished manuscript. State University of New York at Buffalo.

Binford, T. O. (1981). Inferring surfaces from images. Artificial Intelligence, 17, 205-244.

Binford, T. O. (1971). Visual perception by computer. IEEE Systems Science and Cybernetics Conference, Miami, December.

Brady, M. (1983). Criteria for the representations of shape. In J. Beck, B. Hope, & A. Rosenfeld (Eds.) Human and Machine Vision. New York: Academic Press.

Brooks, R. A. (1981). Symbolic reasoning among 3-D models and 2-D images. Artificial Intelligence, 17, 205-244.

Cary, S. (1978) The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.) Linguistic Theory and Psychological Reality. Cambridge, Mass: MIT Press.

Cezanne, P. (1904/1941). Letter to Emile Bernard. In J. Rewald (Ed.) Paul Cezanne's Letters (Translated by M. Kay). B. Cassairrer: London.

Chakravarty, I. (1979). A generalized line and junction labeling scheme with applications to scene analysis. IEEE Transactions. PAMI, April, 202-205.

Checkosky, S. F., & Whitlock, D. (1973). Effects of pattern goodness on recognition time in a memory search task. Journal of Experimental Psychology, 100, 341-348.

Coss, R. G. (1979). Delayed plasticity of an instinct: Recognition and avoidance of 2 facing eyes by the jewel fish. Developmental Psychobiology, 12, 335-345.

Egeth, H., & Pachella, R. (1969). Multidimensional stimulus identification. Perception & Psychophysics, 5, 341-346.

Fildes, B. N., & Triggs, T. J. (1985). The effect of changes in curve geometry on magnitude estimates of road-like perspective curvature. Perception & Psychophysics, 37, 218-224.

Garner, W. R. (1974). The processing of information and structure. New York: Wiley.

Guzman, A. (1971). Analysis of curved line drawings using context and global information. Machine Intelligence 6. Edinburgh: Edinburgh University Press.

Hildebrandt, K. A. (1982). The role of physical appearance in infant and child development. In H. E. Fitzgeral, E. Lester, & M. Youngman (Eds.) uTheory and Research in Behavioral Pediatrics, Vol. 1. New York: Plenum.

Hildebrandt, K. A., & Fitzgerald, H. E. (1983). The infant's physical attractiveness: Its effect on bonding adn attachment. Infant Mental Health Journal, 4, 3-12.

Hochberg, J. E. (1978). Perception, 2nd Ed. Englewood Cliffs, N.J.: Prentice-Hall.

Hoffman, D. D. & Richards, W. (1984). Parts of recognition. Cognition, 18, 65-96.

Humphreys, G. W. (1983). Reference frames and shape perception. Cognitive Psychology, 15, 151-196.

Ittleson, W. H. (1952). The Ames Demonstrations In Perception. New York: Hafner.

Jolicoeur, Gluck, & Kosslyn (1984). Picture and names: Making the connection. Cognitive Psychology, 16, 243-275.

Juleaz, B. (1981). Textons, the elements of texture perception, and their interaction. Nature, 290, 91-97.

Kanade, T. (1981). Recovery of the three-dimensional shape of an object form a single view. Artificial Intelligence, 17, 409-460.

King, M., Meyer, G. E., Tangney, J., & Biederman, I. (1976). Shape constancy and a perceptual bias towards symmetry. Perception & Psychophysics, 19, 129-136.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. Journal of Verbal Learning and Verbal Behavior, 23.

Lowe, D. (1984). Perceptual organization and visual recognition. Unpublished doctoral dissertation, Department of Computer Science, Stanford University.

Mark, L. S., & Todd, J. T. (1985). Describing perception information about human growth in terms of geometric invariants. Perception & Psychophysics, 37, 249-256.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of three dimensional shapes. Proceedings of the Royal Society of London B., 200, 269-294.

Marslen-Wilson, W. (1980). Optimal Efficeincy in human speech processing. Unpublished manuscript, Max Planck Institut fur Psycholinguistik, Nijmegen, The Netherlands, 1980.

