SECURITY CLASSIFICATION OF THIS PAGE

AD-A281 336

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
F49620-92-J-0169

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 94 0413

6a. NAME OF PERFORMING ORGANIZATION
Yale University

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NL

6c. ADDRESS (City, State and ZIP Code)
12 Prospect Place
New Haven, CT 06511

7b. ADDRESS (City, State and ZIP Code)
110 Duncan Ave Suite B115
Bolling AFB, DC 20332-0001

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR/PK

8b. OFFICE SYMBOL (If applicable)
NL

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F49620-92-J-0169

8c. ADDRESS (City, State and ZIP Code)
110 Duncan Avenue
Suite B115
Bolling AFB, DC 20332-0001

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313/AS	2313/BS	

11. TITLE (Include Security Classification)
Representation of Shape in Object Recognition and Long-Term Visual Memory

12. PERSONAL AUTHOR(S)
Tarr, Michael J.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Annual	FROM 1-15-/93 TO 1/14/94	94/4/5	19

16. SUPPLEMENTARY NOTATION

17 COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Object Representation, Object Recognition
Visual Cognition

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

A wide range of psychophysical experiments investigating the mechanisms and representations underlying human object recognition have been conducted. In particular, the focus of this research has been an approach in which object recognition is mediated by at least two systems, one based on an explicit qualitative encoding of viewpoint-invariant features and one based on a metrically specific encoding of shape. Within the literature, this dichotomy has been most often associated with measures of the effect of viewpoint on recognition performance. For the most part, the common assumption has been that viewpoint-dependent patterns of performance are the signature of one recognition mechanism, while viewpoint-invariant patterns of performance are the signature of another recognition mechanism. Reinforcing this distinction, viewpoint-dependent mechanisms have been more broadly associated with metrically-specific representations sensitive to a range of image-based properties, for example, size, handedness, color, or illumination, while viewpoint-invariant mechanisms have been more broadly associated with coarsely-coded representations insensitive to image-based properties. To this point, the majority of work on this project has focused only on the former in recognition tasks where perceivers must discriminate between visually similar objects (e.g., a within-category or subordinate-level judgment). During the past year we have continued this line of research, but have extended our approach to include recognition tasks using objects that are relatively dissimilar in that they may differentiated by a small number of

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. John Tangney

22b. TELEPHONE NUMBER (Include Area Code)
(202) 767-5021

22c. OFFICE SYMBOL
NL

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

SECURITY CLASSIFICATION OF THIS PAGE

17 JUN 1994

AFOSR-TR- 94 0413

AFOSR Grant No. F49620-92-J-0169

REPRESENTATIONS OF SHAPE IN OBJECT RECOGNITION AND LONG-TERM VISUAL MEMORY

Michael J. Tarr
Yale University
Department of Psychology
PO Box 208205 Yale Station
New Haven, CT 06520-8205

April 5, 1994

Annual Technical Report for Period 15 January 1993 - 14 January 1994

Prepared for

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Building 410
Bolling AFB, DC 20332-6448

ANNUAL TECHNICAL REPORT

REPRESENTATIONS OF SHAPE IN OBJECT RECOGNITION AND LONG-TERM VISUAL MEMORY

Michael J. Tarr
Yale University

AFOSR Grant #F49620-92-J-0169

ABSTRACT

A wide range of psychophysical experiments investigating the mechanisms and representations underlying human object recognition have been conducted. In particular, the focus of this research has been an approach in which object recognition is mediated by at least two systems, one based on an explicit qualitative encoding of viewpoint-invariant features and one based on a metrically specific encoding of shape. Within the literature, this dichotomy has been most often associated with measures of the effect of viewpoint on recognition performance. For the most part, the common assumption has been that viewpoint-dependent patterns of performance are the signature of one recognition mechanism, while viewpoint-invariant patterns of performance are the signature of another recognition mechanism. Reinforcing this distinction, viewpoint-dependent mechanisms have been more broadly associated with metrically-specific representations sensitive to a range of image-based properties, for example, size, handedness, color, or illumination, while viewpoint-invariant mechanisms have been more broadly associated with coarsely-coded representations insensitive to image-based properties. To this point, the majority of work on this project has focused only on the former in recognition tasks where perceivers must discriminate between visually similar objects (e.g., a within-category or subordinate-level judgment). During the past year we have continued this line of research, but have extended our approach to include recognition tasks using objects that are relatively dissimilar in that they may differentiated by a small number of qualitatively different parts (e.g., a between-category or entry-level judgment). One track during the past year has been to investigate the concurrent acquisition of viewpoint-dependent and viewpoint-invariant object representations. Results indicate that perceivers learn both types of representations regardless of current recognition task and may subsequently employ either type of mechanism depending on its appropriateness to a given task. A second track has continued with investigations into the specifics of viewpoint-dependent recognition mechanisms. Results here indicate that these mechanisms are specific to image-based properties other than viewpoint, for example, direction of illumination. Moreover, the range of paradigms used to assess the operation of viewpoint-dependent mechanisms has been expanded to include explicit tasks such as same-different judgments and implicit tasks such as priming. One implication of this work has been that performance is not determined by the task *per se*, but rather by the information that is relevant to accomplishing a given task. We have also begun to explore face recognition within the framework of this being the most extreme instance of within-category recognition, thereby providing the most pure measure of viewpoint-dependent mechanisms. To this end, we have systematically explored the effects of misorientation on both recognition memory for faces and face naming, finding that patterns of performance are similar to other within-category discriminations. We have also developed a complex contrast stimulus set for faces. Unlike other contrast sets that have been used (i.e., houses), these objects are both novel and designed to have similar parts in similar configurations so that objects may only be individuated by metric differences in shape. An extensive series of studies are planned to assess recognition performance with these objects as compared to both faces and common familiar objects. In contrast to this work, a final track has focused on entry-level recognition in which objects may be discriminated on the basis of qualitative differences among one or more parts. Results support two conclusions at odds with the predominant viewpoint-invariant theory of entry-level recognition: recognition of single three-dimensional volumes is systematically viewpoint dependent and recognition of multi-part objects that may be differentiated by configurations of small numbers of parts is also viewpoint dependent, with some changes in what constitutes a view mediated by qualitative changes in image structure rather than part configurations.

STATEMENT OF WORK

Work has continued on investigating the mechanisms and underlying representations implicated in human visual cognition, and in particular, in object recognition and visual memory. In addition to the completion of several projects outlined in last year's report (i.e., the encoding of spatial relations and the concurrent encoding of view-based and object-based representations), several new projects have begun that employ a wide range of psychophysical paradigms and stimulus sets. Along the dimension of task, these studies may characterized as extending the novel-object-naming approach introduced by Tarr and Pinker (1989), as utilizing priming or other implicit memory tasks, or as utilizing one of several different explicit memory tasks, including same-different judgments and recognition memory. Along the dimension of stimulus content, these studies may be characterized as using photorealistically rendered novel three-dimensional objects that share similar parts in similar spatial configurations, as using photorealistically rendered novel three-dimensional objects that are composed of qualitatively different parts, or using photorealistically rendered familiar common objects that may be members of the same or different entry-level categories. Additionally, work has begun on assessing the relationship of face recognition to the recognition of other objects. To this end we have applied several paradigms commonly used in object recognition to faces and have developed an extensive set of contrast stimuli that mimic the perceptual category organization found in faces. All of these projects are designed to address specific issues in recognition, and, in particular, to extend what is known about the mechanisms and representations brought to bear under varying conditions.

