Orientation Dependence in Three-Dimensional Object Recognition

by

Michael J. Tarr

B.A., Psychology
Cornell University
(1984)

Submitted to the Department of Brain and Cognitive Sciences
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

September 1989

Signature of Author____
Department of Brain and Cognitive Sciences
May 18, 1989

Certified by____
Steven Pinker
Professor, Brain and Cognitive Sciences
Thesis Supervisor

Accepted by____
Emilio Bizzi
Head, Department of Brain and Cognitive Sciences

Table of Contents

ORIENTATION DEPENDENCE IN THREE-DIMENSIONAL OBJECT RECOGNITION

by

MICHAEL J. TARR

Submitted to the Department of Brain and Cognitive Sciences
on May 18, 1989 in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy in Cognitive Science.

ABSTRACT

Successful vision systems must overcome differences in two-dimensional input shapes arising from orientation changes in three-dimensional objects. How the human visual system solves this problem is the focus of much theoretical and empirical work in visual cognition. One issue central to this research is: are input shapes and stored models involved in recognition described independently of viewpoint? In answer to this question two general classes of theories of object recognition are discussed: viewpoint independent and viewpoint dependent. The major distinction between these classes is that viewpoint-independent recognition is invariant across viewpoint such that input shapes and stored models are encoded free of the orientation from which they arose, while viewpoint-dependent recognition is specific to viewpoint such that input shapes and stored models are encoded in particular orientations, usually those from which they arose.

Five experiments are presented that examine whether the human visual system relies on viewpoint-independent or viewpoint-dependent representations in three-dimensional object recognition. In particular, these experiments address the nature of complex object recognition -- what are the processes and representations used to discriminate between similar objects within the same general class? Two competing theories are tested: a viewpoint-independent theory, best characterized by *object-centered mechanisms*, and a viewpoint-dependent theory, in particular one that relies on the *multiple-views-plus-transformation mechanism*. In the object-centered theory input shapes and stored models are described in a reference frame based on the object itself -- as long as the same features are chosen for both object-centered reference descriptions, the two will match. In the multiple-views-plus-transformation theory input shapes are described relative to a reference frame based on the current position of the viewer, while stored models are described relative to a prior position of the viewer -- when these viewer-centered descriptions correspond, the two may be matched directly, otherwise the input shape must be transformed into the viewpoint of a stored model.

All five experiments tested these competing theories by addressing two questions: (1) Was there an initial effect of orientation on the recognition of novel objects, and if so, did this effect diminish after practice at several orientations; and (2) did diminished effects of orientation at familiar orientations transfer to the same objects in new, unfamiliar orientations? Each of the experiments yielded similar results: initial effects of orientation were found; with practice these effects of orientation diminished; and the diminished effects of orientation did not transfer to unfamiliar orientations. Not only did the effects of orientation return for unfamiliar orientations, but these effects increased with distance from the nearest familiar orientation, suggesting that subjects rotated objects in non-stored orientations through roughly the shortest three-dimensional path to match stored models at familiar orientations. Overall, these results support the existence of a multiple-views-plus-transformation mechanism and suggest that at least for complex discriminations, three-dimensional object recognition is viewpoint dependent.

Thesis Supervisor: Dr. Steven Pinker

Title: Professor of Brain and Cognitive Sciences

Acknowledgments

Two people deserve thanks that can not be easily put into words. Steve Pinker, my advisor, my collaborator, and most of all my friend; and Laurie Heller, my companion, my inspiration, and much more.

David Irwin, who helped me get interested in graduate school in the first place, deserves much of the credit and none of the blame.

Thanks to Irv Biederman and Ellen Hildreth for their helpful comments, support, and advice.

Special thanks to Jacob Feldman, friend and colleague, whose thoughtful discussions have shaped many of my ideas in this thesis (as well as my stereo).

Several other graduate students have been particularly important to me over the past five years. Kyle Cave and Jess Gropen have shared ideas and, more importantly, comradeship. Paul Bloom has been my partner during the sometimes arduous search for employment. In addition, all of the graduate students in our department have been my friends and I will miss them and the community they form.

Thanks to Jan Ellertsen for everything she does for all of the graduate students.

Two terrific UROP's, Jigna Desai and Carmita Signes, deserve thanks for running hundreds of hours of subjects.

I would also like to thank my family, Dad, Tova, Joanna, Maya, and Ilana, for their love and affection.

Finally I wish to acknowledge the financial support provided by the James R. Killian Fellowship sponsored by the James and Lynelle Holden Fund, a Fellowship from the Whitaker Health Sciences Fund, and a NSF Graduate Fellowship. Parts of this research were funded under NSF Grant BNS 8518774 and a grant from the Sloan Foundation to the Center for Cognitive Science.

Orientation Dependence in Three-Dimensional Object Recognition

1. Introduction

How do we recognize objects in three dimensions despite changes in orientation producing different two-dimensional projections? Stored knowledge about objects must be compared to visual input, but the format of this stored knowledge may take many forms. For instance, one might rely on shape-based mechanisms to recognize an object by a small set of unique features, by the two-dimensional input shape, or by the three-dimensional spatial relations between parts. Additionally, recognition might rely on mechanisms using texture, color, or motion. All of these possibilities may play a role in achieving *shape constancy*, the recognition of an object plus its three-dimensional structure from all possible orientations. Furthermore, some of these possibilities may coexist in recognition. For example, unique features might suffice for simple recognition, while complex object recognition, involving discriminations between objects that lack distinguishing or easily located features, might require spatial comparisons between stored representations of objects and input shapes. It is these complex spatial comparisons that this thesis addresses.

2. Viewpoint Dependence In Shape Recognition

2.1. Families of recognition theories

Generally, competing theories of shape-based recognition may be divided into the following four classes (see Pinker, 1984; Tarr and Pinker, 1989a):

1. *Viewpoint-independent theories* in which an observed object is assigned the same representation regardless of its orientation, size, or location. Frequently such theories rely on *structural-description*

models, in which objects are represented as hierarchical descriptions of the three-dimensional spatial relationships between parts, using a coordinate system centered on the object or a part of the object. Prior to describing an input shape, a coordinate system is centered on it, based on its axis of elongation, symmetry, or other geometric properties, and the resulting "object-centered" description is matched directly with stored shape descriptions, which use the same coordinate system (e.g., Marr and Nishihara, 1978).

2. *Single-view-plus-transformation theories* in which objects are represented at a single orientation in a coordinate system determined by the location of the viewer (a "viewer-centered" description). A description of an observed object at its current orientation is mentally transformed (for instance, by mental rotation) to a canonical orientation where it may be matched to stored representations.
3. *Multiple-views theories* in which objects are represented at several familiar orientations. A description of an observed object may be matched to stored representations if its current orientation corresponds to one of the familiar orientations.
4. *Multiple-views-plus-transformation theories* in which objects are represented at several familiar orientations. A description of an observed object may be matched directly to stored representations if its current orientation corresponds to one of the familiar orientations, otherwise it may be mentally transformed from its current orientation to the nearest familiar orientation where it may be matched to stored representations.

Tarr and Pinker (1989a) point out that each type of recognition mechanism makes specific predictions about the effect of orientation on the amount of time required for the recognition of an object. All viewpoint-independent theories predict that the recognition time for a particular object will be invariant across all orientations (assuming that it takes equal time to assign a coordinate system to an input shape at different orientations). The multiple-views theory makes a similar prediction (although only for orientations that correspond to those stored in memory -- at non-stored orientations recognition will fail). In contrast, the single-view-plus-transformation theory, assuming it uses an incremental transformation process, predicts that recognition time will be monotonically dependent on the orientation difference between the observed object and the canonical stored one. Similarly, the multiple-views-plus-transformation theory also predicts that recognition time will vary with orientation, but that recognition time will be monotonically dependent on the orientation difference between the observed object and the nearest of several stored representations.

2.2. Studies of the recognition of shapes at different orientations

An examination of current research on object recognition drawn from both computational vision and experimental psychology makes it apparent that there is little consensus concerning how the human visual system accommodates variations in viewpoint. Several computational theories and empirical studies have argued for viewpoint-independent recognition (Biederman, 1987; Corballis, 1988; Corballis, Zbrodoff, Shetzer, and Butler, 1978; Marr and Nishihara, 1978; Pentland, 1986; Simion, Bagnara, Roncato, and Umilta, 1982), while others have argued for viewpoint-dependent recognition (Jolicoeur, 1985; Koenderink, 1987; Lowe, 1987; Ullman, 1986). Because of this dichotomy, I begin by reviewing experimental findings concerning the role of viewpoint dependence in shape recognition.

2.2.1. Evidence for a mental rotation transformation

Cooper and Shepard (1973) and Metzler and Shepard (1974) found several converging kinds of evidence suggesting the existence of an incremental or analog transformation process, which they called "mental rotation". First, when subjects discriminated standard from mirror-reversed shapes at a variety of orientations, they took monotonically longer for shapes that were further from the upright. Second, when subjects were given information about the orientation and identity of an upcoming stimulus and were allowed to prepare for it, the time they required was related linearly to the orientation; when the stimulus appeared, the time they took to discriminate its handedness was relatively invariant across absolute orientations. Third, when subjects were told to rotate a shape mentally and a probe stimulus was presented at a time and orientation that should have matched the instantaneous orientation of their changing image, the time they took to discriminate the handedness of the probe was relatively insensitive to its absolute orientation. Fourth, when subjects were given extensive practice at rotating shapes in a given direction and then were presented with new orientations a bit past 180° in that direction, their response times were bimodally distributed, with peaks corresponding to the times expected for rotating the image the long and the short way around. These converging results suggest that mental rotation is a genuine transformation process, in which a shape is represented as passing through intermediate orientations before reaching the target orientation (for an extensive review see Shepard and Cooper, 1982).

2.2.2. Evidence interpreted as showing that mental rotation is used to assign handedness but not to recognize shape

Because response times for unpredictable stimuli increase monotonically with increasing orientational disparity from the upright, people must use a mental transformation to a single orientation-specific representation to perform these tasks. However, this does not mean that mental rotation is used to recognize shapes. Cooper and Shepard's task was to distinguish objects from their mirror-image versions, not to recognize or name particular shapes. In fact, Cooper and Shepard argue that in order for subjects to find the top of a shape before rotating it, they must have identified it beforehand. This suggests that an orientation-free representation is used in recognition, and that the mental rotation process is used only to determine handedness.

Subsequent experiments have supported this argument. Corballis, et. al. (1978) had subjects quickly name misoriented letters and digits; they found that the time subjects took to name normal (i.e., not mirror-reversed) versions of characters was largely independent of the orientation of the character. A related study by Corballis and Nagourney (1978) found that when subjects classified misoriented characters as letters or digits there was also only a tiny effect of orientation on decision time. White (1980) also found no effect of orientation on either category or identity judgments preceded by a correct cue, either for standard or mirror-reversed characters, but did find a linear effect of orientation on handedness judgments. Simion, et al. (1982) had subjects perform "same/different" judgments on simultaneously presented letters separated by varying amounts of rotation. In several of their experiments they found significant effects of orientation on reaction time, but the effect was too small to be attributed to mental rotation. Eley (1982) found that letter-like shapes containing a salient diagnostic feature (for example a small closed curve in one corner or an equilateral triangle in the center) were recognized equally quickly at all orientations.

2.2.3. The rotation-for-handedness hypothesis

Based on these effects, Corballis, et al. (1978; see also Corballis, 1988; Hinton and Parsons, 1981) have concluded that under most circumstances recognition (up to but not including the shape's handedness) is accomplished by matching an input shape to an orientation-independent representation. Such a representation does not encode handedness information; it matches both standard and mirror-reversed versions of a shape equally well at any orientation. Therefore subjects must use other means to assess handedness. Hinton and Parsons suggest that handedness is inherently egocentric; observers determine the handedness of a shape by seeing which of its parts corresponds to our left and right sides when the shape is upright. Thus if a shape is misoriented, it must be mentally transformed to the upright. Tarr and Pinker (1989a) call this the "Rotation-for-Handedness" hypothesis.

2.2.4. Three problems for the rotation-for-handedness hypothesis

These findings seem to relegate mental rotation to the highly circumscribed role of assigning handedness. Moreover, this implies that other mechanisms, presumably using object-centered descriptions or other orientation-invariant representations, are used to recognize objects. However, Tarr and Pinker (1989a) cite three serious problems for the rotation-for-handedness hypothesis:

1. Tasks allowing detection of local cues. First, in many experimental demonstrations of the orientation-invariance of shape recognition, the objects could have contained one or more diagnostic local features that allowed subjects to discriminate them without processing their shapes fully. The presence of orientation-free local diagnostic features was deliberate in the design of Eley's (1982) stimuli, and he notes that it is unclear whether detecting such features is a fundamental recognition process or a result of particular aspects of experimental tasks such as extensive familiarization with the stimuli prior to testing and small set sizes.

Similarly in White's (1980) experiment, the presentation of a correct information cue for either identity or category may have allowed subjects to prepare for the task by looking for a diagnostic orientation-free feature. In contrast, the presentation of a cue for handedness would not have allowed subjects to prepare for the handedness judgment, since handedness information does not in general allow any concrete feature or shape representation to be activated beforehand.

2. Persistent small effects of orientation. A second problem for the rotation-for-handedness hypothesis is the repeated finding that orientation does have a significant effect on recognition time, albeit a small one (Corballis, et al., 1978; Corballis and Nagourney, 1978; Simion, et al., 1982). Corballis, et al. note that the rotation rate estimated from their data is far too fast to be caused by consistent use of Cooper and Shepard's mental rotation process; they suggest that it could be due to subjects' occasional use of mental rotation to double-check the results of an orientation-invariant recognition process, resulting in a small number of orientation-sensitive data being averaged with a larger number of unvarying data. However, Jolicoeur and Landau (1984) suggest that normalizing the orientation of simple shapes might be accomplished extremely rapidly, making it hard to detect strong orientation effects in chronometric data. By having subjects identify misoriented letters and digits presented for very brief durations followed by a mask, Jolicoeur and Landau were able to increase subject's identification error rates to 80% on practice letters and digits. When new characters were presented for the same duration with a mask, subjects made systematically more identification errors as characters were rotated further from upright. Jolicoeur and Landau interpret their data as supporting a theory of object recognition based on "time-consuming normalization processes" other than classical mental rotation.[1]

3. *Interaction with familiarity.* A final problem for the rotation-for-handedness hypothesis is that orientation independence in recognition seems to occur only for highly familiar combinations of shapes and orientations; when unfamiliar stimuli must be recognized, orientation effects reminiscent of mental rotation appear. Shinar and Owen (1973) conducted several experiments in which they taught subjects a set of novel polygonal forms at an upright orientation and then had the subjects classify misoriented test shapes as being a member or not being a member of the taught set. The time to perform this old-new judgment for the familiar shapes was in fact dependent on their orientation, and this effect disappeared with practice. Jolicoeur (1985) had subjects name line drawings of natural objects. At first their naming times increased as the drawings were oriented further from the upright, with a slope comparable to those obtained in classic mental rotation tasks. With practice, the effects of orientation diminished, though the diminution did not transfer to a new set of objects. This pattern of results suggests that people indeed use mental rotation

---

[1] A defender of the rotation-for-handedness hypothesis, however, could accommodate these data. Even if representations used in recognition were completely orientation-independent, a perceiver must first find the intrinsic axes or intrinsic top of an object in order to describe it within a coordinate system centered on that object. If the search for the intrinsic axis of an input shape begins at the top of the display, rotations further from the upright would be expected to produce an increase in recognition time, and this axis-finding process could be faster than the rate of mental rotation.

to recognize unfamiliar shapes or examples of shapes. As the objects become increasingly familiar, subjects might become less sensitive to their orientation, for one of two reasons. They could develop an orientation-invariant representation of the object, such as an object-centered structural description or set of features. Alternatively, subjects could come to store a set of orientation-specific representations of the object, one for each orientation it is seen at, at which point recognition of the object at any of these orientation could be done in constant time by a direct match.

These familiarity effects complicate the interpretation of all of the experiments in which subjects were shown alphanumeric characters. Letters and digits are highly familiar shapes that subjects have had a great deal of prior experience recognizing, presumably at many orientations (Koriat and Norman, 1985). Thus it is possible that humans store multiple orientation-specific representations for them; recognition times would be constant across orientations because any orientation would match some stored representation. In fact this hypothesis is consistent with most of the data from the Corballis, et al. studies. In their experiments where subjects named standard and reversed versions of characters, although there was only a small effect of orientation on naming latencies for standard versions, there was a large effect of orientation on naming latencies for reversed versions. On the multiple-views hypothesis, this could be explained by the assumption that people are familiar with multiple orientations of standard characters but only a single orientation of their mirror-reversed versions, which are infrequently seen at orientations other than the upright (Koriat and Norman, 1985). In addition, it is more likely that multiple orientation-specific representations exist for standard characters within +/-90 degrees from upright, since subjects rarely read and write characters beyond these limits. This would explain why mental rotation functions for alphanumeric characters are generally curvilinear, with smaller effects for orientations near the upright (see Koriat and Norman, 1985). With practice, subjects should begin to develop new representations for the presented orientations of the reversed versions and for previously unstored orientations of the standard versions of characters. This would account for Corballis et al.'s (1978) finding of a decrease in the effect of orientation with practice.

One might also expect that as subjects are given increasing practice at determining the handedness of alphanumeric characters at various orientations, they should become less sensitive to orientation, just as is found for recognition. Although Cooper and Shepard (1973) found no change in the rate of mental rotation in their handedness discrimination tasks even with extensive practice, their non-naive subjects may have chosen to stick with the rotation strategy at all times. Kaushall and Parsons (1981) found that when subjects performed same-different judgments on successively presented three-dimensional block structures at different orientations, slopes decreased (the rate of rotation got faster) after extensive practice (504 trials).

2.2.5. Summary

In sum, as originally stated by Tarr and Pinker (1989a), the empirical literature does not clearly support the rotation-for-handedness hypothesis. Unless there is a local diagnostic feature serving to distinguish objects, both handedness judgments and recognition judgments take increasingly more time for orientations farther from the upright when objects are unfamiliar, but become nearly (though not completely) independent of orientation as the objects become familiar. This seems to indicate a role for mental rotation in the recognition of unfamiliar stimuli; the practice/familiarity effect, however, could reflect either the gradual creation of a viewpoint-independent representation for each object, or the storing of a set of viewpoint-dependent representations, one for each object at each orientation. Thus the question of which combination of the four classes of mechanisms humans use to achieve object recognition is unresolved.[2]

---

[2] Jolicoeur's (1985) finding that diminished effects of orientation do not transfer from a set of practiced objects to a set of new objects does not resolve the debate. This lack of transfer demonstrates only *pattern specificity*, not *orientation specificity*. Both orientation-invariant and multiple orientation-specific representations are pattern specific, although only in the latter case are the acquired representations committed to particular orientations.