McClelland, J. L., & Rumelhart, D. E.   (1981).   An interactive
    activation model of context effects in letter perception, Part I:
    An account of basic findings. Psychological Review, 375-407.
Miller, G. A. (1956).   The magical number seven, plus or minus two:
    Some limits on our capacity for processing information.
    Psychological Review, 63, 81-97.
Miller, G. A. (1977). Spontaneous Apprentices: Children and Language.
    New York: Seabury.
Neisser, U. (1963). Decision time without reaction time: Experiments
    in visual scanning. American Journal of Psychology, 76, 376-385.
Oldfield, R. C., & Wingfield, A.   (1965).   Response latencies in
    naming objects. Quarterly Journal of Experimental Psychology, 17,
    273-281.
Oldfield, R. C. (1966).   Things, words, and the Brain. Quarterly
    Journal of Experimental Psychology, 18, 340-353.
Palmer, S, E.   (1980). What makes triangles point: Local and global
    effects in configurations of ambiguous triangles. Cognitive
    Psychology, 12, 285-305.
Palmer, S., Rosch, E., & Chase, P. (1981).   Canonical perspective and
    the perception of objects, Attention & Performance IX, Long, J., &
    Baddeley, A. (Eds.).   Hillsdale, N.J.:   Erlbaum, 1981.
Penrose, L. S., & Penrose, R. (1958).   Impossible objects: A special
    type of illusion. British Journal of Psychology, 49, 31-33.
Perkins, D. N. (1983).   Why the human perceiver is a bad machine.   In
    Beck, J., Hope, B., & Rosenfeld, A. (Eds.) Human and Machine Vision.
    New York: Academic Press.
Perkins, D. N., & Deregowski, J. (1983).   A cross-cultural comparison
    of the use of a Gestalt peceptual strategy. Perception,
Pomerantz, J. R. (1978).   Pattern and speed of encoding. Memory &
    Cognition, 5, 235-241.
Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977).   Perception
    of wholes and their component parts: Some configural superiority
    effects. Journal of Experimental Psychology: Human Perception and
    Performance, 3, 422-435.
Rock, I.   (1984). Perception. New York: W. H. Freeman.
Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem.
    (1976). Basic objects in natural categories. Cognitive Psychology,
    8, 382-439.
Rosenthal, S.(1984). The PF474. Byte, 9, 247-256.
Ryan T., & Schwartz, C. (1956).   Speed of perception as a function of
    mode of representation. American Journal of Psychology, 69, 60-69.
Sugihara, K.   (1984).   An algebraic approach to shape-from-image
    problems. Artificial Intelligence, 23, 59-95.
Shepard, R. N., & Metzler, J. (1971). Mental rotation of three
    dimensional objects. Science, 171, 701-703.
Templin, M. C. (1957).   Certain language skills in children: Their
    development and interrelationship. Minneapolis: University of
    Minnesota Press.
Treisman, A. (1982).   Perceptual grouping and attention in visual
    search for objects. Journal of Experimental Psychology: Human
    Perception and Performance, 8, 194-214.
Treisman, A., & Gelade, G.   (1980).   A feature integration theory of
    attention. Cognitive Psychology, 12, 97-136.

Treisman, A. (1982).   Perceptual grouping and attention in the visual
   search for features and for objects.   Journal of Experimental
   Psychology: Human Perception and Performance, 8, 194-214.
Trivers, R.   (1985).   Social Evolution.   Menlo Park:
   Benjamin/Cummings.
Tversky, A. (1977). Features of similarity.   Psychological Review, 84,
   327-352.
Tversky, B., & Hemenway, K.   (1984).   Objects, parts, and categories.
   Journal of Experimental Psychology: General, 113, 169-193.
Ullman, S. (1983).   Visual routines.   A.I. Memo No. 723.
   Massachusetts Institute of Technology, Artificial Intelligence
   Laboratory.
Virsu, V. (1971a).   Tendencies to eyemovements and misperception of
   curvature, direction, and length.   Perception & Psychophysics, 9,
   65-72.
Virsu, V. (1971b).   Underestimation of curvature and task dependence
   in visual perception of form.   Perception & Psychophysics, 9,
   339-342.
Winston, P. A. (1975).   Learning structural descriptions from
   examples.   In Winston, P. H. (Ed.) The Psychology of Computer
   Vision.   New York: McGraw-Hill.
Witkin, A. P., & Tennenbaum, J. M. (1983). On the role of structure in
   vision.   In Beck, J., Hope, B., & Rosenfeld, A. (Eds.) Human and
   Machine Vision.   New York: Academic Press.