STATUS

1. Lexical and Perceptual Encoding of Spatial Relations

William Hayward (a graduate student at Yale) and I have completed series of experiments to investigate the nature of *qualitative* spatial relations encoded between objects in a scene (or between parts of an object). Such relations are an essential element of many structural-description theories of object representation (i.e., Hummel & Biederman, 1992). Specifically, we have examined the possibility that the restricted meanings of spatial prepositions used in language reflect a similar qualitative encoding of spatial relations in the visual representation system. As detailed in the attached paper (accepted pending revision to *Cognition*), a series of four experiments indicate that linguistic descriptions and the visual encoding of space share common structures for the relations "above" "below" "left" and "right". Across four experiments objects were presented in a scene where one, the reference object, always appeared in the center, and the other, the figural object, appeared in one of many positions on a 7x7 grid surrounding the reference object. Results from the first two experiments indicate that perceivers have a preference to apply spatial terms in a qualitative manner — for example, applying "above" when the figural object is directly vertical relative to the referent. Secondly, while the same spatial terms certainly apply to other relations between objects, they do so in a gradient that decreases in both frequency of application and assessed appropriateness with distance from the preferred axis.

A similar pattern was obtained in two experiments that employed perceptual judgments with scenes configured as in the first two studies. One study required subjects to use spatial memory to recall the position of the figural object relative to the reference object. A second study required subjects to judge whether the figural object was in the same location relative to the reference object in two sequential frames (which shifted randomly in screen position so that subjects could not simply note the absolute position of the figural object between frames). In both studies performance was highest at spatial positions where the figural object was axially

aligned with the reference object. Such results suggest that there is a correspondence between qualitative spatial representations found in the visual system and the categorical form referred to in language (i.e., we refer to objects being simply *above* rather than precisely how far above). Given this correspondence, we may begin to explore the specifics of spatial relations within objects using both linguistic and non-linguistic tasks. For example, one paradigm may employ sequentially presented images of similar objects where the relations between parts vary. While the magnitude of quantitative changes in spatial relations are expected to influence performance, qualitative changes are predicted to have a far greater impact on performance. This and related paradigms may be used to assess the qualitative boundaries of part relations within objects, as well as possible similarities to linguistic descriptions of such relations.

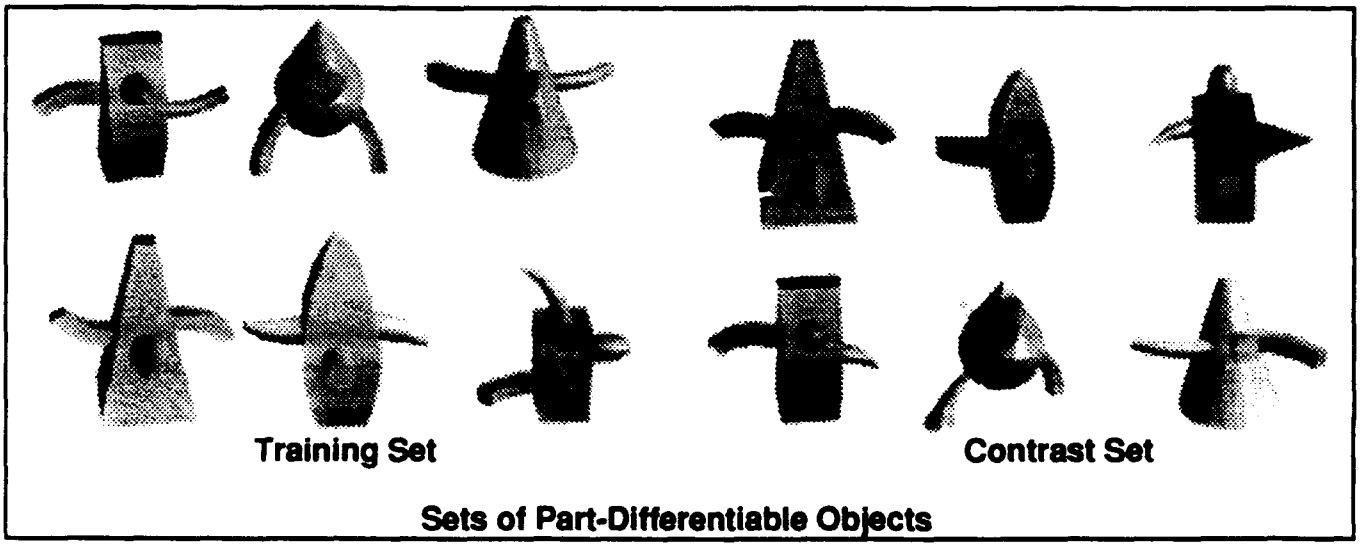
Presentations & Papers:

Hayward, W. G., & Tarr, M. J. Spatial language and spatial representation. *65th Annual Meeting of the Eastern Psychological Association*, Providence, RI, April 15-17, 1994.

Hayward, W. G., & Tarr, M. J. Spatial language and spatial representation. Accepted pending revisions to *Cognition*.

2. Concurrent encoding of viewpoint-dependent and viewpoint-invariant object representations

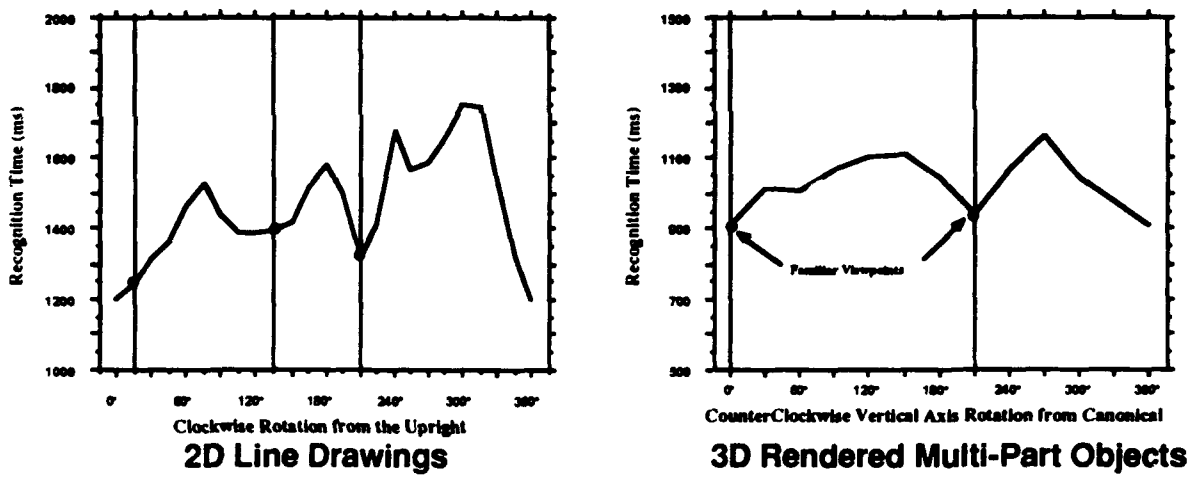
Current theories of object recognition have posited both viewpoint-dependent and viewpoint-invariant modes of object representation. However, it is still unclear as to what conditions determine how perceptual mechanisms apply such representations under different contexts in learning and recognition. We have completed a project in which we have demonstrated that regardless of the role of viewpoint during initial encoding, subjects apparently encode *both* types of representations. Specifically, subjects were initially taught a set of objects, the training set, that could be immediately recognized equally well at all viewpoints: in one case 2D line drawings similar to those used in Tarr and Pinker (1990) and in the other case 3D part-differentiable objects (where a small number of qualitatively different parts is sufficient to discriminate one object from all others in the set).



Sets of Part-Differentiable Objects

After familiarization, subjects were given extensive practice recognizing the objects from a select set of viewpoints generated by rotations in the image-plane or in depth (depending on the stimulus set). As predicted, in both instances, recognition performance was immediately equivalent at all tested viewpoints, indicating that viewpoint-invariant mechanisms and representations were employed during this phase. Following practice at recognition across several days, subjects were taught an equivalent number of new objects, referred to as the contrast set. The critical manipulation is that combined with the objects in the training set, no single object could be differentiated by a qualitative description of parts (as in Biederman's, recognition-by-components theory, Biederman & Gerhardstein, 1993) or by simple one-dimensional ordering of parts (see Tarr & Pinker, 1990). To assess the impact of including these new objects, additional unfamiliar viewpoints were also added during this phase. Two crucial predictions were made: (1) introducing the contrast set would result in a shift to viewpoint-dependent recognition mechanisms; (2) viewpoint-dependent effects would be systematically related to the nearest previously seen viewpoint despite the previous lack of effects of viewpoint.