2.3. Evidence interpreted as showing that viewpoint-dependent mechanisms are used in complex shape recognition

2.3.1. A paradigm for testing whether representations of shape are viewpoint specific

Tarr and Pinker's (1989a) study was designed to examine the viewpoint specificity of representations of shapes used in recognition. In particular, all of their experiments had elements that addressed the problems cited for earlier studies. First, all of Tarr and Pinker's experiments used novel characters that contained similar local features, but different global configurations, and therefore contained no local diagnostic features that might have provided an alternate path to recognition. Second, all of the characters had a salient feature indicating their bottom, and a well-marked intrinsic axis, minimizing effects of finding the top-bottom axis at different orientations. Third, since subjects had no experience with the characters until participating in an experiment, they were able to control which orientations subjects were familiar with.

The paradigm used by Tarr and Pinker was predicated on the different predictions each of the theories of object recognition makes for recognition times. Reviewing briefly, for any particular object, viewpoint-independent theories predict equivalent recognition times across all orientations; multiple-views theories predict equivalent recognition times across all *familiar* orientations (and failure at unfamiliar orientations); single-view theories predict recognition times monotonically dependent on the distance between the current orientation and the canonical one; and multiple-views-plus-transformation theories predict recognition times monotonically dependent on the distance between the current orientation and the *nearest familiar* orientation.

The general paradigm and predictions may be summarized as follows. First, subjects were trained on the characters in a single "canonical" orientation (sometimes the upright and sometimes near the upright). Subjects then were given large amounts of practice naming characters in several "practice" orientations. In early practice trials viewpoint-independent theories predict no effect of orientation on recognition times, while all theories using transformations predict that recognition times will be monotonically dependent on the distance from the canonical orientation. With extensive practice, viewpoint-independent theories predict that orientation will still have no effect on recognition times (even if there were effects of orientation in early trials -- viewpoint-independent representations may be stored only with experience), the multiple-views and the multiple-views-plus-transformation theories also predict that orientation will have no effect on recognition times (because subjects will have stored representations at each familiar orientation), while the single-view-plus-transformation theory predicts that orientation will still be monotonically dependent on the distance from the canonical orientation. Finally, subjects were probed with the same characters in new "surprise" orientations. Tarr and Pinker (1989a) predicted that if diminished effects of orientation after extensive practice were due to subjects storing multiple viewpoint-specific representations during the practice phase, practice effects would not transfer to new orientations and large effects of orientation should be found for the surprise orientations -- however now with reaction times monotonically dependent on the distance from the nearest practice orientation. Alternatively, if diminished effects of orientation were due to subjects storing viewpoint-independent representations of characters during practice, practice effects should transfer to new orientations and no effect of orientation on recognition times should be found for either practice or surprise orientations.

2.3.2. Results and their implications for viewpoint dependency in shape recognition

Tarr and Pinker's major results may be summarized as follows:

- When subjects first had to recognize misoriented characters, the time they required was generally monotonically related to the rotation of the characters from the upright orientation at which they had learned them. This was true despite the fact that no handedness discrimination was required.

- With practice at recognizing the characters in particular orientations, subjects' recognition times became roughly equivalent across all practiced orientations.
- Following training or practice with a character in a specific handedness in particular orientations, recognition times increased with differences in orientation between the stimulus character and a stored orientation, which was most often the nearest familiar orientation.
- Orientation dependence in early practice and surprise trials can probably be attributed to the use of "mental rotation", because the slope of the recognition time functions, estimating the rate of mental rotation, were consistently close to the slopes found in previous mental rotation studies, including ones that used converging techniques to demonstrate the analogue nature of the rotation process, and even closer to the slopes found in a handedness discrimination experiment, where the same stimuli were used in a task that uncontroversially requires mental rotation.

These results are inconsistent with the hypothesis that complex shape recognition is accomplished by matching viewpoint-independent representations such as object-centered descriptions. Such a hypothesis would predict that diminished effects of orientation that come with practice at recognizing shapes at a specific set of orientations should transfer to new orientations, which they do not. Moreover, these results falsify the conjecture that mental rotation is used only when the task requires discriminating handedness. Not only did some of Tarr and Pinker's tasks not require handedness to be assigned, but it is unlikely that subjects surreptitiously tried to determine handedness: one of their tasks made handedness irrelevant in principle by equating standard and reversed patterns, yet rotation to the nearest practice orientation still occurred. Overall these findings suggest that human behavior in complex shape recognition is best accounted for by the multiple-views-plus-transformation theory whereby recognition can be achieved by aligning an input shape with a representation of the shape at one of several stored orientations. Mismatches in orientation are compensated for by a process which requires more time for greater amounts of mismatch, presumably the continuous image transformation process called mental rotation.

2.3.3. Problems for extending the multiple-views-plus-transformation mechanism to three dimensions

One limitation of Tarr and Pinker's (1989a) findings is that their experiments used only two-dimensional stimuli, composed of only of line segments, and rotated only in the frontal plane around the line of sight axis. Thus, Tarr and Pinker present little evidence that people are capable of recognizing three-dimensional (and perhaps even two-dimensional) objects at arbitrary viewpoints in three dimensions by rotating them in depth to the nearest stored view. It is unresolved whether their theory is generalizable to three-dimensional object recognition. In this context, one possible criticism of their theory is that computing the appropriate axis for a single linear rotation may be far more difficult when the axis is not the line of sight. If this is true and people are not able to readily compute the axis of rotation for rotations out of the frontal plane, then the multiple-views-plus-transformation theory is of limited use for object recognition under natural conditions (particularly since objects tend to maintain a constant orientation with respect to gravity around the line of sight).

An alternative account of Tarr and Pinker's results is that rotations in the frontal plane along the line of sight are a special case of misorientation that the default recognition system is incapable of handling. Suppose, for instance, that misoriented objects in all orientations up to, but not including, rotations in the frontal plane, are recognized by matches to stored viewpoint-independent models that represent objects with an explicit top and bottom (see for example, Biederman, 1987, In Press; or Marr, 1982). Such representations certainly would be ecologically valid -- Rock (1973; 1983) has demonstrated that objects are often most easily recognized when seen from their natural orientation with respect to gravity, while Tarr and Pinker (1989a) point out that gravity is an important force affecting objects above a certain size (Haldane, 1927/1985) and that many objects exhibit just such a natural orientation. Thus, stored viewpoint-independent models encoding an explicit top and bottom would be adequate for recognition across most changes in viewpoint, but not for the ecologically rare changes in the location of top and bottom with respect to gravity. Fortunately, this alternative theory is easily tested by observing the effects of rotations in depth on the time to recognize three-dimensional objects -- a manipulation common to all of the experiments presented in this thesis.

2.3.4. Evidence for the use of the multiple-views-plus-transformation mechanism in three-dimensional object recognition

There is some evidence suggesting that a multiple-views-plus-transformation theory may be extended to three-dimensional object recognition. First, in Tarr and Pinker's (1989a) study, subjects familiar only with the normal handedness versions of two-dimensional figures recognized mirror-reversed handedness versions of the same figures by rotating them in depth around an axis contained within the frontal plane. This rotation is the shortest path for aligning the mirror-reversed version of a two-dimensional shape with a normal version at any orientation (see Parsons, 1987a, b). Subjects' willingness to rotate in depth to find a match suggests that humans normally compute the shortest path rotation, regardless of the axis of rotation, even when the shapes themselves are two-dimensional.

Secondly, Parsons (1987a, b) has demonstrated that not only do people rotate through the shortest path in depth to align mirror-image pairs, but that a single mechanism is sufficient for computing both rotations in the picture plane and three-dimensional shortest path rotations. Further, Parsons (1987c) has shown that for same/different discriminations of three-dimensional objects separated by rotations in depth, subjects generally mentally rotate through the shortest path in three dimensions. Thus, it appears that there exists a well documented visual mechanism sufficient for aligning input shapes of three-dimensional objects with stored views of three-dimensional objects, and that this mechanism operates efficiently, computing the shortest possible three-dimensional rotation.

Third, as first pointed out by Palmer, Rosch, and Chase (1981), many real-world objects appear to have a canonical orientation -- a privileged viewpoint in which an object is most familiar and from which it is most commonly described. Further, some objects have more than one canonical view. For instance faces seem to have canonical views head-on and on-profile. The existence of canonical views, in particular multiple views per object, suggests that people do store privileged representations at specific viewpoints. Whether these representations are the same as those used in recognition is an open question, addressed in the experiments presented in this thesis.

Despite this positive evidence, it has not yet been shown conclusively that three-dimensional object recognition is viewpoint dependent. The studies presented in this thesis are designed to test this explicitly in two ways. First, in general, whether the recognition of three-dimensional objects at varying viewpoints in three-dimensional space is viewpoint dependent; and second, more specifically, whether the multiple-views-plus-transformation theory provides a plausible account for this viewpoint dependency.

3. Experiments

3.1. General paradigm

The major aim of this study is to investigate the effects of orientation on three-dimensional object recognition. Tarr and Pinker's (1989a, b) results established that human performance in complex two-dimensional shape recognition tasks is best accounted for by the multiple-views-plus-transformation theory. The experiments presented here are based on the hypothesis that the multiple-views-plus-transformation theory accounts for human performance in three-dimensional object recognition as well. To test this hypothesis the experiments proposed in this thesis are designed as direct three-dimensional counterparts to Tarr and Pinker's (1989a, b) experiments. Although the stimuli and display orientations have been adapted to three dimensions, the basic paradigm and experimental manipulations have been preserved. All of the experiments to be run in this study use a training/practice/surprise paradigm outlined as follows:

- *Training phase.* Subjects are trained on a set of novel objects that are displayed in a single training orientation. Each of the trained objects is given a name, which corresponds to the object and its mirror-reversed image (enantiomorph). Training consists of subjects performing a constructive copying task in which each object is built in this training orientation several times from an example presented in the same orientation. Subjects then recall the objects, again by building them in the training orientation, according to the associated name until they reach the criterion of successfully recalling all target objects twice in succession.
- *Practice phase.* Subjects run through a set of trials in which they name the target objects presented in several different "practice" orientations as quickly as possible, including the training orientation, on a computer display. In addition, subjects must differentiate between the target objects and distractor objects which are drawn from the same set and appear in the same orientations. Subjects respond to target objects by pressing a named key and to distractors by pressing a footpedal. Subjects are given immediate feedback if their response is incorrect.
- *Surprise phase.* After extensive practice (sometimes thousands of trials) and without warning, subjects are presented with both the target and the distractor objects in new, never-before-seen "surprise" orientations as well as the familiar practice orientations. Other than the introduction of new orientations, the trials in this phase are identical to the trials in the practice phase.

3.2. General predictions

To review briefly, using this paradigm Tarr and Pinker (1989a) obtained the following general results: an effect of orientation from the training orientation on the time to recognize target objects in practice orientations in early practice trials; a progressive decrease with practice in the effect of orientation on the time to recognize target objects; and a reemergence of the effect of orientation, this time in terms of the distance from the nearest practice orientation, on the time to recognize target objects in surprise orientations. Tarr and Pinker argue that while many theories of shape recognition, including viewpoint-independent theories, may account for the first two results, only the multiple-views-plus-transformation theory predicts rotation to the nearest stored view for familiar objects presented in unfamiliar orientations.

Similar patterns of results are predicted by the multiple-views-plus-transformation theory for rotations in depth. Initially it is predicted that recognition times will increase with increasing shortest path angular distances from the training orientation of the target (misorientations were generated by rotation around only a single major axis, meaning that the shortest path to a stored view will always be a rotation around the same major axis used to generate the misorientation). With practice, it is predicted that recognition times will become near equivalent at all practice orientations.[3] Finally, it is predicted that recognition times for target objects in surprise orientations will increase monotonically with greater distances from the nearest familiar view (again the shortest path around a single major axis).

In the context of rotation to nearest stored view being the crucial predicted result of these experiments, the predictions of two alternative theories should be considered. Viewpoint-independent theories, such as those proposed by Biederman (1987), Corballis (1988), and Marr and Nishihara (1978), predict no effect of orientation on recognition times for surprise orientations arising from

---

[3] Even with massive amounts of practice the recognition functions of subjects in Tarr and Pinker's (1989a) experiments failed to flatten out completely, always displaying a small residual slope, possibly due to orientation-dependent processes for locating the main axis of shapes.

rotations in depth. These alternative theories also predict no effect of orientation on recognition times for orientations arising from rotations in the frontal plane, while a second alternative viewpoint-independent theory, in which the top and the bottom of objects are explicitly defined, predicts an effect of orientation for rotations in the frontal plane because target objects will be misaligned with the gravitational upright. Although Tarr and Pinker's (1989a) results suggest that these alternative predictions are unlikely to be confirmed, the restriction of their experiments to two-dimensional shapes and to picture plane rotations leaves open the question of whether similar three-dimensional objects are recognized by mechanisms other than the multiple-views-plus-transformation theory.

3.3. Experiment 1

The purpose of Experiment 1 is two-fold: to provide a baseline measure of orientation dependency in making judgments about a particular set of objects; and to examine whether subjects store handedness-specific and viewpoint-specific views for the determination of handedness. Handedness judgments are known to provide robust orientation-dependent reaction time functions in both two and three dimensions that commonly are attributed to mental rotation (Cooper and Shepard, 1973; Metzler and Shepard, 1974). Therefore, the reaction time functions and their slopes from this experiment will be useful for comparison with the results of subsequent experiments, particularly for assessing the effects of mental rotation. Secondly, this experiment provides a measure of the rate of decrease of orientation dependency with practice. These practice effects may be accounted for by the storage of representations that encode handedness in either a viewpoint-independent format or in a viewpoint-dependent format -- this experiment also provides a test of the viewpoint-specificity of these stored representations. If viewpoint-independent representations are stored, then judgments of handedness at unfamiliar orientations will take no longer than judgments at familiar orientations (or at least there will be no systematic effect of orientation on judgments at unfamiliar orientations). Alternatively, if viewpoint-dependent representations are stored, then judgments of handedness at unfamiliar orientations will take increasingly longer with increasing angular distance from familiar orientations. However, this experiment alone does not provide adequate evidence for viewpoint-specific representations in recognition. It is possible that viewpoint-independent representations cannot code for handedness version (Corballis, 1988; Corballis et. al., 1978; Hinton and Parsons, 1981). Under this interpretation, handedness-free viewpoint-independent representations normally used in recognition may not be used for judging handedness. Rather, handedness might be determined by using specialized handedness-specific and, most likely, viewpoint-dependent representations that are stored only to perform this particular task.

3.3.1. Method

Subjects. Twelve students from the Boston area participated in the experiment for pay. In this and subsequent experiments no subject participated more than once in any condition or experiment reported in this thesis or in any other paper by the author.

Materials. The stimuli consisted of seven left/right and front/back asymmetrical objects described in Appendix 1 and illustrated in Figure 1 in their training orientation of a 10° rotation around each axis.[4] Both the standard and reversed versions (enantiomorphs) of an object were used. Stimuli were drawn in 34 orientations (+30° steps around each axis, with rotations around the other two axes fixed at 10° each) on a color CRT with a resolution of 512 x 512 pixels. All rotations were around the center point of the imaginary box defined by the farthest reaching points of the object. The CRT was approximately 38 cm from a chin rest and the objects were drawn in a 13 cm diameter circular area centered within the screen, resulting in a diameter of 19.4 deg of visual angle. The surfaces of the objects were colored a uniform blue with no shading due to lighting and the edges of the faces of each cube comprising an object were colored

---

[4] All rotations are reported in degrees and are measured as follows: x axis rotations starting from the upper vertical and increasing with rotation towards the observer, y axis rotations starting from the right horizontal and increasing with rotation towards the viewer, and z axis rotations starting from the upper vertical and increasing in a clockwise direction (see Figure 2).

red with hidden lines removed. To guard against the idiosyncratic effects of a particular stimulus, the objects were grouped for counterbalancing purposes into three sets of three named objects each: set A was composed of objects 1, 2, and 3; set B of objects 4, 5, and 6; and set C of objects 2, 5, and 7. The first object of each set was named "Kip", the second was named "Kef", and the third was named "Kor". Each of these sets was presented to one third of the subjects, who were not aware of the groupings.

Insert Figures 1 and 2 about here.

Procedure. Subjects were shown both standard and reversed versions of the three objects that were members of the assigned set during the preliminary training session. Subjects were shown the three objects in the assigned set in each version on a computer display. To help them learn the objects, subjects duplicated each version of each object five times by constructing the displayed object out of toy blocks that connected at right angles to each other and to a prebuilt main axis common to all of the objects[5] and connected to a stationary base at the training orientation. For each duplication the subject also was instructed to repeat the name of the object aloud. Subjects were then asked to use the same toy blocks to build from memory a particular version of the object named by the experimenter. Feedback was given and subjects continued to build the objects named until they twice correctly had built all three objects in both standard and reversed versions in sequence.

Throughout the rest of the experiment the objects were shown one at a time on the CRT. Subjects were told that they were to wait for a fixation point (a "+") and then would see one of the objects displayed in one of many orientations. They were instructed to decide as quickly as possible, while minimizing errors, whether it was a standard or reversed version of one of the objects they had learned in the training session. Subjects responded via a labeled two-key

---

[5] Seven blocks stacked vertically with one block attached to either side of the bottom block.

response board with the standard response corresponding to the right key and the reversed response corresponding to the left key. Left-handed subjects had the option of reversing the response board, so that the standard response corresponded to their preferred hand. On trials where subjects made an error, they heard a beep.

Design. Both standard and reversed versions of the three objects were displayed in the four orientations illustrated in Figure 3: the training orientation of 10° around each axis and rotations of 130° around the x, y or z axis. The first part of the experiment consisted of "practice" blocks of trials. Each practice block contained 6 preliminary trials, randomly selected across conditions, followed by 72 trials corresponding to each of the 3 trained objects in their standard and reversed versions in the training orientation 6 times and in each of the other three practice orientations 2 times. In the second part of the experiment trials were organized into a "surprise" block consisting of 6 random preliminary trials, followed by 432 trials corresponding to each of the 3 objects in their standard and reversed versions in the training orientation 6 times and in each of 33 orientations 2 times defined by +30° increments starting at +10° around the x, y, and z axes.