AFOSR Contract (F49620-83-C-0086) Progress Report
HUMAN·INFCRMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES
I. BIEDERMAN

APPENDIX A

Summary Log of Experimental Effort on Object Perception

Experiment	No of Trials	No of Subjects
I. Effect of Number of Components in complete and partial objects.		
1. Component Variation: Object shown Full object Complexity: 2, 3, 4, 6, & 9 comps. Variation (2 to full) in number of components displayed. Order: Start full or w. two components.	120	120
2.* Balanced for No of components and trial order. Name and picture vs. name only familiarity.	99	48
3. Familiarity of components (within Ss) Familiarization (names vs. no names)	99	16
4. Slide duration: 100, 200, 750 msec. Speed not stressed. Error rates only.	99	48
II. Color Slides vs. Line Drawings		
1. High intensity.	52	24
2. High Intensity. Better slides.	60	30
3. Low intensity	60	30
4.* No mask. Low intensity.	60	30
III. Degradation (Contour Deletion). Recoverable vs Nonrecoverable components		
1.* Bet groups 100, 200, 750 msec presentation	70	18
2. W/in groups presentation duration.	70	9
3.* 5000 msec	70	6
4. Verification (bet groups) 100, 200, 750 msec	80	72
5.* 25,45,65% delet at vertex or midsegment 100-700ms	108	30
6.* Removal of components vs contour deletion 65-200ms	108	30

AFOSR Contract (F49620-83-C-0086) Progress Report
HUMAN INFORMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES
I. BIEDERMAN

A-2

Appendix A (Continued)

Summary Log of Experimental Effort on Object Perception

Experiment	No. of Trials	No. of Subjects
IV. Transfer		
1. Rotation: Envelope vs Components altered Same-Diff Same-Diff 0°-225°, Large vs Small z objects	80	120
2. Same vs Different View vs Exemplar Front Left and Right (45° & 315°)	72	64
3. Familiarity. 1st exposure: 100, 300, & 1000 msec. 2nd exposure: 150 msec. Same-Diff view & Exemplar	72	48
4. Familiarity: None, Name, Name+300 msec picture 65 msec exposure on 2nd trial.	72	24
Total No of Usable Subjects		776

Total No of Usable Trials (No. of Subjects X Trials)   64,278

Note.--Data for studies designated with an asterisk are repesented in the
progress report. Unless otherwise ntoed, all studies involved the
identification of object slides at brief exposure durations.

AFOSR Contract (F4962083C0086) Progress Report
HUMAN INFORMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES
I. BIEDERMAN

APPENDIX B

Papers:

Biederman, I. (1985). Human Image Understanding: Recent Research and a Theory. Computer Vision, Graphics, and Image Processing, 32, In Press.

Biederman, I. (1986). The Perceptual Recognition of Objects and Scenes. In G. H. Bower (Ed.) The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 21. New York: Academic Press.

Biederman, I. (1985). Recognition-by-Components: A Theory of Image Interpretation. Ms. submitted to Psychological Review.

Biederman, I., Blickle, T., Teitelbaum, R. C., & Klatsky, G. J. (1985). Object identification in multi-object, non-scene displays. Ms. sumitted to Journal of Experimental Psychology: Human Perception and Performance.

Walters, D. H., & Biederman, I. The combination of spatial frequency and orientation is not effortlessly perceived. (1985). Submitted to Vision Research.

Biederman, I. & Blickle, T. (1985). The perception of degraded objects. Unpublished manuscript. State University of New York at Buffalo.