As shown in the two graphs below, in the final phase of each experiment, both predictions were obtained. For both 2D rotated in the image-plane and 3D objects rotated in depth, there is now a significant effect of viewpoint on naming time. Crucially, this pattern is systematic to the nearest familiar viewpoint, indicating that subjects did encode a viewpoint-specific object representation at each observed viewpoint.



Recognition of Familiar Objects Following Training in Select Viewpoints

A control experiment verified that these viewpoint-dependent effects are not simply due to the addition of viewpoints and objects. This study employed the identical 2D training set used in the previous experiment, but employed a contrast set that did not require subjects to rely on complex part relations across more than a single dimension. Under such conditions it was predicted that, despite the introduction of new viewpoints and objects, viewpoint-dependent patterns would not be obtained. Results for the familiar training objects confirmed this: no systematic pattern of response times across orientation was observed. Overall these results indicate that there is no "default" recognition mechanism. Rather the visual system apparently encodes at least two distinct types of object representations, one viewpoint invariant and one

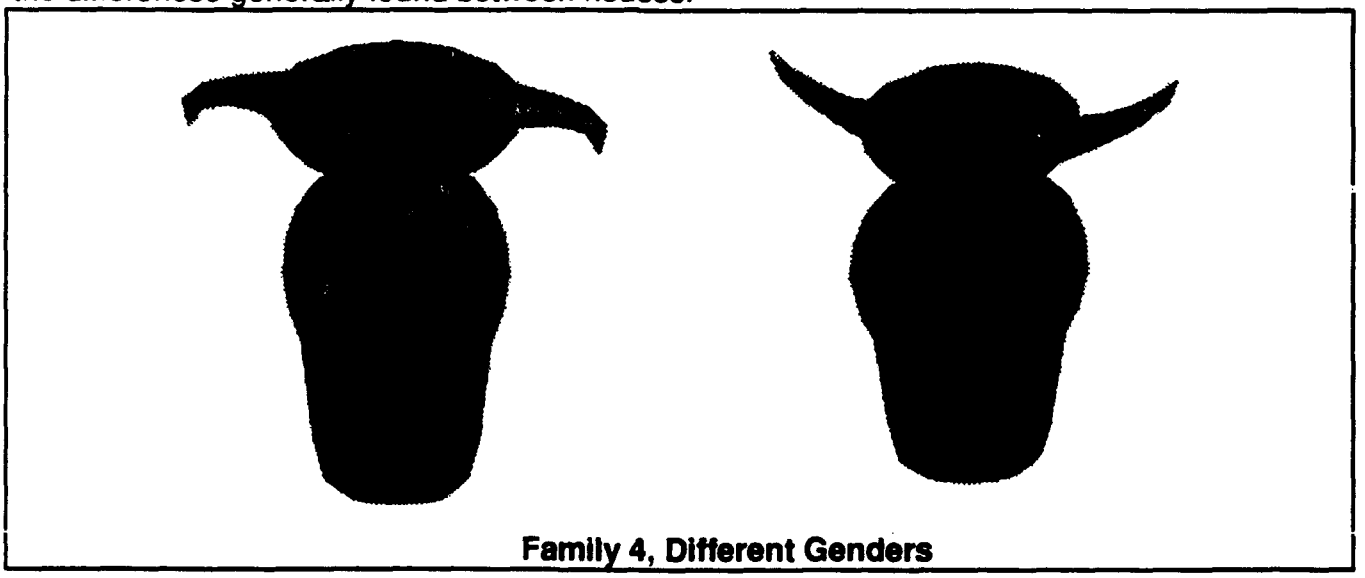
viewpoint dependent, and utilizes each along with appropriate recognition mechanisms in accordance with the perceptual information necessary for accomplishing a given task.

Presentations & Papers:

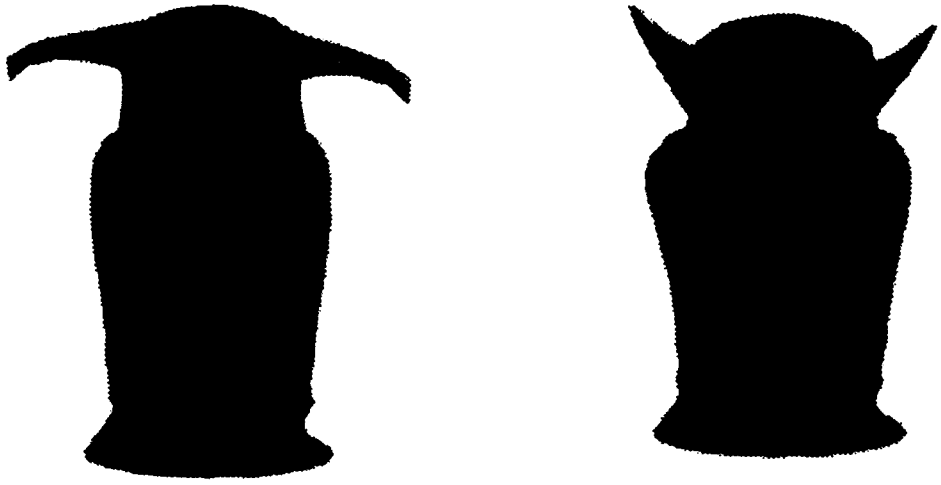
Tarr, M. J., & Chawarski, M. C. (1993). The concurrent encoding of object-based and view-based object representations. Presented at *The 34th Annual Meeting of the Psychonomic Society*, November 5-7. Washington, DC.
Manuscript in preparation.

3. Holistic processing in face and non-face stimuli.

Recent reports by Tanaka and Farah (unpublished, partially reported in Farah, 1992) indicate that face recognition may be *holistic*. Specifically, they tested recognition memory for single facial parts (i.e., a nose) in isolation, in a transformed same-face context (eyes moved relative to study), and in the original face context and found that recognition was best in the original context, poorer when the eyes were moved, and poorest when parts were presented alone (foils were always an unfamiliar part in the identical context). This finding suggests that representations of faces encode all facial features in a single, integrated form, e.g., a holistic representation. In particular, the fact the recognition of parts of a familiar face are better recognized within the task-irrelevant context of the entire face supports this claim. Moreover, it is likely that holistic representations share properties with the viewpoint-dependent image-based representations implicated in the recognition of common objects in that both are posited to preserve metric specificity and image properties from a given viewpoint. What is unclear whether the mechanisms that mediate face recognition are simply similar to or actually the same as the mechanisms that mediate other subordinate-level judgments (as discussed above). To address this issue, Tanaka conducted a control experiment in which houses were used rather than faces. Here subjects recognized single parts, i.e., doors, in isolation, with transformed positions of windows, or in the original training configuration. In this experiment, no evidence was found for holistic representations, rather recognition performance was equivalent across conditions. One possible conclusion is that holistic representations are exclusive to face recognition. Alternatively, it may be that houses offer an inadequate contrast set for faces in that the differences in shape between different faces are somewhat subtler than the differences generally found between houses.