In the surprise block the 6 preliminary trials were composed of only orientations previously used in practice blocks. In all blocks the order of the trials following the preliminary trials was determined randomly for each subject. Subjects were given a self-timed break every 40 trials.

Insert Figure 3 about here.

Subjects were run in a total of four sessions each approximately one hour long. In the first session subjects were first given the training procedure and then were run in two practice blocks. Subjects were run in four practice blocks in both the second and third sessions. In the fourth session subjects were run in two practice blocks prior to the surprise block, to ensure that any effects in the surprise block were not due to a beginning-of-session effect. Not counting preliminary trials, each subject was run in a total of 26 trials for every object at a particular handedness and practice orientation and 78 trials for every object at a particular handedness in the training orientation.

3.3.2. Results

Incorrect responses and responses for the preliminary trials in each block were discarded and reaction time means for each orientation were calculated by block, averaging over all objects. For purposes of data analysis it is generally accepted that in most mental rotation studies clockwise and counterclockwise rotations of the same magnitude produce approximately equal reaction times because subjects rotate through the shortest picture plane path to the upright (see Shepard and Cooper, 1982). This assumption may be extended to include any rotations of equivalent magnitude *around the same axis of rotation*, whether or not the rotation is in the picture plane. Equidistant rotations around a common axis are expected to yield equivalent reaction times even when the target orientation of the rotation is not the upright and may be treated as equivalent. This assumption is supported by Tarr and Pinker's (1989a) finding in many of their experiments and in all of the experiments reported in this thesis that mean reaction times for equivalent orientations around a common axis fall near a single straight line. Thus, the effect of orientation may be characterized by regressing the reaction time means against the distance from a target orientation and calculating the slope, measured in milliseconds per degree, of the best fitting line determined by the method of least squares.

Figure 4 shows mean reaction times for the first (Block 1) and last (Block 12) practice blocks separately for each axis of rotation, collapsed over standard and reversed versions. Three separate ANOVA's, one for each axis of rotation, for data collapsed over blocks 1 to 12 with Version and Orientation as factors revealed significant main effects for Orientation for each axis (x axis: $F(1, 11) = 28.5, p < .01$; y axis: $F(1, 11) = 77.7, p < .01$; z axis: $F(1, 11) = 46.7, p < .01$), a significant main effect for Version only for the z axis ($F(1, 11) = 15.1, p < .01$), as well as a Version by Orientation interaction only for the z axis ($F(1, 11) = 5.2, p$

$< .05$).[6] Standard and reversed versions were collapsed in all subsequent analyses. In Block 1 regressing mean reaction times against the distance from the training orientation reveals a slope of 16.5 msec/deg (61 deg/sec; since there are only 2 practice orientations, the correlation between distance from the training orientation and mean reaction time is always 1.0) for rotations around the x axis, 12.2 msec/deg (82 deg/sec) for the y axis, and 10.6 msec/deg (94 deg/sec) for the z axis. As illustrated in Figure 5, over the next 11 practice blocks the combined slope continued to decrease for rotations around all three axes, with the slopes for Block 12 being 4.0 msec/deg (250 deg/sec) for the x axis, 4.1 msec/deg (244 deg/sec) for the y axis, and 3.3 msec/deg (303 deg/sec) for the z axis. Tarr and Pinker (1989a) argue that practice effects on slope are not due to a speeding up of mental rotation, but rather, are due to the absence of rotation on some trials where input shapes are matched directly against representations at familiar orientations. The finding in Experiment 1, shown in Figure 5, that there are initially different rates depending on the axis of rotation, but with practice there are no significant differences between rates for different axes supports this hypothesis. Separate ANOVA's for slopes from Blocks 1 and 12 revealed a significant effect of axis in Block 1 ($F(2, 22) = 4.1$, $p < .05$), but no significant effect in Block 12. If rotations were simply speeding up with practice, the significant differences between rates for axes should have been preserved over practice. Alternatively, the rates for all axes may have reached a common floor or maximal speed. However, a floor effect would also predict that increased skill in rotating, reflected as faster rotation, would transfer to unfamiliar orientations, which, as we shall see, it did not. Thus the effect of practice on rate of rotation cannot be accounted for by an improvement in a task-specific skill.

From Block 1 to Block 12 there was also a decrease in overall reaction times,

---

[6] Why do the rates of rotation sometimes differ between versions? One explanation is that enantiomorphs of an object rotated through the same paths will reveal opposite parts and visible surfaces. Assuming that the rate of three-dimensional rotation is dependent on which surfaces must be brought into view and occluded (as well as the complexity of the object and the axis of rotation), rates of rotation would not necessarily be equivalent for standard and reversed versions of an object.

reflected as a decrease in the intercept of the reaction time functions. In Block 13, the surprise block, the slopes for the practice orientations were: x axis, 6.0 msec/deg (167 deg/sec); y axis, 4.7 msec/deg (213 deg/sec); and z axis, 4.6 msec/deg (217 deg/sec). In contrast, the slopes for the surprise orientations in the surprise block, computed by averaging across means for orientations at equal distances from the nearest practice orientation, were: x axis, 9.6 msec/deg (104 deg/sec; the correlation between mean reaction time and degrees from the nearest practice orientation is .97); y axis, 11.3 msec/deg (88 deg/sec; r = .98); and z axis, 3.7 msec/deg (270 deg/sec; r = .62). These changes in slope with practice are illustrated in Figure 7 (also see Table 1). These estimates of the rate of rotation for surprise orientations may be somewhat underestimated. As illustrated in Figure 6, objects at surprise orientations, particularly between 40° and 100°, do not always appear to be rotated to the nearest practiced orientation. A post hoc estimate of the rate of rotation may be obtained by including only surprise orientations from 160° to 340° regressed against the distance to the nearest practiced orientation; this analysis yielded slopes of: x axis, 13.6 msec/deg (74 deg/sec); y axis, 11.7 msec/deg (85 deg/sec); z axis, 1.8 msec/deg (556 deg/sec). Another post hoc analysis, including only the surprise orientations from 40° to 100° regressed against the distance to the practice orientation *to which they appear to be rotated* (130° for the x and z axes, and 10° for the y axis), yielded similar estimates: x axis, 1.5 msec/deg (667 deg/sec); y axis, 12.8 msec/deg (78 deg/sec); z axis, 1.0 msec/deg (1000 deg/sec).

Insert Figures 4 and 5 about here.

The effects of practice were examined by three two-way ANOVA's, one for each axis of rotation, on data from all practice blocks (1-12) with Block Number and Orientation as factors. Significant main effects for Block Number for the x axis (F (11, 121) = 24.4, $p < .001$), y axis (F (11, 121) = 26.4, $p < .001$), and z axis (F (11, 121) = 21.7, $p < .001$) were reflected as an overall decrease in reaction times with practice. Significant main effects for Orientation across blocks for the x axis (F (1, 11) = 34.0, $p < .001$), y axis (F (1, 11) = 71.8, $p < .001$), and z axis (F (1, 11) = 47.2, $p < .001$) were found, as were significant interactions between Block and Orientation for the x axis (F (11, 121) = 8.7, $p < .001$), y axis (F (11, 121) = 5.1, $p < .001$), and z axis (F (11, 121) = 6.4, $p < .001$). These interactions indicated that the effect of orientation changed with practice, and as shown in the data, diminished with practice.

The patterns of reaction times in the surprise block separated by the axes of rotation are shown in Figure 6. Unlike most mental rotation studies, the reaction time curves for this and subsequent experiments do not increase with distance from the training orientation. Instead, as predicted, recognition times generally increased with the distance from the nearest practice orientation. This may be seen in the reaction time curves for all three axes of rotation over the range of surprise orientations from 160° to 340°. In most instances minima appear near the practice orientations of 10° and 130°. Peaks appear at the surprise orientations of 220° for the x axis, 280° for the y axis, and 190°/220° for the z axis. The deviations of these peaks from the "ideal" peak of 250°, the orientation furthest from a practice orientation, may be the result of occlusions, singularities, and other changes in the visible parts of objects with rotations in depth. Such factors are also the most likely explanation for the variations in the functions seen for the range of surprise orientations from 40° to 100°. To test this hypothesis, for each of the seven objects a rating of the degree of foreshortening and occlusion at each orientation was compiled for rotations around the x and y axes. These ratings were then used predictors along with distance from the nearest practice orientation and distance from the training orientation in multiple regression analyses on mean reaction times from Block 13 for each object. For both x and y axes, analyses for each object failed to show that the ratings accounted for a significant amount of the variance in mean reaction times. Thus, the hypothesis that some of the variation in reaction times in Block 13 is due to geometric features of the stimuli is not supported and cannot be accounted for by any other obvious explanation. However, the ratings from this analysis do indicate that x axis rotations pass through many more singularities and occlusions than y axis rotations, suggesting that relatively slower rates of rotation for the x axis may be accounted for by this hypothesis. Multiple regression analyses on mean reaction times from Block 13 with distance from the nearest practice orientation and distance from the training orientation as predictors[7] confirmed that for all three axes of rotation the distance from the nearest practice orientation accounted for a significant amount of the variance in reaction times (x axis, $F(1, 9) = 9.9$, $p < .02$; y axis, $F(1, 9) = 31.3$, $p < .001$; z axis, $F(1, 9) = 5.3$, $p < .05$). In addition, although not clearly visible in the reaction time curve, for rotations around the z axis (picture plane rotations) the distance from the training orientation also accounted for a significant amount of the variance ($F(1, 9) = 14.7$, $p < .01$), while the distance from the training orientation was not significant for either the x or y axes. This result is consistent with Tarr and Pinker's (1989a) finding that in some instances subjects rotate to the upright despite the presence of a nearer practice orientation. They suggest that the representation stored at upright is canonical and may "attract" misoriented input shapes to a greater degree than other practice orientations (also see Robertson, Palmer, and Gomez, 1987). This result may help explain the fast rate of rotation for the z axis cited above. Regressing mean reaction times to the nearest practice orientation may not capture the actual rate of rotation since the objects are sometimes rotated to the training orientation near the upright. A post hoc analysis was performed in which mean reaction times for surprise orientations for z axis rotations from Block 13 were regressed against the distance from the training orientation and yielded a slope of 7.4 msec/deg (135 deg/sec) -- a relatively slower rate of rotation than that found in prior analyses.

As shown in Table 2 error rates ranged from about 9-34% in Block 1 to about 1-4% in Block 12. In Block 13 error rates for surprise orientations ranged from about 1-23%. No evidence for a speed/accuracy tradeoff in recognition was found in any block. Rather, here and in subsequent experiments reaction times and

---

[7] The correlation between these predictors is .27

errors rates showed similar trends across orientations.

Insert Figures 6 and 7 about here.

Insert Tables 1 and 2 about here.

3.3.3. Discussion

In large part the results of Experiment 1 are consistent with the results of studies by Metzler and Shepard (1974), Shepard and Cooper (1982), and Parsons (1987c). In particular, three major findings were confirmed: for handedness judgments on misoriented objects, (1) reaction times increased with the distance of the object from its training orientation, (2) reaction time functions were roughly consistent with the rotation of the object through the shortest three-dimensional path to the target orientation, and (3) the rate of rotation, as measured by slope, varied with the apparent axis of the shortest path rotation. In addition, the results of Experiment 1 are consistent with recent findings in experiments by Tarr and Pinker (1989a). The most important of these are: for handedness judgments, (1) with extensive practice, reaction times become near equivalent at all familiar orientations, (2) these practice effects at familiar orientations *do not* transfer to unfamiliar orientations, and (3) objects in these new orientations appear to be rotated through roughly the shortest three-dimensional path to the *nearest familiar orientation*.

These results suggest that Experiment 1 fulfilled its objectives. First, it provides a baseline measure of the rate of rotation for a new set of three-dimensional objects similar to, but not identical to, the objects used in other mental rotation studies (Metzler and Shepard, 1974; Parsons, 1987c).[8] Second,

---

[8] It is also one of the first experiments in the literature to establish that mental rotation is used for handedness judgments of single presentations of three-dimensional stimuli rotated in depth, and in particular around an axis other than the vertical, as opposed to the same/different discriminations used in previous mental rotation studies (see also, Shepard and Metzler, 1988).

the results verify that subjects are capable of rotating this new class of objects in three dimensions through changes in visible surfaces, and that these rotations vary in rate with the axis of rotation. As shown in Table 1, in comparison to the rates of rotation obtained by Metzler and Shepard (1974), Parsons (1987c), and Shepard and Metzler (1988), each using similar three-dimensional objects composed of cubes, the rates of rotation obtained here were of roughly the same order of magnitude. Two discrepancies should be pointed out. First, the relative ordering of speed of rotation around axes (slowest to fastest: x, y, z) is the opposite of the ordering obtained by Parsons (slowest to fastest: z, y, x). Second, the absolute magnitude of rates of rotation is slower than that obtained in the only directly comparable experiment, Shepard and Metzler's (1985) One-Stimulus Condition, where subjects judged whether a singlely presented object was of standard or reversed handedness. All other previous studies involved a same/different comparison between two objects.

There are several possible explanations why the rates of rotation in this experiment differ in relative ordering across axes of rotation from the rates obtained by Parsons. A related question is why the rates of rotation in this experiment, where objects were presented singlely, were slower than those in the One-Stimulus Condition of Shepard and Metzler (1988).[9] The objects used here, while similar to the objects used by Parsons (1987c; and created by Metzler and Shepard, 1974) in being constructed of cubes, differ in their relative level of complexity. Folk and Luce (1987) have presented evidence that for difficult discriminations, the rate of mental rotation speeds up with decreasing complexity. The objects used in this study may have been more difficult to discriminate than those used in prior studies, consequently producing slower rates of rotation. In addition, the presence of different numbers of protruding parts to the sides or front to back may affect the speed of rotation. Depending on the configuration of each object, rotations around different axes might entail traversing either more or fewer unique views and/or singularities. An analysis

---

[9] Although the rates found in subsequent experiments in this thesis are closer to Shepard and Metzler's.

designed to test whether such properties influenced reaction times in this experiment failed to find any significant effects. However, this analysis did suggest that for the stimuli in this experiment there are far more singularities and occlusions around the x axis than around either the y axis or z axis (where there are none). As shown in Table 1, in this and subsequent experiments, x axis rotations are consistently slower than y or z rotations. One last explanation is that Metzler and Shepard's objects did not contain a well-defined bottom, possibly necessitating the use of an orientation-dependent mechanism to locate it before rotation.[10] Depending on the location of the bottom axis in a given orientation, rates of rotation may have been variably affected.

One curious result is that in Metzler and Shepard (1974), Parsons (1987c), Shepard and Metzler (1988), and this study, subjects rotate objects through new configurations of visible surfaces. This finding implies that they have some way of correctly predicting the appearance of an object from a viewpoint other than the displayed or trained viewpoints. However, Rock and Di Vita (1987) have found that for curved wire objects, subjects are incapable of predicting an object's appearance from new viewpoints. The difference here may be attributed to the high predictability of objects constructed of cubes placed only at right angles to each other. Thus, subjects in mental rotation experiments utilizing such objects are able to project the shapes of these objects in intermediate viewpoints between the observed orientation and the target orientation because of the simple geometric relations between the parts of the objects. This suggests that for each new configuration of parts, subjects must predict a new view of the rotating object. One speculation is that if this "mental prediction" takes a significant amount of processing time, then the larger number of new views traversed, the slower the rate of rotation.[11] However, Metzler and Shepard

---

[10] The combination of two additive processes, both orientation-dependent, in this case locating the bottom and mental rotation, produces cumulatively greater slopes because each process will contribute proportionally greater processing times for increasing misorientations.

[11] This speculation coincides with Koenderink and van Doorn's (1979) speculation that reaction times in mental rotation are a product of not only orientation, but also the number of intervening views.

(1974) present evidence indicating that the measured rate of mental rotation passing through singularities is not influenced by the number of singularities. Additionally, in this experiment an analysis to test this hypothesis also failed to find significant effects. Thus, currently there is no evidence to support the argument that rates of rotation are influenced not only by the axis of rotation, but by the number of intervening views that must be predicted -- a property that is dependent on the part configurations within each object.

Experiment 1, while concerned primarily with handedness judgments, also provides some evidence for a multiple-views-plus-transformation mechanism. Primarily, this comes from the finding that subjects seem to rotate objects appearing in unfamiliar viewpoints to familiar viewpoints. This suggests that with practice subjects store viewpoint-specific (and handedness-specific) models of objects. The alternative hypothesis, that with practice, subjects stored handedness-specific viewpoint-independent models, predicts that the departure of the unfamiliar orientations from familiar viewpoints should have had no effect on reaction times. Furthermore, the finding that subject's latencies increased with the distance from familiar viewpoints, with slopes of comparable magnitude to those found in early trials, suggests that the process they used for aligning input shapes with stored models was mental rotation. Similarly, Tarr and Pinker (1989a) found effects of orientation for both handedness discriminations and recognition judgments comparable to both early trials of their experiments and to previous studies that more clearly involved mental rotation. They used this evidence to argue that mental rotation was responsible for orientation effects in their experiments as well.

Is it possible to explain rotation to the nearest familiar viewpoint without appealing to the storage of viewpoint-specific representations? One hypothesis, drawn from Hinton and Parsons (1981) and Corballis (1988), is that objects are represented independent of both orientation and handedness. In performing handedness judgments, objects are first recognized independently of orientation, but the specific handedness version is absent, since stored models are ambiguous with regard to handedness. Because handedness is defined only in the observer's egocentric frame of reference, determining handedness version necessitates the additional step of aligning the object, via mental rotation, with this reference frame. However since representations of objects are independent of orientation, the orientation specificity observed at familiar orientations after extensive practice cannot be due to stored representations. One possibility is that it is not models of the objects that are being stored at familiar orientations, but rather additional reference frames in which left and right are defined. However, there is little evidence supporting the argument that people are capable of reorienting egocentric frames of reference (see Hinton and Parsons, 1981; and Robertson, et al., 1987).