Biederman, I., Ju, G., & Clapper, J. (1985). The perception of partial objects. Unpublished manuscript. State University of New York at Buffalo.

Biederman, I., & Ju, G., (1985). A comparison of the perception of line drawings and colored photography. Unpublished manuscript. State University of New York at Buffalo.

Biederman, I., Ju., J., & Beiring, E. (1985). A comparison of the perception of partial vs degraded objects. Unpublished manuscript. State University of New York at Buffalo.

Biederman, I., & Lloyd, M. (1985). Experimental studies of transfer across different object views and exemplars. Manuscript in preparation. State University of New York at Buffalo.

Biederman, I., Malcus, L., & Mezzanotte, R. J. (1985). The detection of relational violations in very brief exposures of scenes: Evidence for rapid access to semantic relations. Manuscript in preparation.

AFOSR Contract (F4962083C0086) Progress Report
HUMAN INFORMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES
I. BIEDERMAN

APPENDIX B (Continued)

Biederman, I., Tetewsky, S., & Mezzanotte, R. J. (1985).
Unidentifiable objects can become identifiable when brought together
to form a scene: Evidence for scene emergent features. Manuscript
in preparation.

Biederman, I., & Shiffrar, M. (1985). Sex-typing day-old chicks: A
case study of a difficult perceptual learning task. Manuscript in
preparation.

Biederman, I., & Fisher, B. (1985). Spatial attention. Manuscript
in preparation.

Papers at Scientific Meetings:

Walters, D., Biederman, I., & Weisstein, N. The combination of
spatial frequency and orientation is not effortlessly peceived.
Paper presented at the ARVO Meetings, Sarasota, Florida: May,
1983.

Biederman, I. Recognition-by-Components: A theory of image
interpretation. Paper presented at the Air Force Office of
Scientific Research Review of Basic Research in Visual Information
Processing, Sarasota, Florida: May, 1984.

Biederman, I. Recognition-by-Components: Explorations of image
interpretation. Invited address at the Workshop for Human and
Machine Vision, International Conference on Pattern Recognition,
Montreal, Canada: August, 1984.

Biederman, I. Recognition-by-Components: A theory of image
interpretation. Paper to be presented at the Meetings of The
Psychonomic Society, San Antonio, Texas: November, 1984.

Ju, G., Biederman, I., & Clapper, J. Recognition-by-Components: The
minimum number of parts needed for speeded recognition. Paper
presented at the Meetings of the Eastern Psychological Association,
Boston, Mass: April, 1985.

Blickle, T., & Biederman, I. Recognition-by-Components: A principle
of object degradation. Paper presented at the Meetings of the
Eastern Psychological Association, Boston, Mass: April, 1985.

Biederman, I., Blickle, T., & Ju, G. Studies in the speeded
recognition of objects: Evidence for a componential theory of
object perception. Paper to be presented at the Meetings of the
Psychonomic Society, Boston, Mass: November, 1985.

AFOSR Contract (F4962083C0086) Progress Report
HUMAN INFORMATION PROCESSING OF TARGETS AND REAL-WORLD SCENES
I. BIEDERMAN

APPENDIX B (Continued)

Invited Colloquiua:

Rutgers University
Emory University
Naval Aerospace Medical Corps, Pensacola, Fl.
Temple University
Air Force Office of Scientific Research, Bolling, AFB, D.C.
Human Resources Laboratory, Williams AFB, AZ
University of Illinois at Chicago
University of Delaware
Center for Adaptive Systems, Boston University
University of California, Santa Cruz.
University of California, Berkeley (Psychology Department)
NASA Ames
Clairmont Graduate Schools
University of California, Santa Barbara, Department of Psychology
University of California, Santa Barbara, Computer Science
Stanford University (Psychology Department)
University of California, Davis
Human Resources Laboratory, Williams, AFB, AZ
University of California, Oxyopia
Massachusetts Institute of Technology
Stanford University (Artificial Intelligence Laboratory)
Bucknell University

END

10-86

DTIC