Family 4, Different Genders



Family 5, Different Genders

In collaboration with Tanaka, we have begun a project to address these two possibilities, and more generally, to address the problem of identifying an adequate set of contrast stimuli for faces. In particular, faces are drawn from a set of highly similar objects that share common parts and spatial configurations. However, faces are also somewhat unique in that they may be organized into subclasses such as gender or family. Our goal then, was to develop a novel set of realistic three-dimensional stimuli that enforced these two constraints. The initial result of this work has been the generation of 60 novel 3D objects that share similar parts and configurations, but may also be subdivided into shape defined subclasses corresponding to family and gender (which cuts across families). Sample objects are displayed above. Note that we have arbitrarily selected the displayed texture and viewpoint, but that the objects may be displayed from any viewpoint with a wide range of textures, colors, and illuminations. After familiarization with such objects and name-object training for a subset, subjects will perform a 2AFC recognition task on single parts similar to that used by Tanaka. Again the critical manipulation will be whether the part in question appears in isolation, in a transformed context, or in the original training context. Here, given the structure of the stimulus set, it is predicted that evidence for holistic representations will be found (e.g., better recognition performance in the original training context). Such a result would support the hypothesis that holistic representations are not exclusive to faces, but rather, that such representations are employed in recognition discriminations where precise metric information is essential (a claim consistent with the results of the experiments discussed in the previous section).

While our initial intention was to use these objects in a paradigm similar to that used by Tanaka, we have come to realize these stimuli have great potential for serving as controls for faces in many domains. For example, we are currently exploring the use of these in assessing general recognition performance in lesion patients with face recognition deficits, in single cell recording studies of macaque monkeys that have previously been trained with face stimuli, in normal adults performing a variety of face recognition tasks (see #4 below), and in social stereotyping situations. In order to use the stimuli in these studies it is first necessary that we verify that the objects are naturally categorized into the subsets intended in their design. Several different sorting tasks, perceptual confusion judgments, and explicit comparisons will be used for this purpose. Additionally, one other factor that makes faces such an atypical stimulus set is the degree of expertise humans have with faces as compared to most other classes of objects. To address this issue, we are developing several methods for training

subjects to be "experts" at discriminating and identifying these novel objects. Such studies will also provide data on acquisition of expertise with entirely novel objects (as compared with the familiar objects used by Diamond & Carey, 1986).

Presentations & Papers:

Organized and chaired symposium on Complex Object Recognition at the *65th Annual Meeting of the Eastern Psychological Association*, Providence, RI, April 15-17, 1994. Participants: J. Tanaka, B. Gibson, S. Carey, & M. Farah.

4. Recognition of faces in implicit and explicit tasks

The inversion effect, poorer recognition performance for faces misoriented 180° in the image-plane, is often cited as one of the properties that makes face recognition unlike "normal" object recognition. However, several recent studies using both familiar (Jolicoeur, 1985) and novel (Tarr & Pinker, 1989) objects have demonstrated an inversion effect for the recognition of non-face stimuli. Notable in these findings is the *systematic* pattern of performance, with increasingly poorer performance as objects are misoriented farther from the canonical upright (with sometimes a small "dip" at precisely 180°, see McMullen & Jolicoeur, 1992). One possibility is that this pattern is indicative of a mental transformation in which the stimulus object is normalized to a canonical viewpoint for purposes of matching in recognition. Thus, it is possible that the facial inversion effect is simply another instance of this normalization effect, and as such is not indicative of any specialized face recognition mechanism. However, because, for the most part, face recognition has only been tested at the upright and at complete inversion, it is currently unknown whether the same systematic pattern of performance is obtained for face stimuli or whether there is a categorical pattern in which faces are recognized equally well up to some threshold (as might be predicted if faces are encoded as structural descriptions). A second issue addressed by this project is the recent proposed dichotomy between the representations mediating implicit and explicit memory, with structural-descriptions underlying the former and image-like episodic representations underlying the latter (Cooper & Schacter, 1992). In particular, if the same systematic patterns of performance are found for both implicit and explicit tasks, one inference may be that the same representations mediate both tasks and it is the nature of the recognition discrimination (in this instance, individual face recognition), not the task *per se*, that determines the recognition mechanisms employed.

Alan Ashworth (a graduate student at Yale) and I have conducted several experiments to address these issues. In the first block of experiments designed to examine the issue of implicit versus explicit memory, the basic design was the same throughout. There is an initial learning phase in which subjects study a set of faces. In the following testing phase, subjects perform an implicit memory task followed by an explicit memory task. For the implicit task, the studied faces are paired with unstudied faces and presented to the subject in a 2AFC format. The subject indicated which of the two faces they preferred. For the explicit task, the same pairings were once again shown in a 2AFC format; however, this time the subject indicated which one of the faces was previously studied.

Our initial study revealed that subjects have significant memory sensitivity in both conditions, indicating that they recognized the studied faces and preferred the studied faces at above chance levels. Moreover, the explicit and implicit effects were found to be stochastically independent at the trial level, indicating that subjects were not simply liking the faces they recognized, nor recognizing the faces they liked. These results indicate that the memory representations underlying implicit and explicit memory are both sufficiently sensitive to

support subordinate-level discriminations, for instance, recognizing individual faces. Such a conclusion is at odds with claims that implicit memory is mediated exclusively by a structural description (Cooper & Schacter, 1992), presumably too coarse to encode differences between faces. Thus, these results lend support to the argument that it is not the nature of the memory task, but rather information relevant to the task that determines the recognition mechanisms used.

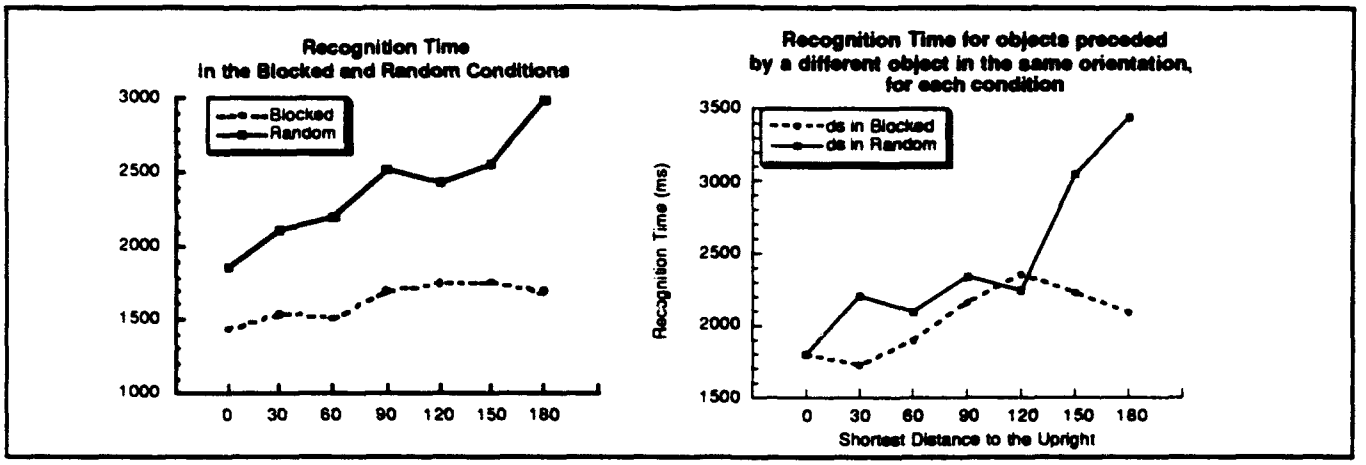
To control for the possibility that memory effects would not be found in the first experiment, an identical experiment using houses as stimuli was conducted. Such a study addresses the concern that a failure to find implicit memory for faces in the first study is not evidence for a coarse structural description, but rather simply that the procedure was inappropriate for finding priming effects. Here, failure to find priming would support this argument, while a finding of priming for houses would support the alternative, that the representations mediating implicit memory are insufficiently sensitive to encode subtle differences between faces. As in the first experiment, a significant explicit memory effect for houses was found, but in contrast to the first experiment, no significant implicit effect was obtained. While the fact that an implicit effect for faces was found in the first experiment renders these results somewhat moot, it is interesting that there was a failure to find priming for the control stimuli (where better priming was actually predicted). One possible factor is that the faces were a much more homogenous stimulus set than the houses. That is, they looked very similar, while the houses were a variety of styles. The implicit tasks used, based on "preferring" one stimulus to another, involves a subtle affect bias. If there is appreciable variability in the attitude that a subject displays towards the individual items in the stimulus set, the bias will be masked. For example, if a subject prefers two-story houses more than one-story houses, then that subject will prefer an unstudied two-story house over a studied one-story house, and the subtle nature of the implicit effect will be lost. Because the faces looked highly similar — in fact had been rated as such in a pilot study — the variability of the subject's attitude toward them was minimized, thus allowing the bias to be measured. Such findings reinforce the need for the development of an adequate contrast set for faces. In particular, the stimulus set discussed in the previous section is far more homogeneous than other contrast sets and, being entirely novel, is unlikely to invoke any preconceived biases.