In summary, the results of Experiment 1 establish that the set of asymmetrical three-dimensional objects may be rotated in three-dimensional space and that these rotations often appear to take the shortest path. Further, it replicates Tarr and Pinker's (1989a) finding that practice effects at familiar orientations do not transfer to unfamiliar orientations. Most importantly, Experiment 1 demonstrates that objects at these unfamiliar orientations appear to be rotated to the nearest familiar orientation with a rate of rotation of comparable magnitude to early trials and to other mental rotation studies. All of these findings are consistent with the multiple-views-plus-transformation theory.

3.4. Experiment 2

What role does mental rotation play in human object recognition? Experiment 1 demonstrates that mental rotation is necessary to make handedness discriminations on three-dimensional objects. However, this experiment alone does not rule out the alternative that objects are stored in a viewpoint-invariant format that does not encode handedness, relegating mental rotation to the circumscribed role of determining handedness. This hypothesis, which Tarr and Pinker (1989a) call the *rotation-for-handedness hypothesis*, proposes that mental rotation is needed only for resolving handedness, but not for recognition (Corballis, 1988; Corballis, et. al., 1978; Hinton and Parsons, 1981). Indeed, there is some evidence that judgments other than handedness, including naming misoriented shapes, occur independently of orientation (Corballis, et. al., 1978; Simion et. al., 1982; White, 1980). These results have been cited in support of the argument that human object recognition is object-centered, suggesting that mental rotation is unnecessary for successful recognition. Alternatively, Tarr and Pinker (1989a, b; see also Jolicoeur, 1985) present evidence for orientation dependence in complex object recognition -- recognition in which members from within a class of similar shapes must be differentiated. Their results may be cited in support of the argument that human object recognition is viewer centered, suggesting that mental rotation is necessary to align input shapes with stored views in order to achieve successful recognition.

One way to test these competing hypotheses is to have subjects simply name novel rotated objects -- a task which should not require explicit handedness information. Experiment 2 does this by replicating the basic experimental paradigm of Tarr and Pinker's (1989a) earlier two-dimensional recognition study in three dimensions. As in Experiment 1, there are three crucial results concerning the effects of orientation on the recognition of objects: when objects are novel, when objects are familiar and seen in familiar views, and when objects are familiar but seen in unfamiliar views. The multiple-views-plus-transformation theory predicts, as in Experiment 1, that there initially will be an effect of orientation, that this effect will diminish with practice, and that this effect will return for the same objects in unfamiliar views. In contrast, viewpoint-independent theories, which may have predicted orientation dependency in Experiment 1, predict that practice will lead to equivalent performance at all familiar orientations, and that there will be no effect of orientation for the same objects in unfamiliar orientations. Thus, the surprise phase of Experiment 2, in which familiar objects are presented in unfamiliar views, distinguishes among viewpoint-dependent and viewpoint-independent theories of recognition and extends the earlier findings of Tarr and Pinker (1989a) to three-dimensional objects and rotations in depth.

3.4.1. Method

Subjects. Twelve students from the Boston area participated in the experiment for pay.

Materials. The stimulus items, computer display, stimulus sets, and experimental conditions were identical to those used in Experiment 1.

Procedure. The training procedure was the same as in Experiment 1 except that subjects were never shown the reversed versions of the objects and the toy blocks used to build the objects were faceted along all edges (these commercially available blocks were used to obtain more rigid models). In addition, the four objects not used in the named set were presented during testing to subjects (the subjects were not shown these distractors during the training phase), at the same orientations as the three trained objects. The four distractor objects contained many of the same local features as the trained objects, but in different configurations. These distractors were included to minimize the possibility that subjects would find some local feature that uniquely identified the three objects in the training set.

Subjects responded via a three key response board with the left key labeled "Kip", the center key labeled "Kef", and the right key labeled "Kor". Subjects were told they could use either hand or both hands to respond. They were informed that the objects would appear in many orientations and that sometimes an object they had not been taught would be displayed. In this case they were to press a footpedal.

Design. Trials were organized into practice blocks consisting of 6 randomly selected preliminary trials, followed by 96 trials corresponding to the 3 named objects in their standard versions in the training orientation 12 times and 4 times in each of the other three orientations, and the 4 distractor objects in the training orientation 3 times and in each of the other three orientations once each. In addition, trials were organized into a surprise block consisting of 8 random preliminary trials, followed by 576 trials corresponding to each of the 3 objects in their standard versions in the training orientation 12 times and in each of 33 orientations 4 times, and the 4 distractor objects in the training orientation 3 times and in each of 33 orientations once each. In the surprise block the 8 preliminary trials were composed of only orientations previously used in practice blocks. In all blocks the order of the trials following the preliminary trials was determined randomly for each subject. Subjects were given a self-timed break every 53 trials. Subjects were run in a total of four sessions identical to those in Experiment 1.

3.4.2. Results

As in Experiment 1, incorrect responses and responses for the preliminary trials in each block were discarded and reaction time means for each orientation were calculated by block, averaging over all objects. Also as in Experiment 1, equidistant rotations from a practice orientation around a common axis are expected to yield equivalent reaction times. As before, the effect of orientation will be characterized by regressing the reaction time means against the distance from a target orientation and calculating the slope of the best fitting line.

A three-way ANOVA on data from Block 13 with Group, Axis, and Orientation as factors revealed no significant main effect or interactions for the Group factor, indicating that counterbalancing of target objects across subject groups can be disregarded in further analyses. Three additional two-way ANOVA's on data from Block 13 within each group (A, B, or C) separately with Target Object and Axis as factors revealed no main effect or interaction for Target Object, indicating that individual objects also may be disregarded in these analyses.

In Block 1 regressing mean reaction times against the distance from the training orientation reveals a slope of 13.2 msec/deg (76 deg/sec) for rotations around the x axis, 6.0 msec/deg (167 deg/sec) for the y axis, and 8.0 msec/deg (125 deg/sec) for the z axis. Over the next 11 practice blocks the slope continued to decrease for rotations around all three axes, with the slopes for Block 12 being 2.2 msec/deg (455 deg/sec) for the x axis, 1.7 msec/deg (588 deg/sec) for the y axis, and 2.4 msec/deg (417 deg/sec) for the z axis (see Figure 9). From Block 1 to Block 12 there was also a decrease in overall reaction times, reflected as a decrease in the intercept of the reaction time functions. In Block 13, the surprise block, the slopes for the practice orientations were: x axis, 3.4 msec/deg (294 deg/sec); y axis, 1.3 msec/deg (769 deg/sec); and z axis, 2.2 msec/deg (455 deg/sec). In contrast, the slopes for the surprise orientations in the surprise block, computed by averaging across means for orientations at equal distances from a practice orientation, were: x axis, 8.8 msec/deg (114 deg/sec; r = .99); y axis, 7.2 msec/deg (139 deg/sec; r = .98); and z axis, 5.7 msec/deg (175 deg/sec; r = .84). Again these estimates of the rate of rotation for surprise orientations may be somewhat underestimated. Figure 8 shows that the surprise orientations between 40° and 100° do not appear to be rotated to the nearest practiced orientation. Rather these orientations exhibit curves similar to the curves found in Experiment 1 over the same range of orientations. A post hoc estimate of the rate of rotation was obtained by including only surprise orientations from 160° to 340° regressed against the distance to the nearest practiced orientation; this analysis yielded slopes of: x axis, 13.1 msec/deg (76 deg/sec); y axis, 8.4 msec/deg (119 deg/sec); z axis, 5.9 msec/deg (169 deg/sec). Another post hoc analysis, including only the surprise orientations from 40° to 100° regressed against the distance to the practice orientation to which they appear to be rotated (130° for the x and z axes, and 10° for the y axis), yielded similar estimates: x axis, 6.6 msec/deg (152 deg/sec); y axis, 7.4 msec/deg (135 deg/sec); z axis, 4.8 msec/deg (208 deg/sec).

Insert Figure 8 about here.

The effects of practice were examined by three two-way ANOVA's, one for each axis of rotation, on data from all practice blocks (1-12) with Block Number and Orientation as factors. Significant main effects for Block Number for the x axis (F (11, 121) = 54.1, $p < .001$), y axis (F (11, 121) = 37.1, $p < .001$), and z axis (F (11, 121) = 33.0, $p < .001$) were reflected as an overall decrease in reaction times with practice. Significant main effects for Orientation across blocks for the x axis (F (1, 11) = 72.1, $p < .001$), y axis (F (1, 11) = 16.9, $p < .01$), and z axis (F (1, 11) = 30.3, $p < .001$) were found, as were significant interactions between

Block and Orientation for the x axis (*F* (11, 121) = 20.8, $p < .001$), y axis (*F* (11, 121) = 8.8, $p < .001$), and z axis (*F* (11, 121) = 5.4, $p < .001$). These interactions indicated that the effect of orientation changed with practice, and as shown in the data, diminished with practice.

The patterns of reaction times in the surprise block separated by the axes of rotation are shown in Figure 8. As in the surprise block of Experiment 1, recognition times generally increased with the distance from the nearest practice orientation. This may be seen in the reaction time curves for all three axes of rotation over the range of surprise orientations from 160° to 340°. As before minima appear at or near the practice orientations of 10° and 130°. Peaks appear at the surprise orientations of 250° for the x axis, 250° for the y axis, and 220° for the z axis. Only the peak for z axis rotations deviates from the "ideal" peak of 250°, the orientation furthest from a practice orientation. Interestingly, variations in the monotonicity of the functions, attributed in Experiment 1 to changes in the visible parts of objects with rotations in depth, fall at roughly the same orientations in Experiments 1 and 2. This was confirmed by three multiple regressions, one for each axis of rotation, on mean reaction times from Block 13 of Experiment 2 with the mean reaction times from Block 13 of Experiment 1 and distance from the nearest practice orientation as predictors. These analyses revealed that the variation in reaction times in Experiment 1, beyond that correlated with distance from the nearest practice orientation, accounted for a significant amount of the variance in reaction times in Experiment 2 for the x axis (*F*(1, 9) = 11.9, $p < .01$), y axis (*F*(1, 9) = 7.9, $p < .05$), and z axis (*F*(1, 9) = 11.4, $p < .01$). Variation in the distance from the nearest practice orientation uncorrelated with reaction times from Experiment 1 accounted for a significant amount of the variance in Experiment 2 for the z axis (*F*(1, 9) = 10.4, $p < .05$), while not being significant for the x or y axes. These results suggest that the reaction time functions display highly similar variations across orientation. In addition, as in Experiment 1, three multiple regression analyses on mean reaction times from Block 13 with distance from the nearest practice orientation and distance from the training orientation as predictors confirmed that for all three axes of rotation the distance from the nearest practice orientation accounted for a significant amount of the variance in reaction times (x axis, $F(1, 9) = 12.0, p < .01$; y axis, $F(1, 9) = 23.1, p < .001$; z axis, $F(1, 9) = 16.2, p < .01$), while the distance from the training orientation was not significant for any axis of rotation.

As shown in Table 3 error rates ranged from about 4-31% in Block 1 to about 1-3% in Block 12. In Block 13 error rates for surprise orientations ranged from about 1-26%. No evidence for a speed/accuracy tradeoff in recognition was found in any block.

Insert Figure 9 about here.

Insert Table 3 about here.

3.4.3. Discussion

The results of Experiment 2 almost perfectly replicate the results of Experiment 1 for recognition rather than handedness judgments. The major findings may be summarized as follows: (1) When subjects first encounter objects in orientations other than the training orientation, their recognition times increase monotonically with the angular distance from the training orientation; (2) with practice recognizing the objects in several orientations, subjects' performance becomes roughly equivalent at all familiar orientations; and (3) when subjects encounter the same objects in unfamiliar orientations, their recognition times increase with the angular distance from the nearest familiar orientation. Further, as shown in Table 1, the effect of orientation on recognition times for objects in unfamiliar orientations is comparable to that found in early trials of Experiment 2, to the effects found in Experiment 1, and to the effects reported in other mental rotation studies (Metzler and Shepard, 1974; Parsons, 1987c, Shepard and Metzler, 1988). These findings suggest that mental rotation is responsible for effects of orientation not only in handedness discriminations, but in recognition as well.

There is additional evidence in Experiment 2 that the same transformation process, mental rotation, accounts for orientation dependency in recognition and in handedness judgments. Not only did rates of rotation differ between axes in Experiment 1, but, as shown in Figure 6, the shapes of the reaction time functions differed between axes (although they all displayed the similar general trend of increasing with distance from a familiar orientation). Similar variations in reaction time functions may be observed in Experiment 2 (Figure 8) and are found to be predicted by the reaction times from Experiment 1. Logically, both handedness judgments and recognition should exhibit similar effects of orientation for each axis if the same mechanism is used to compensate for misorientation in both cases. This is found in Experiments 1 and 2, implicating a role for mental rotation in recognition. Importantly, this role is no longer limited to picture plane rotations of two-dimensional objects. Experiment 2 establishes that not only is mental rotation involved in recognition, but that objects are rotated in three dimensions, including rotations through new configurations, and that these rotations often take the shortest path to target orientations (at least for rotations around the major axes tested here).

The results of Experiment 2 also provide evidence that subjects store multiple views of objects and that recognition from unfamiliar views is accomplished by shortest path rotations to these stored views. The effect of orientation in early trials suggests that subjects have stored viewpoint-specific models of the objects at the training orientation and recognize objects in unfamiliar practice orientations by rotating them to this "canonical" orientation. Subject's more nearly equivalent performance with practice at familiar orientations suggest that they have stored either multiple viewpoint-specific viewer-centered models or a single viewpoint-invariant model. The crucial finding is that in the surprise block, objects appearing in unfamiliar views exhibit reaction time functions that are best predicted by distance from the nearest familiar view, rather than either distance from the training view or no effect of orientation. This finding is inconsistent with both the single-view-plus-transformation theory and viewpoint-independent object-centered theories. In contrast, this finding is consistent with the multiple-views-plus-transformation theory, suggesting that complex three-dimensional object recognition is viewpoint-dependent and that objects are represented as a collection of views corresponding to familiar orientations of objects.

Finally, another implication of Experiment 2 is that the rotation-for-handedness hypothesis is discredited -- in the recognition task in this experiment handedness was irrelevant (since subjects only saw one version of an object), yet reaction times still exhibited viewpoint dependency. Thus, mental rotation is not used exclusively for the discrimination of handedness. In addition, because similar effects of orientation were found for rotations that preserve the position of an object's top and bottom with respect to the upright, the use of mental rotation is not limited to those "ecologically unusual" misorientations of objects where an object's normal orientation with respect to gravity must be located (see Rock, Di Vita, and Barbeito, 1981). Taken together, these interpretations suggest that orientation dependency in recognition is not a "special case" limited to uncommon tasks or unusual orientations, but rather the result of general mechanisms for complex object recognition.

3.5. Experiment 3

Tarr and Pinker (1989a) point out that it is possible for the rotation-for-handedness hypothesis to accommodate findings of viewpoint-dependent recognition in experiments involving only one handedness version per object. Such was the case in Experiment 2, where subjects were trained, practiced, and surprised with only the "standard" version of each stimulus item. Under this account (suggested by Corballis, 1988; Corballis, et. al., 1978; and Hinton and Parsons, 1981), objects are represented as viewpoint-independent models that are ambiguous with regard to handedness version and mental rotation is used to determine handedness surreptitiously on the chance that the version of the object will be incorrect for the recognition task. Thus, although in Experiment 2 subjects recognized objects independently of orientation, they rotated objects into a frame of reference, either egocentric or stored at a familiar viewpoint, where left and right were defined. This is because subjects are conservative; despite never having seen a reversed version of any object, the possibility of a reversed version occurring and matching incorrectly against the target standard version prompted them to verify the standard handedness of each stimulus object.

Tarr and Pinker (1989a) addressed this criticism by designing a recognition task in which handedness was explicitly irrelevant, informing subjects of this fact, and giving them practice treating mirror pairs equivalently. Specifically, both handedness versions of a shape were given the same name and classified as a single shape in both training and recognition (they point out that this is analogous to learning the label "glove" in connection with a right-hand glove, and applying it to left-hand gloves as well). Here, since handedness is by definition irrelevant to the task, subjects rotating only to verify handedness version should no longer do so. However, Tarr and Pinker found that this was not the case -- in several experiments where handedness was irrelevant, in both early trials of the experiment and in later trials when shapes appeared in unfamiliar orientations, subjects recognized shapes by rotating them. These results are interpreted as supporting the multiple-views-plus-transformation theory and further disconfirming the rotation-for-handedness hypothesis. Thus, since similar criticisms may be levied against Experiment 2, Experiment 3 addresses these criticisms in a similar fashion, making handedness explicitly irrelevant by assigning the same name to both standard and reversed versions of each target object.

Tarr and Pinker (1989a) also uncovered an effect of training that lends further support to the multiple-views-plus-transformation theory. In their experiments, although subjects were informed that the same name referred to both versions of a figure and initially subjects were shown both versions, during training subjects only learned the standard -- presumably storing representations of the shapes at the training orientation only in this standard handedness. Tarr and Pinker found that for reversed versions of the shapes in early trials, seen extensively for the first time, reaction time functions appeared flat, indicating equivalent performance in recognizing untrained reversed shapes at any orientation. Rather than proposing an orientation-invariant mechanism, they argue that subjects are rotating or "flipping" reversed shapes in depth around an axis lying within the picture plane. Such a rotation is the shortest path to align a mirror-reversed two-dimensional shape in any picture plane orientation with its standard counterpart (see Parsons, 1987a, b). Further, this rotation will always be exactly 180°, producing equivalent reaction times for all possible misorientations of reversed shapes. Such a rotation is unlikely to occur as a result of the need to determine handedness, since a shorter picture plane rotation will align the reversed shape with the upright egocentric frame of reference. However, this is precisely the sort of rotation predicted by a theory in which mirror-reversed input shapes must be aligned with the nearest matching stored shape; a prediction made only by the multiple-views-plus-transformation theory.[12]

Unlike two-dimensional shapes there is no rigid three-dimensional transformation that will align a three-dimensional object and its enantiomorph. At least two possible alignments do exist: a nonrigid deformation *pulling the object through itself*; and a rigid rotation in four-dimensional space. In both cases it seems unlikely that such transformations are commonly used or even computable within the visual system. Thus, it seems improbable that untrained reversed versions of three-dimensional objects will be recognized in constant time. Instead, reversed versions might be aligned with stored standard versions by rotating them into an orientation of maximum congruence and then comparing non-congruent parts to see if they appear in exactly opposite positions. Experiment 3 addresses this hypothesis by manipulating training

---

[12] A control experiment in which subjects were trained on both standard and reversed versions produced increasing functions with distance from the upright for both versions, suggesting that when subjects are given the opportunity to store reversed versions, they no longer use a flip in depth to recognize them.

between two conditions: the *Both-Versions Condition*, in which subjects are trained on both versions; and the *One-Version Condition*, in which subjects are trained only on standard versions (although they are still shown reversed versions during training, as well as throughout the rest of the experiment). One possible result of this manipulation is that subjects will again exhibit equivalent performance for all untrained reversed versions. Such a finding would suggest that, contrary to the "flip in depth" hypothesis, subjects are utilizing a viewpoint-independent recognition mechanism here and in Tarr and Pinker's (1989a) two-dimensional experiments. In contrast, the multiple-views-plus-transformation theory makes the same predictions for both conditions -- increasing reaction times with distance from the training orientation for both standard and reversed versions, regardless of training condition. Such a finding would support the "flip in depth" hypothesis, suggesting that here subjects rotated reversed versions to the training orientation and in Tarr and Pinker's two-dimensional experiments in depth to a standard version.