Following up on these results, we have recently conducted two studies in which we systematically explored the nature of the inversion effect in face recognition. Here we contrasted implicit (face naming) and explicit (recognition memory) tasks using faces rotated in the image-plane away from the studied upright canonical orientation. Unlike earlier studies, finer rotation increments of 30° were used in order to assess the pattern of performance in a more systematic fashion. In both experiments, each studied face was tested at 12 different orientations. Consistent with the hypothesis that the inversion effect is the manifestation of a normalization procedure not unique to faces, a systematic pattern of decreasing performance with increasing misorientation was obtained. The magnitude of this effect was comparable to that found in studies employing non-face stimuli such as novel stick figures (Tarr & Pinker, 1989), cube objects (Tarr, 1989), and line drawings of familiar objects (Jolicoeur, 1985). Also crucial was that a dip (better performance than the surrounding orientations) was found at 180° — a pattern consistent with that found by McMullen and Jolicoeur (1992) and indicative of a similar process for the recognition of both face and non-face stimuli. Finally, there was a practice effect in which the effect of orientation on performance diminished with repeated presentations of the stimuli. This effect is similar to that observed by Tarr and Pinker (1989; Tarr, 1989) and suggests that subjects encode orientation-specific representations of faces if given training at non-canonical orientations. Thus, the inversion effect in face recognition may not be evidence for specialized processing, rather it may result from the fact that faces are

rarely viewed at orientations far from the upright and almost never with complete inversion. To test this more directly, we are beginning several experiments in which faces will be used as stimuli in the paradigm used by Tarr and Pinker (1989) in which some orientations are withheld during initial practice, but later are introduced to assess whether practice effects are due to the encoding of orientation-specific representations. Additionally, several control experiments are planned employing the homogeneous stimulus set discussed in the previous section. Such controls will allow us to better equate recognition performance across changes in orientation for face and non-face stimuli.

5. Reference frame transformations in recognition

One recurring issue in the recognition via mental transformations has been whether it is actually an image of the object that is normalized or whether the perceiver simply transforms their frame of reference, thereby aligning the input shape with all objects encoded at their canonical orientation (both mechanisms would predict the viewpoint-dependent patterns of response times obtained in many mental transformation experiments). At present, the majority of evidence indicates that subjects are incapable of rotating their egocentric frame of reference in a variety of perceptual tasks (Shepard & Cooper, 1982; Robertson, Palmer, & Gomez, 1987). For the most part, studies investigating this issue have focused on tasks other than recognition, for instance employing handedness judgments or judging the top of the stimulus. Indeed, in handedness discriminations of familiar letters and digits, by probing intermediate orientations during a putative rotation, Shepard and Cooper demonstrated that it is an image of the stimulus that is rotated, not an egocentric reference frame. However, it is an open question whether this result generalizes to recognition. In particular, because handedness judgments require an egocentric reference frame in which left and right are defined, they may be less amenable to transformations that the viewer-centered references frames involved in recognition. To test this possibility, Isabel Gauthier (a graduate student at Yale) and I have developed several paradigms in which subjects are given some foreknowledge of the coming orientation of the stimulus and are asked to name the presented object (using novel CVC names arbitrarily associated with each object). This was done in one of two ways: without explicitly informing subjects, a certain number of trials in a seeming random sequence were ordered so that the stimulus object appeared at the same orientation as the object displayed on the previous trial (the "Random" condition); alternatively, subjects were explicitly informed that all trials in a blocked sequence would have the stimulus object appear at the same orientation (the "Blocked" condition). Stimuli were novel 2D stick figures similar to those used by Tarr and Pinker (1989). Subjects were run in both conditions so that learning effects and orientation generalization could be assessed in different training conditions.



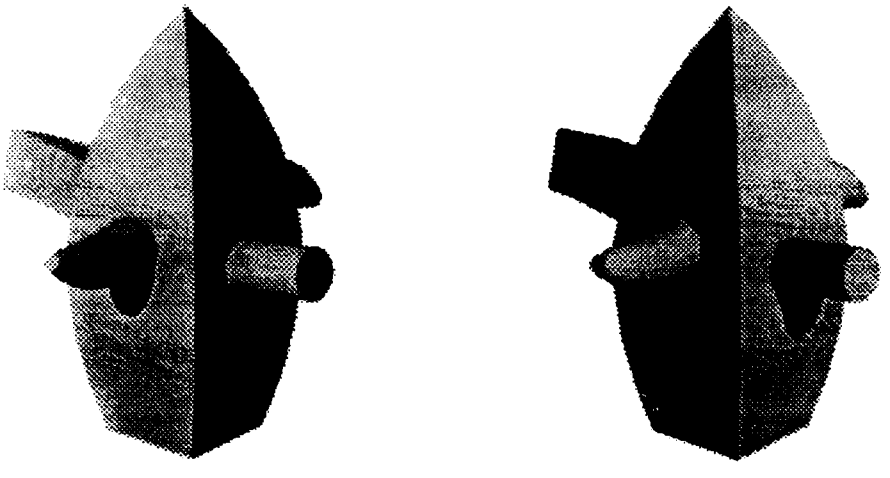
At this point we have only considered the results from the first condition of each subject (where no previous learning has occurred). Response times were found to vary significantly between conditions. Examining results across all trial types (four types: different object/different orientation; different object/same orientation; same object/different orientation; same object/same orientation) from both conditions, some evidence was found for reference frame transformations in the Blocked condition relative to the Random condition (left graph above). Specifically, the Random condition displays monotonically increasing response times with increasing misorientation — a pattern similar to that observed in Tarr and Pinker (1989). In contrast, the Blocked condition displays a much smaller magnitude increase (lower slope) of response time with increasing misorientation. This difference suggests that subjects were able to make use of foreknowledge about orientation to facilitate responses (foreknowledge did not include the actual response — different objects appeared *on* each trial) and therefore supports the hypothesis that viewer-centered reference frames may be transformed for purposes of recognition. A second issue within the data analysis involves directly comparing those trials that are most diagnostic for observing orientation priming: trials where a different stimulus object appears in the same orientation as the object in previous trial. Here, far fewer differences were observed between conditions (right graph above). Specifically, response times for both conditions were not reliably difference for misorientations up to 120°. This result suggests that egocentric reference frames may be transformed, and, in particular, even in instances where subjects have no explicit foreknowledge of that the next trial was to use the same reference orientation. However, at greater misorientations, in this instance, 150° and 180°, lack of explicit foreknowledge regarding stimulus orientation apparently hindered the transformation of the frame of reference. Overall these results provide some support for the ability to transform egocentric reference frames for purposes of recognition. However, the strongest evidence comes from conditions in which subjects had explicit foreknowledge regarding the coming stimulus orientation — a context that may occur only rarely in "everyday" recognition. On the other hand, based on the similarities between the two conditions for different object/same orientation trials, there is some evidence that reference frames may be routinely aligned with the orientation of the stimulus object. In particular, data from the Random condition indicates that this transformation may occur even in instances where subjects are unaware of the possible advantages such a transformation may confer.

We are now exploring several issues raised by these studies. For example, it is unclear in the Blocked condition whether it was the explicit foreknowledge of orientation or repeated occurrence of same-oriented objects that prompted the apparent transformation of reference frame. The former implies a conscious process of transformation, while the latter implies an

unconscious mechanism. To address this issue we have designed a paradigm in which different length "runs" of same-orientation objects appear imbedded in an apparently random sequence. In this design subjects will not have explicit foreknowledge of the repeated orientations and we be able to manipulate the length of each run to explore the degree of redundancy necessary to prompt reference frame shifts (given the mechanism is unconscious). A second experiment is designed to examine performance given perfect foreknowledge: essentially a variant of Shepard and Cooper's (1982) orientation cueing studies, subjects will be given a correct orientation cue prior to each trial. The critical difference between this and earlier studies being the use of recognition rather than handedness as the task. We are also investigating possible differences in how orientation-specific object representations develop under conditions where orientation is predictable (Blocked) versus unpredictable (Random).

6. Specificity of encoding of in image-based representations

One extension of the multiple-views theory of recognition is the possibility that object representations are *image-specific* rather than simply viewpoint-specific. In particular, it may be that image properties such as illumination, texture, and color are encoded along with shape information. Indeed, recent evidence suggests that object memory may be more specific to color than previously thought (Wurm, Legge, Isenberg, & Luebker, 1993). In collaboration with Dan Kersten and Heinrich Bülthoff, we have been investigating whether the same is true for illumination (or consequential shading). Illumination is particularly interesting in that it is often assumed to be a source of a information immediately discounted in recognition (although extremely useful in inferring shape). This is true for two reasons: first, it is a source of extreme variability in that images of the same object illuminated from orthogonal directions may correlate less than images of two different objects illuminated from the same direction; second, illumination is generally not diagnostic for identity (as is color, i.e., yellow bananas).



Same Object in Orthogonal Lighting Conditions

Across several experiments we have been investigating whether extreme changes in illumination between encoding and test (as in the object shown above) affect recognition performance. The first experiment employed 6 part-differentiable objects (the objects used in the training set in Section 2) in a 2AFC design. Objects were shown sequentially with an intervening mask and isi's of 500 ms or greater (a range generally thought to remove image persistence as the explanation for any performance cost between encoding and test; see, Ellis

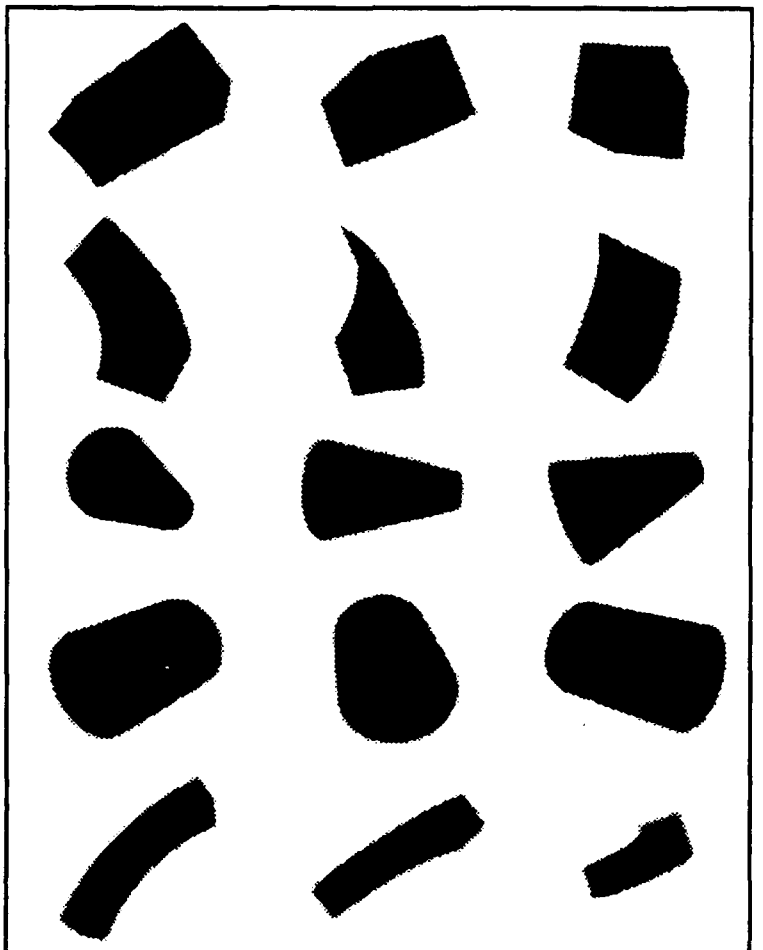
& Allport, 1986). The subjects' task was simply to judge whether the two objects displayed were the same or different. In this straightforward study, a reliable 20 ms response time cost was found for transforming illumination to an orthogonal position, but accuracy was equivalent in the changed and unchanged illumination conditions. Thus, while there was some evidence for sensitivity to illumination direction, suggesting that such information may have been encoded in the object representation, it is relatively small and does not affect successful recognition. Based on evidence that more complex object discriminations prompt increased reliance on viewpoint-specific mechanisms (i.e., Tarr & Pinker, 1990), we reasoned that a more complex discrimination here would result in increased reliance on image-specific mechanisms and consequently greater costs for changes in illumination. This was done by introducing an additional 6 objects into the design (the objects used in the contrast set in Section 2) that shared many parts and configurations with the previously used objects. Otherwise the design was identical to that used previously. As in the first experiment, a reliable 20 ms cost was found for changing illumination between study and test and, crucially, a significant decrease in sensitivity (as measured by d') was found for those trials in which illumination changed versus those where it remained constant. This result provides some evidence that image-specific properties such as illumination are encoded in object representations, and, much as with viewpoint-specific effects, are most likely to mediate complex recognition in which subtle shape discriminations must be made. We are continuing with this work in several directions, including using naming tasks and a range of new stimuli. We are also develop paradigms to assess both texture- and color-specificity in recognition.

7. The role of viewpoint in entry-level recognition

While there is evidence for view-based mechanisms in recognition, for the most part they have been assigned the circumscribed role of subordinate-level within-class recognition (i.e., Biederman, 1987). Indeed, there have recently been claims that "everyday" entry-level recognition is mediated by part-based structural descriptions that are viewpoint-invariant up to changes in visible and occluded parts (Biederman & Gerhardstein, 1993). In particular, Biederman and Gerhardstein list three conditions for *immediate* viewpoint invariance: objects must be decomposable into parts; objects must be composed of distinct parts; and identical part configurations must be visible. Importantly, these conditions lead to several specific predictions concerning entry-level recognition performance. First, the recognition of single parts (geons) is predicted to be *completely* viewpoint invariant in that individual parts function as the invariant features of the structural description. Second, the recognition of multi-part objects is predicted to be viewpoint invariant so long as the same configuration of parts is visible (no occlusions or new parts). These assumptions were tested in several experiments employing either single 3D parts or multi-part objects. In both instances, objects were differentiable on the basis of qualitative differences between the parts (adapted from Biederman & Gerhardstein, 1993), thereby resulting in an entry-level discrimination.