Overall Experiment 3 has three major purposes: first, to address the possibility that subjects rotated in Experiment 2 to verify handedness, despite its irrelevancy; second, to examine the prediction of the multiple-views-plus-transformation theory that the recognition of untrained reversed three-dimensional objects should exhibit orientation dependency (in contrast to similar two-dimensional objects); and third, to provide another test of the general predictions of the multiple-views-plus-transformation theory, that there will be an effect of orientation in early trials, that this effect will diminish with practice, and that this effect will return for unfamiliar orientations.

3.5.1. Method

Subjects. Twenty-four students from the Boston area participated in the experiment for pay: 12 in the first condition and 12 in the second condition.

Materials. The stimulus items, computer display, stimulus sets, and general experimental conditions were identical to those used in Experiment 2.

Two conditions were employed. In the first, the Both-Versions Condition, the practice (10° and 130°) and surprise orientations were identical to those used in

Experiment 2. In the second, the One-Version Condition, the objects were displayed in 13 new practice orientations (the training orientation and 40°, 70°, 100°, and 190° around each axis) and in the same 34 surprise orientations used in Experiment 2. New practice orientations were introduced primarily to examine the effects of a decreased range between practice orientations and were not part of the manipulation of training between conditions.

Procedure. The training procedure was the same as in Experiment 2 except that subjects were shown the reversed versions of the objects as well as the standard versions and were told that the name of an object referred to both versions.

In the Both-Versions Condition, subjects were trained on both versions of each named object, including duplicating and then building both versions from memory. In the One-Version Condition, subjects were shown both versions of each named object, but only duplicated and then built the objects in their standard versions.

Design. In the Both-Versions Condition trials were organized into practice and surprise blocks similar to those used in Experiment 2. However, trials for the 3 named objects were divided equally between standard and reversed versions. Trials for the distractor objects were similarly divided, except that to preserve the 3:1 ratio of named objects to distractors, in practice blocks each distractor was presented in one version at the training orientation and the other version at the 130° orientations, while in the surprise block each distractor was presented in one version at 60° increments beginning at 10° (10°, 70°, ...) and the other version at 60° increments beginning at 40° (40°, 100°, ...). Which version was presented at even or odd orientations was also varied by the axis of rotation. Named objects were still presented in both versions at all orientations. Subjects were given a self-timed break every 53 trials. Subjects were run in a total of four sessions identical to those in Experiment 1.

In the One-Version Condition trials were organized into practice blocks consisting of 6 randomly selected preliminary trials, followed by 240 trials corresponding to the 3 named objects in the training orientation 6 times in their standard version and 6 times in their reversed version and 2 times in each version in each of the other 12 orientations, and the distractor objects in an additional 60 trials. As in the Both-Versions Condition, to preserve the ratio of named objects to distractors, each distractor was presented in one version at some practice orientations and the other version in the remaining practice orientations. Surprise blocks were identical to those used in the Both-Versions Condition. Subjects were given a self-timed break every 62 trials. Subjects were run in a total of four sessions identical to those in Experiment 1.

3.5.2. Results

In both conditions incorrect responses and responses for the preliminary trials in each block were discarded and reaction time means for each orientation were calculated by block, averaging over all objects.

In the Both-Versions Condition the slopes for Block 1 collapsed over standard and reversed versions were: x axis, 9.6 msec/deg (104 deg/sec); y axis, 5.7 msec/deg (175 deg/sec); and z axis, 7.2 msec/deg (139 deg/sec). Computed separately, the slopes for standard versions were: x axis, 6.3 msec/deg (159 deg/sec); y axis, 4.3 msec/deg (233 deg/sec); and z axis, 6.0 msec/deg (167 deg/sec); and the slopes for reversed versions were: x axis, 13.0 msec/deg (77 deg/sec); y axis, 7.1 msec/deg (141 deg/sec); and z axis, 8.4 msec/deg (119 deg/sec). ANOVA's on data from Block 1 with Orientation and Version as factors revealed, as illustrated in Figure 10a, no significant main effects for Version or significant interactions between Orientation and Version for any axis of rotation with the exception of a significant Orientation by Version interaction for the x axis (F (1, 11) = 5.0, $p < .05$). In this instance the rates of rotation for both versions were still relatively slow with a slope for standard versions of 6.3 msec/deg and for reversed versions of 13.0 msec/deg. Three ANOVA's, one for each axis of rotation, for data from blocks 1 to 12 with Orientation and Version as factors revealed for the y axis only a significant main effect for Version (F (1, 11) = 6.8, $p < .05$) and a Version by Orientation interaction (F (1, 11) = 5.2, $p < .05$). For the x and z axes slopes broken down by standard and reversed versions did not vary significantly. Standard and reversed versions were collapsed in all subsequent analyses. Over the next 11 practice blocks the combined slope continued to decrease for rotations around all three axes, with the slopes for Block 12 being 1.1 msec/deg (909 deg/sec) for the x axis, 0.7 msec/deg (1429 deg/sec) for the y axis, and 1.1 msec/deg (909 deg/sec) for the z axis (see Figure 12). From Block 1 to Block 12 there was also a decrease in overall reaction times, reflected as a decrease in the intercept of the reaction time functions. In Block 13, the surprise block, the slopes for the practice orientations collapsed over version were: x axis, 1.7 msec/deg (588 deg/sec); y axis, 0.2 msec/deg (5000 deg/sec); and z axis, 1.4 msec/deg (714 deg/sec). In contrast, the combined slopes for the surprise orientations in the surprise block, computed by averaging across means for orientations at equal distances from a practice orientation, were: x axis, 3.9 msec/deg (256 deg/sec; $r = 1.0$); y axis, 5.4 msec/deg (185 deg/sec; $r = .99$); and z axis, 4.3 msec/deg (233 deg/sec; $r = .92$). As in previous experiments these estimates of the rate of rotation for surprise orientations may be somewhat underestimated. Figure 11 shows that the surprise orientations between 40° and 100° do not appear to be rotated to the nearest practiced orientation. Rather these orientations exhibit curves similar to the curves found in Experiments 1 and 2 over the same range of orientations. A post hoc estimate of the rate of rotation was obtained by including only surprise orientations from 160° to 340° regressed against the distance to the nearest practiced orientation; this analysis yielded slopes of: x axis, 6.7 msec/deg (149 deg/sec); y axis, 6.4 msec/deg (156 deg/sec); z axis, 4.8 msec/deg (208 deg/sec). Another post hoc analysis, including only the surprise orientations from 40° to 100° regressed against the distance to the practice orientation to which they appear to be rotated (130° for the x and z axes, and 10° for the y axis), yielded similar estimates: x axis, 0.9 msec/deg (1111 deg/sec); y axis, 5.6 msec/deg (179 deg/sec); z axis, 4.5 msec/deg (222 deg/sec).

Insert Figures 10, 11, and 12 about here.

The effects reported for the Both-Versions Condition were replicated in the One-Version Condition, where the slopes for Block 1 collapsed over standard and reversed versions were: x axis, 2.4 msec/deg (414 deg/sec); y axis, 2.1 msec/deg (469 deg/sec); and z axis, 3.7 msec/deg (268 deg/sec). Computed separately, the slopes for standard versions were: x axis, 2.6 msec/deg (385 deg/sec); y axis, 3.0 msec/deg (333 deg/sec); and z axis, 4.0 msec/deg (253 deg/sec); and the slopes for reversed versions were: x axis, 2.2 msec/deg (449 deg/sec); y axis, 1.2 msec/deg (839 deg/sec); and z axis, 3.5 msec/deg (283 deg/sec). ANOVA's on data from Block 1 with Orientation and Version as factors revealed, as illustrated in Figure 10b, no significant main effects for Version or significant interactions between Orientation and Version for any axis of rotation. ANOVA's for data from blocks 1 to 12 with Orientation and Version as factors revealed a significant main effect for Version (F (1, 11) = 11.4, $p < .01$) and a Version by Orientation interaction (F (4, 44) = 4.1, $p < .01$) for y axis rotations, and a Version by Orientation interaction (F (4, 44) = 4.0, $p < .01$) for x axis rotations. Thus, slopes broken down by standard and reversed versions sometimes varied significantly from each other. However, these differences were unsystematic and do not appear to reflect the use of different mechanisms in the recognition of different versions. In particular, reaction times for both versions generally increased with distance from from the training orientation. Standard and reversed versions were collapsed in all subsequent analyses. Over the next 11 practice blocks the combined slope continued to decrease for rotations around all three axes, with the slopes for Block 12 being 1.0 msec/deg (1039 deg/sec) for the x axis, 0.3 msec/deg (2946 deg/sec) for the y axis, and 1.2 msec/deg (841 deg/sec) for the z axis (see Figure 13). From Block 1 to Block 12 there was also a decrease in overall reaction times, reflected as a decrease in the intercept of the reaction time functions. In Block 13, the surprise block, the slopes for the practice orientations collapsed over version were: x axis, 0.5 msec/deg (2000 deg/sec); y axis, 0.5 msec/deg (2000 deg/sec); and z axis, 1.4 msec/deg (714 deg/sec). In contrast, the combined slopes for the surprise orientations in the surprise block, computed by averaging across means for orientations at equal distances from a practice orientation, were: x axis, 5.2 msec/deg (192 deg/sec; r = .74); y axis, 6.2 msec/deg (161 deg/sec; r = .96); and z axis, 5.4 msec/deg (185 deg/sec; r = .76). As in previous experiments these estimates of the rate of rotation for surprise orientations may be somewhat underestimated. As in the Both-Versions Condition some surprise orientations (here at 100° and 160°) do not appear to be rotated to the nearest practiced orientation. Rather these orientations appear to be rotated to the practice orientation of 190°. A post hoc estimate of the rate of rotation was obtained by including only the surprise orientations from 220° to 340° regressed against the distance to the nearest practiced orientation; this analysis yielded slopes of: x axis, 6.7 msec/deg (149 deg/sec); y axis, 6.5 msec/deg (154 deg/sec); z axis, 6.0 msec/deg (167 deg/sec).

Insert Figure 13 about here.

In both conditions the effects of practice were examined by three two-way ANOVA's, one for each axis of rotation, on data from all practice blocks (1-12) with Block Number and Orientation as factors. For the Both-Versions Condition significant main effects for Block Number for the x axis (F (11, 121) = 44.1, $p < .001$), y axis (F (11, 121) = 39.8, $p < .001$), and z axis (F (11, 121) = 36.1, $p < .001$) were reflected as an overall decrease in reaction times with practice. Significant main effects for Orientation across blocks for the x axis (F (1, 11) = 46.7, $p < .001$), y axis (F (1, 11) = 32.4, $p < .01$), and z axis (F (1, 11) = 85.7, $p < .001$) were found, as were significant interactions between Block and Orientation for the x axis (F (11, 121) = 16.5, $p < .001$), y axis (F (11, 121) = 4.9, $p < .001$), and z axis (F (11, 121) = 19.1, $p < .001$). Similar effects were found for the One-Version Condition, where significant main effects for Block Number for the x axis (F (11, 121) = 53.2, $p < .001$), y axis (F (11, 121) = 45.9, $p < .001$), and z axis (F (11, 121) = 46.8, $p < .001$) were reflected as an overall decrease in reaction times with practice. Significant main effects for Orientation across blocks for the x axis (F (4, 44) = 16.9, $p < .001$), y axis (F (4, 44) = 12.6, $p < .01$), and z axis (F (4, 44) = 13.6, $p < .001$) were found, as were significant interactions between Block and Orientation for the x axis (F (44, 484) = 4.1, $p < .001$), y axis (F (44, 484) = 2.2, $p < .001$), and z axis (F (44, 484) = 2.3, $p < .001$). In both conditions these interactions indicated that the effect of orientation changed with practice, and as shown in the data, diminished with practice.

The patterns of reaction times in the surprise block separated by the axes of rotation are shown in Figure 11 for the Both-Versions Condition. As in the surprise blocks of previous experiments, recognition times generally increased with the distance from the nearest practice orientation. For the Both-Versions Condition this may be seen in the reaction time curves for all three axes of rotation over the range of surprise orientations from 160° to 340°. Minima appear at or near the practice orientations of 10° and 130°. Peaks appear at the surprise orientations of 250°/280° for the x axis, 250° for the y axis, and 220° for the z axis. Only the peak for z axis rotations deviates from the "ideal" peak of 250°, the orientation furthest from a practice orientation. Variations in the monotonicity of the functions, attributed in Experiment 1 to changes in the visible parts of objects with rotations in depth, fall at roughly the same orientations in Experiment 1 and the Both-Versions Condition of Experiment 3. This was partly confirmed by three multiple regressions, one for each axis of rotation, on mean reaction times from Block 13 of the Both-Versions Condition of Experiment 3 with the mean reaction times from Block 13 of Experiment 1 and distance from the nearest practice orientation as predictors. These analyses revealed that the variation in reaction times in Experiment 1, beyond that correlated with distance from the nearest practice orientation, accounted for a significant amount of the variance in reaction times in the Both-Versions Condition of Experiment 3 for the z axis ($F(1, 9) = 10.6, p < .01$), while not being significant for the x or y axes. Variation in the distance from the nearest practice orientation uncorrelated with reaction times from Experiment 1 accounted for a significant amount of the variance in Both-Versions Condition of Experiment 3 for the z axis ($F(1, 9) = 9.8, p < .05$), while not being significant for the x or y axes. These findings suggest that the reaction time functions display highly similar variations across orientation. In addition, three multiple regression analyses on mean reaction times from Block 13 with distance from the nearest practice orientation and distance from the training orientation as predictors confirmed that for all three axes of rotation the distance from the nearest practice orientation accounted for a significant amount of the variance in reaction times (x axis, $F(1, 9) = 10.4$, $p < .01$; y axis, $F(1, 9) = 22.6$, $p < .001$; z axis, $F(1, 9) = 17.5$, $p < .01$), while the distance from the training orientation was not significant for any axis of rotation.

In the One-Version Condition recognition times generally increased with the distance from the nearest practice orientation and, in particular, for all three axes of rotation over the range of surprise orientations from 220° to 340°. Minima are found at the practice orientations of 10° and 190°. Peaks generally are near the ideal midpoint of 280°. The surprise orientations of 100° and 160° which were interspersed with practice orientations seem to have been rotated to the practice orientation of 190° for all three axes. Multiple regression analyses on mean reaction times from Block 13 with distance from the nearest practice orientation and distance from the training orientation as predictors[13] confirmed that for all three axes of rotation the distance from the nearest practice orientation accounted for a significant amount of the variance in reaction times (x axis, $F(1, 9) = 20.1$, $p < .001$; y axis, $F(1, 9) = 75.2$, $p < .001$; z axis, $F(1, 9) = 30.9$, $p < .001$), while the distance from the training orientation was not significant for any axis of rotation.

As shown in Tables 4 and 5 error rates in both conditions ranged from about 5-23% in Block 1 to about 1-5% in Block 12. In Block 13 error rates for surprise orientations ranged from about 1-22%. No evidence for a speed/accuracy tradeoff in recognition was found in any block of either condition.

Insert Tables 4 and 5 about here.

---

[13] The correlation between these predictors is -.02

3.5.3. Discussion

The results of both conditions in Experiment 3 are consistent with the multiple-views-plus-transformation theory. Moreover, the same pattern of results found in Experiments 1 and 2 are found here. First, recognition times in early trials increased with the distance from the training orientation. Second, these effects of orientation diminished with practice until subjects' recognition performance was approximately equivalent at all familiar orientations. Third, the effects of orientation returned for unfamiliar orientations, but now increasing with the distance from the nearest familiar orientation. In particular, the reaction time functions for Block 13 of the Both-Versions Condition are similar in appearance to the reaction time functions for Block 13 of Experiment 1 where it is uncontroversial that mental rotation was used. In addition, as in Experiments 1 and 2, x axis rotations were slowest in both conditions of this experiment. As before, these findings are not only most consistent with a multiple-views-plus-transformation mechanism in object recognition, but confirm that the mechanism extends to three-dimensional objects rotated through shortest paths in three dimensions.

The results of the first manipulation unique to Experiment 3, the inclusion of both standard and reversed versions of an object, both assigned the same name, are also consistent with conclusions drawn from the other experiments. Specifically, they support the existence of a multiple-views-plus-transformation mechanism in that the rotation-for-handedness hypothesis cannot account for the systematic effects of orientation found for recognition tasks where handedness is explicitly irrelevant. Under these circumstances the rotation-for-handedness hypothesis predicts that there should be no effects of orientation since the surreptitious determination of handedness no longer compensates for possible incorrect matches. In contrast, the multiple-views-plus-transformation theory is consistent with effects of orientation since input shapes must be rotated to the nearest stored view of an object.