7a. Single 3D volumes

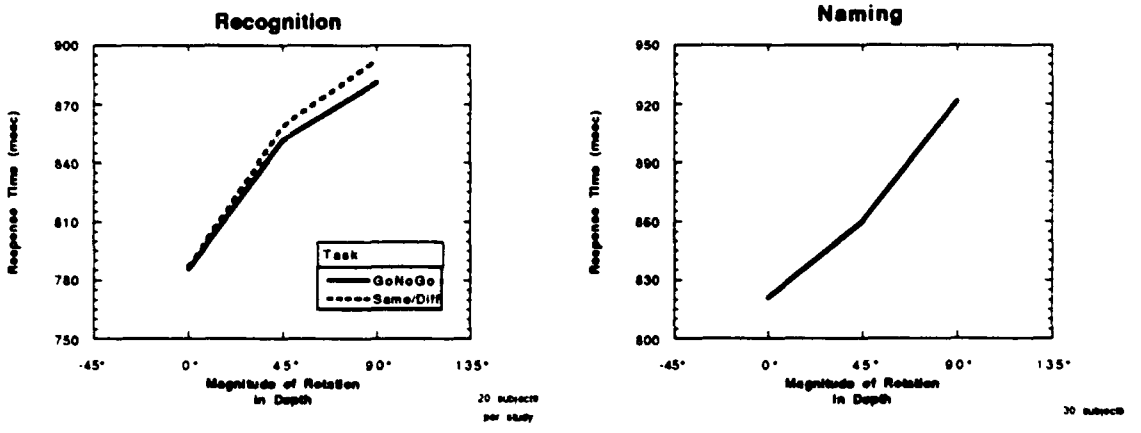
Ten three-dimensional volumes (five of which are shown below) were photorealistically rendered from three viewpoints each. Objects were rendered with texture and shading (held constant across rotations) to enhance their realistic three-dimensional appearance. Stimuli were used in a series of experiments in an attempt to both replicate and test the data and theory presented in Biederman and Gerhardstein (1993; also, Hummel & Biederman, 1992). Two experiments employed an 2IFC design in which a volume was displayed, followed by a mask (composed of random parts), followed by another volume, and, finally, the same mask. Isi's were greater than the interval found by Ellis and Allport (1986) to remove all effects of image persistence. The subject's task was simply to judge whether the two volumes were the same or different regardless of changes in viewpoint. In the first study, a "go/no-go" task was used; subjects were to respond only when the two volumes were believed to be the same and to do nothing when they were believed to be different. This task was used to match Biederman and Gerhardstein (1993) as closely as possible. Results were straightforward: response times (and error rates) for same responses were observed to be monotonically dependent on the rotational separation between the two views. In order to obtain converging evidence and to assess the impact of employing the somewhat idiosyncratic go/no-go task, the same study was run using a same/different task in which subjects responded both when they believed the objects to be the same and when they believed the objects to be different. The primary finding of this manipulation was a pattern of performance almost identical to that observed in the go/no-go task. However, there are two additional points of note: First, variance in the go/no-go response task was actually *greater* than in the same/different — a finding that contradicts Biederman and Gerhardstein's claim of reduced variance in such tasks and generally indicates that there is little reason to use such a task in that it is somewhat atypical of "normal" recognition and provides less information about cognitive processing (no different responses). Second, performance for different responses was equivalent at all rotational disparities, a finding consistent with viewpoint-dependent mechanisms in which a normalization is only executed subsequent to a precomputation of a valid transformation.



Sample Single 3D Volumes (Geons)

The same objects and viewpoints were also used in a naming study in which each volume was assigned a somewhat diagnostic name. For example, for the five volumes shown above, from top to bottom: "brick", "claw", "cone", "cylinder", "fry". Subjects were taught the names and were then given practice naming each object from a single viewpoint (the leftmost viewpoint in

the above figure). Subjects then named all ten objects from all viewpoints and their voice onset times were recorded (errors were almost nonexistent). As shown in the right-hand graph above, response times in a verbal naming task equivalent to entry-level naming were found to increase monotonically with increasing rotations in depth away from the initially learned viewpoint. Such a result is crucial in that it is inconsistent with the predictions of geon-structural description theory, yet provides a far more stringent test than those used by Biederman and Gerhardstein. While these results are inconsistent with viewpoint-invariant theories of recognition, they support a viewpoint-dependent account — in particular, one in which objects are represented at specific viewpoints and recognized via normalization between the observed viewpoint and the nearest familiar viewpoint encoded in object memory.

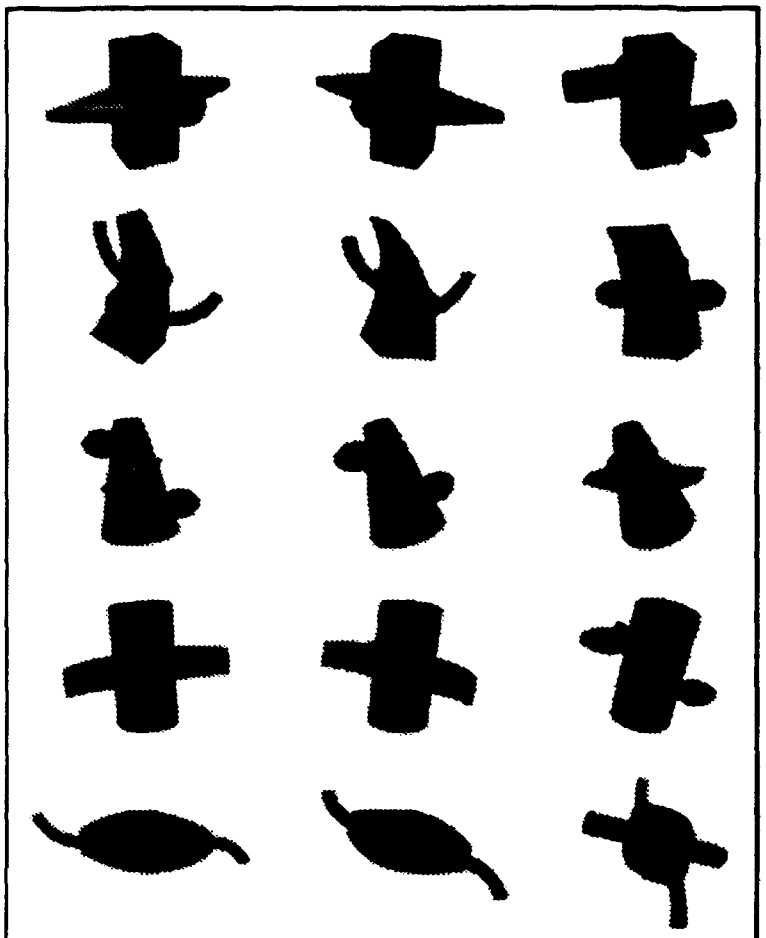


Recognition of Single 3D Volumes

7b. Multi-Part objects

Ten three-dimensional multi-part objects (five of which are shown below) were photorealistically rendered from many viewpoints. Objects were rendered with texture and shading (held constant across rotations) to enhance their realistic three-dimensional appearance. Subjects were run in a 2IFC design using a same/different response task and random-part masks following both stimulus presentations. Isi's were greater than the interval found by Ellis and Allport (1986) to remove all effects of image persistence. Two variables were manipulated in this study: the rotational disparity in depth between the two objects, and, the particular viewpoints that were displayed in each view pair. Consequently, several different view pairs yielded same-magnitude rotational disparities, but were considered separately because of Biederman and Gerhardstein's prediction that changes in visible part configuration would result in viewpoint effects, but that identical part configurations would not result in viewpoint effects. Thus, equal magnitude rotations were predicted to have reliably different performance characteristics depending on the qualitative changes in part structure. Alternatively, it is possible that viewpoint effects will be constant across rotations, predicted by the magnitude of the rotation rather than any changes in the image. Such a result would be problematic for part-based structural description theories and would instead provide evidence for viewpoint-dependent normalization theories.

Results in this experiment are straightforward. Performance patterns monotonically related to the magnitude of rotational disparity were obtained in all instances, *regardless* of whether the two views crossed one or more boundaries where parts became visible or occluded. In particular, for all possible view pairs (graphed by placing all view pairs sharing a common view

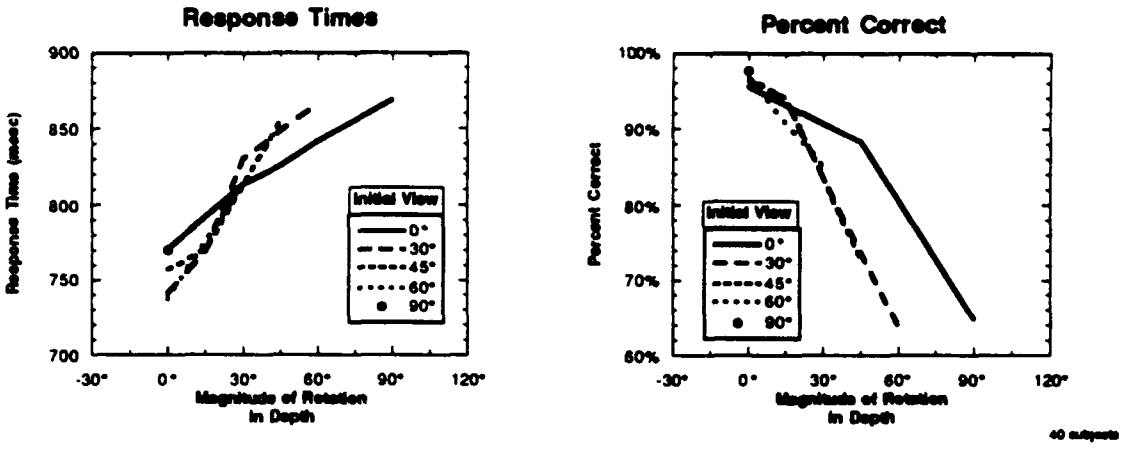


Multi-Part 3D Objects (Geon-Differentiable)

on the same line; for example, the 45° line plots performance for all view pairs where the 45° viewpoint, the middle view as shown in the figure to the left, is the nearest to canonical; points on that line represent data for rotations away from the 45° viewpoint, the magnitude of the rotation specified by the horizontal axis) response times were equivalent and predicted solely by the rotational disparity of the views. Crucially, this experiment is somewhat more "ecologically valid" as compared to those presented by Biederman & Gerhardstein: not only were the objects more realistic in both shading and texture, but the far wider range of viewpoints is more typical of everyday recognition where viewpoints are unlikely to be predictable or restricted. The results of this experiment provide two additional challenges to viewpoint-invariant structural description theories: first, a consistent effect of viewpoint was observed to part-differentiable objects; second, this effect was obtained regardless of the changes that occurred in the visible configuration of parts.

Overall, the findings obtained with both single volumes and multi-part objects are remarkably consistent. In each instance, the most reliable predictor of subject performance, regardless of task, was rotational disparity between familiar views of the objects. Such results are problematic for viewpoint-invariant theories of recognition, even given the limits placed on obtaining invariance proposed by Biederman and Gerhardstein. Moreover, such results provide strong evidence for viewpoint-dependent mechanisms, and, in particular, object representations organized on the basis on multiple viewpoint-specific views along with normalization procedures for matching perceived objects with those encoded in memory. Finally, the work outlined here extends this approach to a variety of tasks and stimulus conditions heretofore assumed to be based on alternative mechanisms.

We are currently exploring a range of issues raised by these studies. For example, one area of interest is the difference in processing between line drawings and realistic renderings — a question that has implications for many areas of cognitive psychology in that many studies have employed line drawings as stimuli. Other investigations are exploring what models of qualitative change best account for the delineation of views in object representations. In particular, we have designed several studies to directly compare part-based models to feature-based models — the prediction being that it is changes in image structure (i.e., aspect graphs) that mediate subject performance across changes in view.



Recognition of Multi-Part 3D Objects

Presentations & Papers:

Tarr, M. J., & Bülthoff, H. H. (1993). *Conditions for viewpoint dependence and viewpoint invariance: What mechanisms are used to recognize an object?* (Tech Report Memo 3). Arbeitsgruppe Bülthoff, Max-Planck-Institut für biologische Kybernetik. Submitted to *Journal of Experimental Psychology: Human Perception and Performance.*

REFERENCES

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(6), 1162-1182.

Cooper, L. A., & Schacter, D. L. (1992). Dissociations between structural and episodic representations of visual objects. *Current Directions in Psychological Science*, *1*(5), 141-146.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(27), 107-117.

Ellis, R., & Allport, D. A. (1986). Multiple levels of representation for visual objects: A behavioural study. In A. G. Cohn & J. R. Thomas (Eds.), *Artificial intelligence and its applications* (pp. 245-247). New York: Wiley.

Farah, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain-specificity in visual object recognition. *Current Directions in Psychological Science*, *1*(5), 164-169.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*(3), 480-517.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, *13*, 289-303.

McMullen, P. A., & Jolicoeur, P. (1992). Reference frame and effects of orientation on the finding the tops of rotated objects. *Journal of Experiment Psychology: Human Perception and Performance*, *18*(3), 807-820.

Robertson, L. C., Palmer, S. E., & Gomez, L. M. (1987). Reference frames in mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(37), 368-379.

Shepard, R. N., & Cooper, L. A. (1982). *Mental Images and Their Transformations*. Cambridge, MA: The MIT Press.

Tarr, M. J. (1989). *Orientation dependence in three-dimensional object recognition*. Unpublished Doctoral Dissertation. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*(28), 233-282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, *1*(42), 253-256.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 899-911.

OTHER PUBLICATIONS

Tarr, M. J., & Black, M. J. (In Press). A computational and evolutionary perspective on the role of representation in vision. *Computer Vision, Graphics, and Image Processing: Image Understanding, 59*(3).

Tarr, M. J., & Black, M. J. (In Press). Reconstruction and Purpose. *Computer Vision, Graphics, and Image Processing: Image Understanding, 59*(3).

Tarr, M. J. (1993). From perception to cognition. *Behavioral and Brain Sciences, 16*(2), 251-252.

Tarr, M. J. (1993). Is a picture really worth a thousand words? *Computational Intelligence, 9*(4), 356-359.

Tarr, M. J. (In Press). Visual representation. *Encyclopedia of Human Behavior.* San Diego, CA: Academic Press.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (In Press). *How are three-dimensional objects represented in the brain?* (Tech Report Memo 5). Arbeitsgruppe Bülthoff, Max-Planck-Institut für biologische Kybernetik. In Press, *Cerebral Cortex.*

Tarr, M. J. (Under Review). Behavioral and computational constraints in human object representation. Submitted to *Psychological Bulletin.*

Tarr, M. J. (Under Review). Rotating objects to recognize them: A case study of the role of mental transformations in the recognition of three-dimensional objects. Submitted to *Psychonomic Bulletin and Review.*

Tarr, M. J., & Kriegman, D. J. (Under Review). Toward understanding human object recognition: Aspect graphs and view-based representations. *Submitted to Psychological Review.*

PERSONNEL

Doctoral Students	Isabel Gauthier	1st Year
	Pepper Williams	1st Year
	William Hayward	3rd Year
	Alan Ashworth	4th Year
Undergraduates	Scott Yu	BA in May 1994
	KaRin Turner	"

OTHER CONFERENCE PRESENTATIONS AND INVITED COLLOQUIA

Tarr, M. J., & Kriegman, D. J. A formal basis for understanding view-based representations in humans. *Workshop on Visual Perception: Computation and Psychophysics*, Cape Cod, MA, January, 1993.

Tarr, M. J. Invited panel member, special session on purposive vision. *International Joint Conference on Artificial Intelligence*, Chambery, France, August, 1993.

COLLOQUIA: Department of Cognitive and Neural Systems, Boston University, February, 1993; Department of Psychology, University of Toronto, March, 1993; Department of Cognitive Science, Brown University, April 1993; ONR Workshop on Cognitive Neuroscience, Pittsburgh, PA, October 1993; Department of Psychology, Columbia University, November 1993; Max-Planck-Institut für biologische Kybernetik, Tübingen, Germany, December 1993.

OTHER

Organized first annual meeting of the Pre-Psychonomics Workshop on Object Perception and Memory (OPAM).