The results of the second manipulation unique to Experiment 2, the differential training of handedness versions in the Both-Versions and the One-

Version Conditions, once again support the existence of a multiple-views-plus-transformation mechanism. In contrast with earlier two-dimensional studies, both untrained and trained reversed versions of three-dimensional objects apparently were rotated to the training orientation. This rules out subjects' use of a viewpoint-independent recognition mechanism, since such a mechanism would be equally effective for untrained two- and three-dimensional objects. Rather, a viewpoint-dependent mechanism is used -- aligning two-dimensional shapes by a flip in depth to trained standard versions and three-dimensional objects by a shortest path three-dimensional rotation to the training orientation.[14]

3.6. Experiment 4

The manipulation of training in Experiment 3 is predicated on the assumption that during training subjects are storing viewpoint-specific representations of the objects at the training orientation. The primary evidence for this is that reaction times from early practice trials of Experiments 1, 2, and 3 increase with distance from the training orientation suggesting that misoriented objects are being rotated to this orientation. However, the representations stored during initial training may differ from those that develop with repeated exposure to objects -- this is particularly likely to be true if ultimately people store viewpoint-independent representations, but that they take some time to develop. The finding that diminished effects of orientation after practice do not transfer to objects in unfamiliar orientations suggests that this is not the case; both because this is evidence for the orientation-specificity of

---

[14] How do subjects recognize untrained reversed versions if no stored match is available? As suggested one hypothesis is that subjects rotate reversed versions into orientations of the greatest partial congruence with stored standard versions and then look for identical parts connected on opposite sides (at least one subject reported doing this). This leads to the prediction that overall recognition times for untrained reversed versions should be longer than for trained standard versions due to the additional comparison. As shown in Figure 12b this prediction is not confirmed; rather, trained and untrained versions exhibit no significant differences in overall reaction times. However, as may be seen in Figure 1, enantiomorphs of the stimuli differ in no more than the location of one or two parts, suggesting that a comparison between them may occur quite rapidly (reversed versions may be seen by flipping Figure 1 over and holding it to the light).

representations stored with practice and because this is evidence that representations stored during training and during practice function equivalently in recognition. Despite these arguments, it is important to test whether presumably unstored reversed versions are recognized through viewpoint-independent mechanisms.

Tarr and Pinker (1989a) tested this possibility by not only manipulating training, but the amount of practice subjects received recognizing particular versions of each shape. Their crucial manipulation was the withholding of reversed versions of shapes throughout both *training and practice*, not presenting them until the surprise phase of the experiment. In the surprise block, where subjects were exposed to reversed versions for the first time, Tarr and Pinker (1989) once again found equivalent performance for the reversed versions at most orientations, suggesting that a flip in depth of a constant 180° was used to align reversed versions with representations of standard versions that came to be stored with extensive practice. This result replicated their original finding that in early trials subjects performed equivalently on all orientations of untrained versions. As in Experiment 3, a similar manipulation in Experiment 4 with three-dimensional objects is predicted to produce increasing rather than flat reaction time functions. As before, this is accounted for by the fact that, unlike two-dimensional mirror-image pairs, enantiomorphs of three-dimensional objects cannot be aligned by any rigid three-dimensional transformation. Instead, it is predicted that reversed versions will be rotated into partial congruence with stored standard versions, where a part-by-part comparison of mismatching parts may be made to verify that they are connected on exactly opposite sides. However, since stored models of standard versions presumably exist at all familiar orientations, it is predicted that rotations will be to the nearest familiar orientation rather than exclusively to the training orientation.

3.6.1. Method

Subjects. Twelve students from the Boston area participated in the experiment for pay.

Materials. The stimulus items, computer display, stimulus sets, and experimental conditions were identical to those used in Experiment 2.

Procedure. The training procedure was the same as in Experiment 2. Subjects were never shown the reversed versions of the objects.

Design. Practice blocks were identical to those used in Experiment 2, where no reversed versions were presented. Surprise blocks were identical to those used in Experiment 3, where the trials for both named objects and distractors were divided equally between standard and reversed versions. Subjects were given a self-timed break every 53 trials. Subjects were run in a total of four sessions identical to those in Experiment 1.

3.6.2. Results

Incorrect responses and responses for the preliminary trials in each block were discarded and reaction time means for each orientation were calculated by block, averaging over all objects. Equidistant rotations from a practice orientation around a common axis are expected to yield equivalent reaction times; therefore the effect of orientation will be characterized by regressing the reaction time means against the distance from a target orientation and calculating the slope of the best fitting line.

In Block 1 regressing mean reaction times against the distance from the training orientation reveals a slope of 12.6 msec/deg (79 deg/sec) for rotations around the x axis, 4.3 msec/deg (233 deg/sec) for the y axis, and 7.1 msec/deg (141 deg/sec) for the z axis. Over the next 11 practice blocks the slope continued to decrease for rotations around all three axes, with the slopes for Block 12 being: x axis, 1.7 msec/deg (588 deg/sec); y axis, 1.1 msec/deg (909 deg/sec); and z axis, 1.9 msec/deg (526 deg/sec; see Figure 15). From Block 1 to Block 12 there was also a decrease in overall reaction times, reflected as a decrease in the intercept of the reaction time functions. In Block 13, the surprise block, the slopes for standard versions in the practice orientations were: x axis, 3.3

msec/deg (303 deg/sec); y axis, 0.6 msec/deg (1667 deg/sec); and z axis, 1.9 msec/deg (526 deg/sec); while the slopes for reversed versions in the practice orientations were: x axis, 3.9 msec/deg (256 deg/sec); y axis, 1.8 msec/deg (556 deg/sec); and z axis, 2.6 msec/deg (385 deg/sec). Although some of these rates of rotation are within the range found in other mental rotation studies, they are much faster than the rates found in Block 1 of this experiment. In contrast, the slopes for standard versions in the surprise orientations, computed by averaging across means for orientations at equal distances from a practice orientation, were: x axis, 6.3 msec/deg (159 deg/sec; r = .98); y axis, 3.3 msec/deg (303 deg/sec; r = .90); and z axis, 7.4 msec/deg (135 deg/sec; r = 1.0); while the slopes for reversed versions were: x axis, 5.6 msec/deg (179 deg/sec; r = .96); y axis, 4.1 msec/deg (244 deg/sec; r = .98); and z axis, 2.4 msec/deg (417 deg/sec; r = .72). Once again, these estimates of the rate of rotation for surprise orientations may be somewhat underestimated. Figure 14 illustrates that the surprise orientations between 40° and 100° do not appear to be rotated to the nearest practiced orientation. A post hoc estimate of the rate of rotation was obtained by including only surprise orientations from 160° to 340° regressed against the distance to the nearest practiced orientation; for standard versions this analysis yielded slopes of: x axis, 9.3 msec/deg (108 deg/sec); y axis, 4.3 msec/deg (233 deg/sec); z axis, 8.1 msec/deg (123 deg/sec); while slopes for reversed versions were: x axis, 9.7 msec/deg (103 deg/sec); y axis, 4.1 msec/deg (244 deg/sec); z axis, 3.5 msec/deg (286 deg/sec). It is interesting to note that comparing slopes between standard and reversed versions, the crucial manipulation in this experiment, there is little difference for either the x and y axes; however there is a large difference for the z axis where standard versions exhibited a rate of rotation over twice as slow as that for reversed versions.

Insert Figure 14 about here.

The effects of practice were examined by three two-way ANOVA's separated by axis of rotation, on data from all practice blocks (1-12) with Block Number and Orientation as factors. Significant main effects for Block Number for the x axis (*F* (11, 121) = 29.1, $p < .001$), y axis (*F* (11, 121) = 22.8, $p < .001$), and z axis (*F* (11, 121) = 23.5, $p < .001$) were reflected as an overall decrease in reaction times with practice. Significant main effects for Orientation across blocks for the x axis (*F* (1, 11) = 33.4, $p < .001$), y axis (*F* (1, 11) = 25.0, $p < .01$), and z axis (*F* (1, 11) = 27.5, $p < .001$) were found, as were significant interactions between Block and Orientation for the x axis (*F* (11, 121) = 6.7, $p < .001$), y axis (*F* (11, 121) = 4.9, $p < .001$), and z axis (*F* (11, 121) = 4.9, $p < .001$). These interactions indicated that the effect of orientation changed with practice, and as shown in the data, diminished with practice.

The patterns of reaction times in the surprise block separated by the axes of rotation are shown in Figure 14. As in other experiments, recognition times generally increased with the distance from the nearest practice orientation. This may be seen in the reaction time curves for all three axes of rotation over the range of surprise orientations from 160° to 340°. With the exception of y axis rotations for reversed versions where the minima are 10° and 190°, minima appear near the practice orientations of 10° and 130°. For standard versions peaks appear at the surprise orientation of 250° for all three axes of rotation; for reversed versions peaks appear at 250° for the x axis, 280° for the y axis, and 220° for the z axis. Variations in the monotonicity of the functions, attributed in Experiment 1 to changes in the visible parts of objects with rotations in depth, fall at roughly the same orientations in Experiments 1 and 4. This was confirmed by three multiple regressions, one for each axis of rotation, on mean reaction times from Block 13 of Experiment 4 (collapsed across standard and reversed versions) with the mean reaction times from Block 13 of Experiment 1 and distance from the nearest practice orientation as predictors. These analyses revealed that the variation in reaction times in Experiment 1, beyond that correlated with distance from the nearest practice orientation, accounted for a significant amount of the variance in reaction times in Experiment 4 for the y axis ($F(1, 9) = 11.9$, $p < .01$) and the z axis ($F(1, 9) = 7.3$ $p < .05$), while not being significant for the x axis. Variation in the distance from the nearest practice orientation uncorrelated with reaction times from Experiment 1 accounted for a significant amount of the variance in Experiment 4 for the z axis ($F(1, 9) = 13.2$, $p < .01$), while not being significant for the x or y axes. These findings suggest that the reaction time functions display similar variations across orientation. In addition, six multiple regression analyses on mean reaction times from Block 13 with distance from the nearest practice orientation and distance from the training orientation as predictors confirmed that for each version around all three axes of rotation the distance from the nearest practice orientation accounted for a significant amount of the variance in reaction times: standard versions, x axis, $F(1, 9) = 12.4$, $p < .01$; y axis, $F(1, 9) = 16.1$, $p < .01$; z axis, $F(1, 9) = 33.8$, $p < .001$; reversed versions: x axis, $F(1, 9) = 7.8$, $p < .05$; y axis, $F(1, 9) = 5.5$, $p < .01$; z axis, $F(1, 9) = 7.4$, $p < .05$). Reversed versions rotated around the z axis the distance from the training orientation also accounted for a significant amount of the variance ($F(1, 9) = 5.4$, $p < .05$), while the distance from the training orientation was not significant for either version for the x or y axes. This finding is consistent with the results of Experiment 1, where z axis rotations exhibited a two components: rotation to the training orientation and rotation to the nearest practice orientation.

As shown in Table 6 error rates ranged from about 11-45% in Block 1 to about 0-2% in Block 12. In Block 13 error rates for surprise orientations ranged from about 3-24% for standard versions and 2-33% for reversed versions. No evidence for a speed/accuracy tradeoff in recognition was found in any block.

Insert Figure 15 about here.

Insert Table 6 about here.

3.6.3. Discussion

The same basic pattern of results found in Experiments 1, 2, and 3 is found in Experiment 4. To reiterate, in early practice trials there are effects of orientation from the distance to the training orientation; the effects of orientation diminish with practice at all familiar orientations; and the effects of orientation return for the same objects in unfamiliar orientations, now from the distance to the nearest familiar orientation. Moreover, as indicated by the similarity of reaction time functions for both standard and reversed versions in Block 13 with the reaction time function for Block 13 of Experiment 1, similar patterns of rotation are found for each axis. Finally, rotations in three dimensions sometimes appear to be through the shortest path to align the input shape with stored models. Again, all of these findings are consistent with a multiple-views-plus-transformation mechanism in recognition.

Secondly, the introduction of reversed versions of the objects in Block 13 produced few changes in reaction time patterns -- as shown in Figure 14b, for the x and y axes of rotation it appears that reversed versions were rotated to the nearest familiar orientation. There is one inconsistency in these results -- in Block 13 the function for reversed versions rotated around the z axis exhibited a faster rate of rotation than that found for standard versions. One explanation for this finding is that subjects sometimes ignored the fronts and backs of the objects, treating them as flat shapes, and used a 180° flip in depth, similar to that found by Tarr and Pinker (1989a), to align reversed versions with standard versions. An example of this is aligning your left and right hands by holding them in front of you and then rotating the left hand 180° around the vertical axis -- the fronts and backs are different, but the two-dimensional contours match. In Experiment 4, subjects could use this strategy because for picture plane rotations the two-dimensional contours of objects do not change; however, for rotations in depth the two-dimensional contours change at each orientation. Thus, the fast rate of z axis rotations for reversed versions may be accounted for by subjects sometimes rotating objects through a constant 180° flip in depth.

Generally, these findings are consistent with the One-Version Condition of

Experiment 3, where *untrained* reversed versions were rotated to the training orientation. Here *unpracticed* reversed versions are rotated to the nearest familiar orientation. Overall, these findings suggest that a viewpoint-dependent mechanism is still used in the recognition of reversed versions of objects for which no representation has been stored, and that the same mechanism was responsible for subjects' equivalent recognition performance at all orientations in Tarr and Pinker (1989a). Moreover, these findings help eliminate the possibility that viewpoint-independent mechanisms were responsible for this equivalent performance.

3.7. Experiment 5

In their recognition studies, Tarr and Pinker (1989a, b) suggest that there may be more than one path to object recognition. Although they show that for two-dimensional recognition (and extended here to three dimensions) subjects use a viewpoint-dependent process that includes the multiple-views-plus-transformation mechanism, they limit this process to *complex* recognition in which objects may be differentiated only on the basis of two-dimensional (or greater) spatial relations. For instance, in complex recognition simply knowing that part 1 is on top of part 2 is insufficient; one must also know how far to the side part 1 is relative to part 2. For simpler recognition, for instance knowing that part 1 is on top of part 2 or simply knowing that part 1 and part 2 are present in the object, they suggest that viewpoint-independent processes, such as locating unique features (see Eley, 1982), suffice. In another series of experiments (Tarr and Pinker, 1989b), they present evidence that viewpoint-independent processes are used in recognition and are capable of capturing some spatial relations between parts (for instance "on top of"). Specifically, they suggest that when shapes may be described in a one-dimensional ordering of parts along a single vector, including the ordering of parts to either side, but not including the side such parts are located on or their distance from the vector (a "one-and-one-half-dimensional" coordinate system), recognition is independent of viewpoint. This hypothesis is supported by three recognition experiments: in the first, two-dimensional shapes that were visually symmetric about the vertical axis were recognized independently of orientation; in the second, two-dimensional shapes that had the same configuration of parts at different metric distances on either side of the vertical axis axis were recognized independently of orientation; and in the third, two-dimensional shapes that had different parts on either side of the vertical axis, but in which the parts on one side predicted the parts on the other side perfectly, also were recognized independently of orientation. The common property between each of these sets of shapes is that it is sufficient to keep track of the one-dimensional ordering of parts on either side of the shape from bottom to top or vice-versa. Apparently perceivers can assign a one-dimensional vector to an object's axis defining a top-to-bottom ordering of parts equally quickly regardless of the the object's orientation. In contrast, for asymmetrical shapes where recognition was dependent on orientation the special property that requires rotation is that the locations of parts must be specified along two dimensions simultaneously (a "two-dimensional" or greater coordinate system), for instance part 1 being on top of part 2 and to the left of part 3. The mere requirement that two sides be kept distinct (even if neither has to be labeled as left or right) is enough to require that subjects mentally rotate, producing a systematic effect of orientation on recognition.

One possible criticism of this hypothesis is that Tarr and Pinker (1989b) demonstrate viewpoint-independent recognition only for shapes that exhibit symmetry across the vertical axis (although not only visual symmetry). A supporter of the rotation-for-handedness hypothesis might argue, despite evidence to the contrary, that effects of orientation in recognition occur only for asymmetrical objects where handedness version is possibly construed by the subject as relevant to the task. In Experiment 5, this hypothesis is examined by the introduction of two new classes of three-dimensional objects, both exhibiting left/right symmetry, one of which exhibits front/back asymmetry and one of which exhibits front/back symmetry (although not the same symmetry as left/right). Unlike symmetrical two-dimensional shapes, these three-dimensional objects can not be described in a one-and-one-half-dimensional coordinate system since the locations of parts attached to the front and back are necessary for differentiating among them (although subjects might take it upon themselves to ignore these parts and compress the objects along the line of sight). Additionally, these objects may be misoriented in three dimensions, producing multiple unique visible part configurations for each object, thus rendering a single one-dimensional ordering of parts ineffective for recognition at all viewpoints. Therefore, it is predicted that, despite their symmetry, objects within both classes will be recognized through viewpoint-dependent mechanisms involved in complex object recognition yielding results much the same as in Experiments 1-4. Alternatively, the rotation-for-handedness hypothesis predicts that, precisely because of their symmetry, objects in both classes will be recognized through orientation-independent mechanisms yielding results much the same as those found in Tarr and Pinker (1989b).

3.7.1. Method

Subjects. Twenty-four students from the Boston area participated in the experiment for pay: 12 in each of two conditions.

Materials. The computer display and experimental conditions were identical to those used in Experiment 2.

Two conditions corresponding to two stimulus sets were employed. In the first, *Condition AS*, the stimuli consisted of seven left/right symmetrical and front/back asymmetrical objects illustrated in Figure 16 in their training positions. In the second, *Condition SS*, the stimuli consisted of seven left/right and front/back symmetrical objects illustrated in Figure 17 in their training positions. As in previous experiments, in each condition the objects were grouped into three sets of three named objects each.

Insert Figures 16 and 17 about here.

Procedure. In both conditions the training procedure was the same as in Experiment 2. Because of the left/right symmetry present in all stimulus objects, no mirror-reversed versions were used (or possible).

Design. In both of the conditions, practice and surprise blocks were identical to those used in Experiment 2, where no reversed versions were presented. Subjects were given a self-timed break every 53 trials. In each condition, subjects were run in a total of four sessions identical to those in Experiment 1.

3.7.2. Results

For each condition, incorrect responses and responses for the preliminary trials in each block were discarded and reaction time means for each orientation were calculated by block, averaging over all objects.

Condition AS. In Block 1 regressing mean reaction times against the distance from the training orientation reveals slopes of: x axis, 15.7 msec/deg (64 deg/sec); y axis, 6.5 msec/deg (154 deg/sec); and z axis, 8.8 msec/deg (114 deg/sec). The effect of orientation on the reaction time functions for Block 1 is shown in Figure 18. Over the next 11 practice blocks the slope continued to decrease for rotations around all three axes, with the slopes for Block 12 being: x axis, 1.3 msec/deg (769 deg/sec); y axis, 0.4 msec/deg (2500 deg/sec); and z axis, 0.8 msec/deg (1250 deg/sec; see Figure 20). From Block 1 to Block 12 there was also a decrease in overall reaction times, reflected as a decrease in the intercept of the reaction time functions. In Block 13, the surprise block, the slopes for the practice orientations were: x axis, 2.2 msec/deg (455 deg/sec); y axis, 0.6 msec/deg (1667 deg/sec); and z axis, 1.0 msec/deg (1000 deg/sec). Slopes for the surprise orientations, computed by averaging across means for orientations at equal distances from a practice orientation, were: x axis, 4.2 msec/deg (238 deg/sec; r = .87); y axis, 2.2 msec/deg (455 deg/sec; r = .99); and z axis, 3.5 msec/deg (286 deg/sec; r = .55). Once again, these estimates of the rate of rotation for surprise orientations may be somewhat underestimated. Figure 19 shows that the surprise orientations between 40° and 100° do not appear to be rotated to the nearest practiced orientation. A post hoc estimate of the rate of rotation was obtained by including only surprise orientations from 160° to 340° regressed against the distance to the nearest practiced orientation; this analysis yielded slopes of: x axis, 9.4 msec/deg (106 deg/sec); y axis, 2.0 msec/deg (500 deg/sec); z axis, 5.1 msec/deg (196 deg/sec).

Insert Figure 18 about here.

The effects of practice were examined by three two-way ANOVA's, one for each axis of rotation, on data from all practice blocks (1-12) with Block Number and Orientation as factors. Significant main effects for Block Number for the x axis (F (11, 121) = 79.3, $p < .001$), y axis (F (11, 121) = 116.7, $p < .001$), and z axis (F (11, 121) = 73.3, $p < .001$) were reflected as an overall decrease in reaction times with practice. Significant main effects for Orientation across blocks for the x axis (F (1, 11) = 88.6, $p < .001$), y axis (F (1, 11) = 44.8, $p < .001$), and z axis (F (1, 11) = 84.4, $p < .001$) were found, as were significant interactions between Block and Orientation for the x axis (F (11, 121) = 19.3, $p < .001$), y axis (F (11, 121) = 14.9, $p < .001$), and z axis (F (11, 121) = 19.3, $p < .001$). These interactions indicated that the effect of orientation changed with practice, and as shown in the data, diminished with practice.

The patterns of reaction times in the surprise block separated by the axes of rotation are shown in Figure 19. As in other experiments, recognition times generally increased with the distance from the nearest practice orientation. This may be seen in the reaction time curves for the x and z axes of rotation over the range of surprise orientations from 160° to 340°. In contrast, the reaction time curve for the y axis exhibits roughly equivalent recognition times no matter what the distance from a practice orientation. In most instances, minima appear near the practice orientations of 10° and 130°. Peaks appear at the surprise orientation of 220° for the x and z axes of rotation, while a small peak appears at 280° for the y axis. As expected, because new stimuli were used, variations in the monotonicity of the functions fall at different orientations than in Experiments 1-4. Multiple regression analyses on mean reaction times from Block 13 with distance from the nearest practice orientation and distance from the training orientation as predictors confirmed that for the x and, surprisingly, the y axes of rotation the distance from the nearest practice orientation accounted for a significant amount of the variance in reaction times (x axis, $F(1, 9) = 4.0$, $p < .08$; y axis, $F(1, 9) = 16.0$, $p < .01$), but was not significant for the z axis, while the distance from the training orientation was not significant for any axis of rotation.

As shown in Table 7 error rates ranged from about 5-42% in Block 1 to about 0-2% in Block 12. In Block 13 error rates for surprise orientations ranged from about 1-32%. No evidence for a speed/accuracy tradeoff in recognition was found in any block of any condition.

Insert Figures 19 and 20 about here.

Insert Table 7 about here.

Condition SS. In Block 1 regressing mean reaction times against the distance from the training orientation reveals slopes of: x axis, 7.3 msec/deg (137 deg/sec); y axis, 3.8 msec/deg (263 deg/sec); and z axis, 6.5 msec/deg (154 deg/sec). The effect of orientation on the reaction time functions for Block 1 is shown in Figure 18. Over the next 11 practice blocks the slope continued to decrease for rotations around all three axes, with the slopes for Block 12 being: x axis, 1.2 msec/deg (833 deg/sec); y axis, -0.2 msec/deg (-5000 deg/sec); and z axis, 1.0 msec/deg (1000 deg/sec; see Figure 22). From Block 1 to Block 12 there was also a decrease in overall reaction times, reflected as a decrease in the intercept of the reaction time functions. In Block 13, the surprise block, the slopes for the practice orientations were: x axis, 2.8 msec/deg (357 deg/sec); y axis, 0.2 msec/deg (5000 deg/sec); and z axis, 0.9 msec/deg (1111 deg/sec). Slopes for the surprise orientations, computed by averaging across means for orientations at equal distances from a practice orientation, were: x axis, 10.0 msec/deg (100 deg/sec; $r = .91$); y axis, 7.2 msec/deg (139 deg/sec; $r = .96$); and z axis, 1.4 msec/deg (714 deg/sec; $r = .31$). Once again, these estimates of the rate of rotation for surprise orientations may be somewhat underestimated. Figure 21 shows that the surprise orientations between 40° and 100° do not appear to be rotated to the nearest practiced orientation. A post hoc estimate of the rate of rotation was obtained by including only surprise orientations from 160° to 340° regressed against the distance to the nearest practiced orientation; this analysis yielded slopes of: x axis, 15.6 msec/deg (64 deg/sec); y axis, 7.9 msec/deg (127 deg/sec); z axis, 2.9 msec/deg (345 deg/sec).

Insert Figure 21 about here.

The effects of practice were examined by three two-way ANOVA's, one for each axis of rotation, on data from all practice blocks (1-12) with Block Number and Orientation as factors. Significant main effects for Block Number for the x axis (F (11, 121) = 79.0, $p < .001$), y axis (F (11, 121) = 53.8, $p < .001$), and z axis (F (11, 121) = 61.8, $p < .001$) were reflected as an overall decrease in reaction times with practice. Significant main effects for Orientation across blocks for the x axis (F (1, 11) = 80.4, $p < .001$), y axis (F (1, 11) = 8.5, $p < .02$), and z axis (F (1, 11) = 60.9, $p < .001$) were found, as were significant interactions between Block and Orientation for the x axis (F (11, 121) = 11.6, $p < .001$), y axis (F (11, 121) = 3.4, $p < .001$), and z axis (F (11, 121) = 6.5, $p < .001$). These interactions indicated that the effect of orientation changed with practice, and as shown in the data, diminished with practice.

The patterns of reaction times in the surprise block separated by the axes of rotation are shown in Figure 21. As in other experiments, recognition times generally increased with the distance from the nearest practice orientation. This may be seen in the reaction time curves for all axes of rotation over the range of surprise orientations from 160° to 340°. Minima appear at or near the practice orientations of 10° and 130°. Peaks appear at the surprise orientation of 220° for the x and z axes of rotation, while a peak appears at 250° for the y axis. As in Condition AS, the use of new stimuli is hypothesized to account for variations in the monotonicity of the functions falling at different orientations than in

Experiments 1-4. Similarly, multiple regression analyses on mean reaction times from Block 13 with distance from the nearest practice orientation and distance from the training orientation as predictors confirmed that for the x and y axes of rotation the distance from the nearest practice orientation accounted for a significant amount of the variance in reaction times (x axis, $F(1, 9) = 5.9$, $p < .05$; y axis, $F(1, 9) = 17.8$, $p < .01$), but was not significant for the z axis, while the distance from the training orientation was not significant for any axis of rotation.

As shown in Table 8 error rates ranged from about 5-17% in Block 1 to about 1-4% in Block 12. In Block 13 error rates for surprise orientations ranged from about 1-42%. No evidence for a speed/accuracy tradeoff in recognition was found in any block of any condition.

Insert Figure 22 about here.

Insert Table 8 about here.

3.7.3. Discussion

The major new finding of Experiment 5 is that recognition of two different classes of left/right symmetrical three-dimensional objects exhibits viewpoint dependency. This is clearly seen in Figure 18, showing the recognition times in Block 1 for both conditions. This is in contrast to Tarr and Pinker's (1989b) finding that the recognition of two-dimensional "symmetrical" shapes did not exhibit viewpoint dependency. Moreover, as in Experiments 1-4, an orientation-dependent recognition process was used both when the objects were first encountered, after training, in unfamiliar orientations and again, after extensive practice, when the objects were encountered in additional unfamiliar orientations.

This discrepancy between the recognition of symmetrical two-dimensional shapes and three-dimensional objects is consistent with Tarr and Pinker's (1989b) hypothesis that it is only when humans must describe the spatial relations within objects along two or more dimensions that viewpoint-dependent recognition mechanisms are used. While members within each class of two-dimensional shapes could be differentiated by a single vector describing the ordering of parts, members within each class of three-dimensional objects could be differentiated only by the addition of at least one more vector describing the relationship of the parts connected to the front and back of each object to the vector describing the ordering of symmetric left/right parts. This is true whether or not the front/back parts are symmetric as well -- in either case these parts must still be described relative to the ordering of left/right parts.

Overall, the strongest result from Experiment 5 is that the recognition of three-dimensional left/right symmetrical objects is viewpoint dependent. This finding completely rules out the rotation-for-handedness hypothesis, since subjects in this experiment *could not* assign handedness to the stimuli, yet they still rotated them for recognition. Further, this viewpoint dependency exhibits a pattern common to Experiments 1-4, supporting the existence of a multiple-views-plus-transformation mechanism in complex object recognition. In addition, taken together with Tarr and Pinker's (1989b) results, Experiment 5 provides evidence for two distinct processes in the recognition of objects within human vision -- a viewpoint-independent process based on mechanisms utilizing a "one-and-one-half-dimensional" coordinate system in which simple features or an ordering of features are described and a viewpoint-dependent process based on the multiple-views-plus-transformation mechanism and a two-dimensional or greater coordinate system in which complex spatial relations between parts are described.

4. General Discussion

4.1. Major results

The study presented here tested whether complex three-dimensional object recognition is viewpoint-dependent by having subjects learn novel objects in a single orientation and then having them recognize the same objects in several new orientations generated by three-dimensional rotations. In addition, the study was designed to help disentangle viewpoint-dependent theories from viewpoint-independent theories by having subjects familiarize themselves with the stimulus objects in this small set of orientations and then having them recognize the same objects in unfamiliar orientations. Generally, the key results of the four recognition experiments (Experiments 2-5) may be summarized as follows:

- In early trials, when subjects first recognized objects misoriented in depth around one of the three major axes, their recognition times increased monotonically with distance from the trained near upright orientation. Moreover, this effect was found even when handedness was explicitly irrelevant to the recognition task either because enantiomorphs of an object were referred to by the same name or because the objects were symmetrical.
- With extensive practice at recognizing the objects at several orientations, the effect of orientation diminished to near equivalent performance at each of the familiar orientations.
- When subjects were presented with the same objects in new, unfamiliar orientations these diminished effects of orientation for familiar orientations did not transfer to unfamiliar orientations. Rather, recognition times were once again orientation dependent, this time increasing with distance from the nearest familiar orientation or the training orientation.

4.2. Major conclusions

These results suggest that subjects first recognize misoriented objects by rotating them to the training orientation; with practice subjects store viewpoint-specific representations of the objects at each familiar orientation, allowing subjects to recognize misoriented objects without rotation; and that subjects recognize misoriented objects in unfamiliar orientations by rotating them to the nearest familiar orientation. These results are consistent with the results of experiments by Tarr and Pinker (1989a), supporting the multiple-views-plus-transformation theory. In addition, results unique to these three-dimensional experiments extend the theory with the following implications:

- Orientation dependence in both early trials after training and in surprise trials after extensive practice (where familiar objects were presented in unfamiliar orientations) may be accounted for by the use of mental rotation. This is supported by the similarity between the measured rates of rotation in these conditions and the rates found in handedness discriminations that uncontroversially involve mental rotation. Moreover, these rates are of roughly the same order of magnitude as those found in other three-dimensional mental rotation studies (Metzler and Shepard, 1974; Parsons, 1987c; Shepard and Metzler, 1988; see Table 1). In addition, the variations within reaction time functions and the relative ordering of slopes across axes of rotation for recognition were somewhat consistent with the variations and ordering of slopes found for handedness discriminations on the same objects. For handedness discriminations, the relative ordering of rates of rotation around each axis was (slowest to fastest): x, y, z (with one exception -- the One-Version Condition of Experiment 3); for all experiments involving recognition judgments the relative ordering in Block 1 was: x, z, y. In addition, with the same exception, for all experiments in both Blocks 1 and 13 rotations around the x axis were always the slowest (perhaps because only x axis rotations involve changes in the location of the top and the visible features of objects; y axis rotations only involve changes in features, while z axis rotations only involve changes in the location of the top). These findings suggest that the same viewpoint-dependent process, mental rotation, was used in both recognition and handedness judgments.

- Reaction time functions were generally consistent with mental rotations around the major axis of rotation originally used to generate the misorientation, suggesting that subjects usually rotated objects through the shortest path in three dimensions to match input shapes with stored models.

- When subjects did not have an opportunity to learn mirror-reversed versions of three-dimensional objects, which cannot be aligned with their enantiomorphs by any rigid three-dimensional rotation, their recognition times still exhibited the same pattern of orientation dependency as standard versions (with the exception of y axis rotations for surprise orientations in Experiment 4). This suggests that untrained mirror-reversed versions *of three-dimensional objects* are recognized by invoking mental rotation. Furthermore, this suggests that untrained mirror-reversed versions *of two-dimensional shapes* are recognized by a similar viewpoint-dependent process, and that the constant reaction times across orientations found by Tarr and Pinker (1989a) for such shapes are due to a consistent 180° shortest path rotation in depth, rather than viewpoint-independent mechanisms.

- The recognition of bilaterally symmetrical three-dimensional objects (symmetrical across the sagittal plane) yields orientation-dependent reaction time functions similar to those produced by the recognition of asymmetrical three-dimensional objects. This suggests that the rotation-for-handedness hypothesis cannot be correct, since for these objects it was *impossible for subjects to assign handedness*, yet effects of orientation on recognition were still found. Furthermore, this indicates that symmetry alone provides insufficient redundancy for viewpoint-independent recognition. Rather, unless objects may be discriminated on the basis of a simple ordering of parts, recognition must rely on viewpoint-dependent mechanisms such as multiple-views-plus-transformation.

Overall, these findings are inconsistent with a solely viewpoint-independent theory of object recognition. In particular, theories that argue for the viewpoint-invariant recovery of volumetric primitives and subsequent reconstruction of three-dimensional structure of objects cannot account for the large effects of orientation in recognition. Furthermore, even quasi-viewpoint-independent theories, in which primitives are recoverable over a limited range of orientations encompassing a single view of an object, cannot account for the increase in the reaction times with the distance from the training or nearest familiar orientation. Both of these theories predict that the diminished effects of orientation that arise with extensive practice should transfer from familiar to unfamiliar orientations (in the first case, to all unfamiliar orientations, and in the second case, to all unfamiliar orientations within the same view). The results here do not confirm this prediction -- diminished effects of orientation *never* transfer to unfamiliar orientations and, as stated, reaction times for unfamiliar orientations increase with distance from the nearest familiar orientation.

These results are also inconsistent with two proposals that viewpoint dependence in three-dimensional object recognition is a special case -- first, the rotation-for-handedness hypothesis, which suggests that mental rotation is used only when handedness is possibly relevant to the judgment, and second, the hypothesis that mental rotation is used only when the top of an object with respect to gravitational upright must be located. In none of the experiments did determining handedness facilitate recognition, yet effects of orientation were found consistently. Furthermore, even making handedness explicitly irrelevant or impossible to assign failed to eradicate effects of orientation on recognition. Additionally, recognition of objects in orientations that preserved the position of the top of an object with respect to gravity (y axis rotations) also failed to eradicate effects of orientation.

In contrast, the results of all of the experiments presented here are consistent with a viewpoint-dependent theory of complex object recognition, specifically, the multiple-views-plus-transformation theory. In particular, this theory accounts for initial effects of orientation, diminished effects of orientation with practice, and the lack of transfer between familiar and unfamiliar orientations. Additionally, the multiple-views-plus-transformation mechanism predicts that alignments between input shapes and stored models should be relatively efficient, rotating input shapes through approximately the shortest path in three dimensions to the nearest orientation at which a candidate stored models exists -- a finding supported in some of the results of this study. Cumulatively these findings demonstrate that the stored representations involved in complex object recognition are viewpoint-specific and that mental rotation through the shortest three-dimensional path is used to compensate for this specificity.

5. Computational Issues in the Multiple-Views-Plus-Transformation Theory

All theories of object recognition are aimed at addressing the fundamental problem of how objects are recognized despite differences in their two-dimensional retinal projections that arise from varying and frequently novel viewpoints. The multiple-views-plus-transformation theory provides a solution to this problem by proposing that objects are represented in visual long-term memory as a collection of viewpoint-specific representations (referred to as "viewer-centered" if the coordinate system is determined by the perspective of the viewer -- see Marr and Nishihara, 1978). As in multiple-views theories an object will be recognized directly if it is observed at a viewpoint that matches a stored viewpoint-specific representation. However, since it is both theoretically and practically impossible to store views of an object at every possible orientation, there can never exist a stored representation for every view of an object. Therefore to recognize novel and otherwise non-stored views of an object, the multiple-views-plus-transformation theory includes the possibility of a mental transformation to align the current view of an object with a stored view. The addition of a mechanism for mental transformation greatly reduces the need for a large number of stored representations, thus making the theory more parsimonious. However, the inclusion of mental rotation also allows the extreme case of reducing the theory to the storage of a single view sufficient for the recognition of an object across any rotation which preserves the visible surfaces of an object (the equivalent of the single-view-plus-transformation theory). Although there is some empirical evidence suggesting that this may sometimes occur in human shape recognition (Tarr and Pinker, 1989a, Experiment 1), the desire to minimize the magnitude of rotations dictates that more than one view of an object be stored for efficient recognition. For instance, considering rotations in the frontal plane only, storing only two equally spaced views of an object reduces the maximum possible rotation required for a match to 90°, and storing four equally spaced views of an object reduces the maximum possible rotation to 45°.

5.1. How are the direction, magnitude, and target of rotations to the nearest stored view determined prior to recognition?

One of the most persuasive arguments against the use of mental rotation for recognition is the paradox of needing to identify an object in order to know the correct direction and distance of the rotation needed to match it with a target representation (Corballis, et al., 1978; Shepard and Cooper, 1982). One solution to this paradox is that only a small portion of the information available in the object's input shape is used to determine the rotation. Huttenlocher and Ullman (1987a, b, 1988; Ullman, 1986) present a computational theory of object recognition that, similar to the multiple-views-plus-transformation theory, relies on the alignment of objects in observed viewpoints to viewpoint-specific representations. In this theory, Ullman (1986) suggests that recognition is dependent on "alignment keys" -- cues to an object's orientation in the current viewpoint that are independent of the identity of the object. Ullman demonstrates that if three non-collinear landmarks features are located on both the observed object and the stored representation, the two-dimensional coordinates of these landmarks are sufficient to compute the direction and the magnitude of the rotation (as well as the translation and size scaling) necessary to bring the two into alignment.

Huttenlocher and Ullman suggest that the correct target representation for a given transformation may be determined by comparing the alignment keys of the observed object with all stored representations, performing the necessary alignment, and then comparing all possible matches in parallel. However, considering that humans must store an extremely large number of representations, this is not an entirely satisfactory solution since no attempt is made to reduce the search space of stored representations prior to comparison -- the sheer number of possible matches is handled simply by appealing to parallel mechanisms. A more efficient alternative is to use an overconstrained alignment key (Ullman, personal communication). While three landmarks on an observed object can always be aligned exactly with three landmarks in stored representations, four or more landmarks may only be aligned approximately (unless of course there is perfect match), for instance by a least-squares algorithm. The resulting goodness-of-fit measure provides an indication of whether the actual alignment is worth carrying out, thereby providing a technique for reducing the search space of stored representations.

5.1.1. Object recognition by alignment (ORA)

Huttenlocher and Ullman (1987b, 1988) have implemented this "Object Recognition by Alignment" (ORA) theory in a working computer recognition system that successfully recognizes three-dimensional objects in arbitrary orientations, including rotations in depth, from stored two-dimensional views. This implementation uses viewpoint-specific models as the basis for matches with input shapes. Only surfaces of the object visible from the viewpoint of the model are stored. Therefore a given object will be represented by several models, each representing an object from a different viewpoint. Although not discussed explicitly, Huttenlocher and Ullman imply that for a given object each unique set of surfaces must be represented in at least one view, thereby representing all possible alignment keys to be found on a given object. If this condition is satisfied, the overall representation of the object, consisting of several viewpoint-specific models, will be sufficient for the recognition of that object from any arbitrary two-dimensional view.

5.1.2. The low-level representation of shape in ORA

Huttenlocher and Ullman's implementation of ORA makes explicit the format for the low-level primitives that are extracted from the input shape and that comprise the stored viewpoint-specific representations. Each model consists of the three-dimensional locations of the edges of an object's surface derived from the intensity edges in grey-level images. Within this edge-based representation two distinct classes of image features may be identified: Class I features that define three points and Class II features that define an oriented point. Rather than using three or more landmarks to compute an alignment, the ORA system uses single Class I features and pairs of Class II features, each of which define more than a single point. Matches between the input shape and stored models are determined by solving for all possible alignments between Class I features, performing these alignments, and then verifying the match for all features of the input shape by comparing these features with nearby model features. Any features of the input shape that are not successfully matched to Class I features in a stored model are then used for potential alignments with Class II features. Verification of a successful match between a stored model and an input shape is based on sufficient correspondence between the stored model's Class I and Class II features and the input shape's features. This last step implements is similar to what Lowe (1987) calls the *viewpoint consistency constraint*, so called because low-level perceptually organized properties projected by an object should be consistent across instances of an object from a particular viewpoint (for instance, parallel lines should be parallel in both a stored model and an input shape from the same viewpoint). After a tentative match has been made and the input shape has been aligned with a candidate stored model, the recognition system verifies that the majority of groupings of features in the image and in the stored model coincide.

The edge-based representations proposed by Huttenlocher and Ullman (1988) contain sufficient information about the input shape to be used in a functional computer system for successful recognition by alignment. However, as they point out, no surface properties of the object are represented -- only bounding contours are stored in models of objects. Thus, for objects that are characterized by more complex properties, such as surface characteristics, edge-based representations are not sufficient for successful recognition by alignment. When one considers human performance, for instance with human faces where shading is important for differentiating individuals, it seems that the stored representations should contain more than just edge information. However, arguing against this claim, there is some evidence that edge-based representations may be sufficient for recognition. Biederman and Ju (1988) present several experiments that demonstrate that surface characteristics such as a diagnostic color do not facilitate faster recognition, leading them to suggest that objects are recognized primarily through the use of edged-based representations, with surface characteristics playing only a secondary role. Furthermore, our accurate recognition of caricatures and cartoons suggests that edges provide sufficient information for successful recognition. At a minimum, it is likely that stored representations of objects contain edge-based information, although not exclusive of other type of information.

Considering how such edge-based representations might be organized in human vision, a first hypothesis might be provided by adopting those features used by Huttenlocher and Ullman. Certainly the Class I and Class II features incorporated into the ORA theory are adequate for the implementation of a functional recognition system; however they seem somewhat arbitrary as representational primitives in human vision. There is no conclusive experimental evidence to suggest that these features are fundamental in the perceptual organization of images. This is not a serious problem for ORA, since Huttenlocher and Ullman were concerned with implementing a working computer recognition system rather than a psychologically plausible theory of human vision. However, when considered in the context of human performance it is not obvious that the features used in ORA are sufficient.

5.2. What conditions determine which views are stored?

Another computational issue concerns what criteria should be used by a multiple-views-plus-transformation mechanism to determine when a new viewpoint-specific representation of an object should be stored. This is referred to as the *generality of views* problem. This problem arises from there being a possibility of observing an object from any three-dimensional viewpoint -- for even a single object there exists a far greater number of possible three-dimensional viewpoints than can conceivably be stored efficiently by the human recognition system. The question is how to differentiate between a new view that is different enough from stored views for it to be worth storing and a new view that is near enough to an already stored view for it to remain unstored. The same problem exists for views of objects limited to rotations in the frontal plane where the same parts are always visible; however, because of the single degree-of-freedom in such rotations, it is not as significant a problem as for rotations in three dimensions.

5.2.1. Probabilistic factors

Tarr and Pinker (1989a) address this problem by proposing that views of an object are stored at any frequently seen and therefore familiar orientation of an object. Whether a view is actually stored or not is probabilistic: the more often an object is seen from a particular viewpoint, the greater the likelihood that it will be represented at that viewpoint and the higher the probability that it will be observed from that viewpoint in the future. Thus, the most common views of an object will be the views recognized most easily -- by a direct match or by a minimal transformation to a stored view.

Tarr and Pinker point out that storing objects in familiar orientations makes ecological sense -- one efficient strategy for recognition is to concentrate on doing a good job recognizing objects in their most commonly observed viewpoints, for instance facing the dial on a phone. This argument is supported by several pieces of evidence. First, the results of Tarr and Pinker and this thesis demonstrate that subjects recognize familiar characters in unfamiliar views by rotating them to the nearest familiar orientation, suggesting that humans store representations of objects at familiar orientations. Second, Kendrick and Baldwin (1987) found that some neurons in monkeys are maximally responsive to upright monkey faces and that other neurons are maximally responsive to upside-down monkey faces, but that neurons in sheep are maximally responsive only to upright sheep faces. They argued that the difference is related to the fact that monkeys, which are arboreal, often view other monkeys upside down, but that sheep almost never view other sheep upside down. Third, Perrett, Mistlin, and Chitty's (1987) found that there exist separate cells in monkeys maximally sensitive to full-face views of monkey faces and other cells maximally sensitive to profiles. Finally, Rock (1973, 1983) has pointed out that humans have difficulty recognizing objects at other orientations, perhaps because humans most frequently observe objects from an upright position or because many objects themselves have a common orientation with respect to gravity. Thus, there is evidence from both human performance and from ecological observations of monkeys, sheep, and humans that views of an object are stored based on their actual frequency of observation in the environment.

5.2.2. Configurational factors

A completely probabilistic mechanism for determining which views of an object are stored may not be entirely satisfactory when extended to three dimensions. The sheer number of commonly seen views might easily overwhelm the representational system. To extend the multiple-views mechanism to three dimensions, a new mechanism is needed for determining when a view of an object is unique enough to merit being stored. One constraint that such a mechanism might use to its advantage is the fact that most three-dimensional objects are self-occluding. Therefore, unlike views of two-dimensional shapes, no view of a three-dimensional object will represent all of the surfaces of that object -- or put another way, stored views of three-dimensional objects may vary in their information content as well as their viewpoint. As a consequence, this new mechanism may operate by determining when the configuration of visible surfaces in a viewpoint is unique and therefore should be stored. Thus, a view of an object is only considered new and worth storing if it depicts a new configuration of the surfaces of an object.

5.2.3. A topologically-based representation of objects

What properties of surfaces might such a mechanism rely on to classify the visible surfaces of objects into common or distinctive views? One possibility is that consistent geometric properties of surfaces are extracted. Koenderink (1987) has proposed such a model, suggesting three-dimensional objects are represented as a set of two-dimensional *generic views* classified by topological properties of surfaces. For instance, Koenderink suggests that a house might be represented by eight generic views -- four for the views from each corner and four for each side alone (ignoring the top and bottom of the house). Generic views are defined as those views of an object that are qualitatively distinct in terms of the topological properties of the image. Thus, each new generic view of an object will be identified by a new configuration of surface features derived from the image. Views exhibiting previously seen configurations will be considered simply as an instance of a stored generic view. Koenderink also precisely specifies the exact set of surface features that comprise his "language" for describing surface shape in terms of its topological structure. This includes local features based on points, curves, and regions, and more global features based on planes, surfaces, and lines. Further, he suggests that when you partition surfaces based on these features you obtain a decomposition of shape with well specified generic views. This is an important point -- otherwise there would be no way to determine which combinations of features constitute a generic view of an object.

Koenderink's (1987) theory provides a formal basis for determining the minimal number of viewpoint-specific representations sufficient for recognizing an object from any viewpoint. In particular, it seems probable that recognition by alignment generally occurs *within* a generic view rather than *across* generic views. This is because within a generic view the visible parts of an object remain constant, but across generic views they change. For instance, it may be that an alignment process similar to that proposed by Ullman (1986) only computes alignments between landmarks in the input shape and landmarks in stored generic views of similar topology (to each other and to the input shape).[15]

One implication of Koenderink's theory is that no more than one representation need be stored for each generic view of an object. However, his theory is intended as the minimal case for the representation of structure rather than a comprehensive theory of human performance -- it would be surprising if people *only stored* those views uniquely defined by topological features. Clearly humans perceive many topologically equivalent views as different -- for instance views produced by rotation around the line of sight axis are topologically equivalent, but differ in their orientation with respect to gravity and are often perceived differently.

5.2.4. The representation of three-dimensional structure in generic views

Koenderink suggests that a unique decomposition of generic views of an object are organized into a graph structure that describes the spatial relations between views and in doing so, the three-dimensional structure of the object (for instance imagine a collection of flat paper cutouts with labels such as "insert Tab A into Slot B" that link them together to form a model house). Thus, when the boundary of one generic view is encountered, a new generic view may be made visible by following the graph representing the particular object. Koenderink and van Doorn (1979) even speculate that one important variable in Metzler and Shepard's (1974) experiments other than the angle of rotation may be the number of generic views an object must pass through (although when this hypothesis was tested by Metzler and Shepard, 1974, the results were negative). However, this speculation does not address how it was possible for subjects here

---

[15] Koenderink's proposal that generic views are determined by their unique geometric features is similar to Ullman's suggestion that overconstrained alignment keys are used to determine appropriate views for alignment. The difference is that Koenderink proposes a theory for how to carve up stored representations of objects, while Ullman proposes a theory for how to differentiate between already stored representations. To some extent, Ullman addresses this issue as well -- in Huttenlocher and Ullman's (1987b) computer implementation of their alignment theory they suggest that individual views are stored for each position from which a different set of object surfaces is visible. However, they fail to provide a precise definition of what constitutes "different".

and in many mental rotation experiments (Metzler and Shepard, 1974; Parsons, 1987c) to rotate objects into unfamiliar views (since the stimuli used in all of these experiments were novel and subjects were never given the opportunity to view the objects from intermediate viewpoints). One possibility is that all of these experiments used stimuli composed of strings of cubes, connected at right angles. The three-dimensional structure of these objects was therefore highly predictable to the extent that subjects were even able to predict new unfamiliar generic views.[16] However, Koenderink's theory appears to predict that for less predictable smoothly varying objects rotations and/or alignments will be restricted to within generic views.

5.3. Why the multiple-views-plus-transformation mechanism is not a template theory

Any theory based on viewpoint-dependent representations may end up proposing that humans store a new representation for every slight variation in viewpoint, of which there are an infinite number. This is a new example of the classic template matching problem (for instance, see Lindsay and Norman, 1977). One way around this problem is to suggest that each stored representation has generality for objects that vary in position or in form. However it is important that the source of this generality be rigorously specified, otherwise one can always simply push back the mechanism for generality to another level, where eventually once again the theory will run into trouble explaining the template matching problem. Viewpoint-independent theories (Biederman, 1987; Marr and Nishihara, 1978; Pentland, 1986) avoid this problem by abstracting away from the input shape. This abstraction, realized as an object-centered structural description, produces a viewpoint-invariant representation composed of parameterized volumetric primitives that approximate not only the input object, but many members that share the same

---

[16] The ability of subjects to predict the nature of unfamiliar generic views may provide clues as to how people recognize known objects from novel never-before-seen viewpoints and, in general, how people learn about the three-dimensional structure of objects.

hierarchy of primitives. Thus, no further mechanism for generality in either position or form need be specified -- many members of a class of objects at arbitrary orientations lead to like representations. In contrast, any theory that relies on a multiple-views-plus-transformation mechanism must be careful to specify the limits of generality in viewpoint-specific representations. There are two components to this. First, generality across viewpoint is specified by using geometric surface properties of objects to classify them into generic views (Koenderink, 1987). Second, generality within generic views is specified by the matching of identity-independent landmarks to compute an alignment between the input shape and stored views (Huttenlocher and Ullman, 1987a, b).

5.4. The role of the recovery of three-dimensional structure in object recognition

The existence of a multiple-views-plus-transformation mechanism raises the issue of whether the recovery of the three-dimensional structure of objects is necessary for the successful recognition of objects. It is entirely plausible that the stored views of objects used in recognition contain no three-dimensional information or at the most viewpoint-specific depth information. In contrast, several of the most important theories of object recognition are based on the premise that recognition is achieved by the recovery of the three-dimensional structure of an object from two-dimensional retinal input and the subsequent comparison of the resulting three-dimensional model with similar stored representations (Biederman, 1987; Marr and Nishihara, 1978; Pentland, 1986). These theories may be referred to as *object model* based theories because they reconstruct a three-dimensional model of the observed object. Object models are structural descriptions composed of volumetric primitives (for instance, cylinders or cubes) and describe the structure of an object through the spatial relations between these primitives. In general, the recovery of individual primitives is accomplished by extracting particular features within the image that provide constraints on the possible structure of the object (for instance, a cylinder might be recovered by locating its axis of elongation; see Marr and Nishihara, 1978).

Such descriptions, both recovered from images and stored for use in matching, are viewpoint invariant in that three-dimensional object models composed of the same primitives will be recovered from all possible two-dimensional projections of an object. It is interesting that representations of this type coincide with our introspections about the three-dimensional nature of objects -- it is not hard to believe that we have stored faithful three-dimensional replicas of objects. Object-model based theories are also appealing in that most solid-modeling CAD systems (which implement the inverse problem of recognition -- what three-dimensional representations should be used to produce realistic two-dimensional projections of objects?) use three-dimensional object models as underlying representations (for instance, see Pentland, 1986).

A representational format sufficient for the representation of three-dimensional structure does not automatically qualify as sufficient for recognition. Although there are few *a priori* reasons to think that human object recognition is actually based on parameterized volumetric primitives, Marr and Nishihara (1978) argue that such descriptions meet their criterion of *sensitivity and stability*. Specifically, they suggest that object models capture information that is stable over similar objects (allowing them to be recognized), yet that is also sensitive to finer differences between similar objects (allowing them to be differentiated). In their view, because object models satisfy this and other criteria, they are well suited as representations for object recognition. However, the properties of sensitivity and stability are not limited to models composed of volumetric primitives; rather any part-based representation, either two or three dimensional, may share these properties. Indeed, the reconstruction of three-dimensional structure is distinct computationally from object recognition. The recovery of three-dimensional structure will play a role in recognition only if recognition is based on three-dimensional representations.

Thus, investigators are faced with two separate puzzles: the three-dimensional perception of the world, the *reconstruction problem*, and the ability to recognize objects in the world, the *recognition problem*. Many theorists attempt to solve the latter problem by first solving the former. However, as Marr and Nishihara (1978) point out, even if object models satisfy their criterion of stability and sensitivity, such representations must still satisfy their criterion of *accessibility* -- whether the desired description can be computed relatively efficiently from the image. Although Marr and Nishihara appear to suggest that this is possible for the primitives in their structural descriptions (also see Biederman, 1987; and Pentland, 1986), the reconstruction of three-dimensional structure is computationally difficult, possibly more so than recognition. The evidence presented in this thesis for a multiple-views-plus-transformation mechanism in recognition supports this argument, suggesting that at least some of the time the recognition process does not rely on the reconstruction of three-dimensional structure.

5.5. An alternative theory of object recognition

As mentioned, Biederman (1987) is one of several theorists that have hypothesized that the reconstruction of three-dimensional structure is a prerequisite to recognition. Since Biederman's theory, which he refers to as *Recognition By Components* or RBC, is considered to be one of the preeminent theories of human object recognition, it is worthwhile to examine some aspects of it in greater detail (as well as providing an example of this family of viewpoint-independent theories; see also Marr and Nishihara, 1978; and Pentland, 1986).

The foundations of RBC theory are *non-accidental properties*, which are so named because they occur in the two-dimensional image or edge map as a result of three-dimensional structure, rather than as an accident of the particular two-dimensional projection (non-accidental properties were first enumerated by Lowe, 1985). For example, parallel lines in the two-dimensional image are likely to be parallel in three dimensions as well. Thus, whenever a non-accidental property is detected in a two-dimensional image, there is a high probability that the property is present in the three-dimensional structure. RBC theory assumes that recognition is based on five non-accidental properties: collinearity of points or lines, curvilinearity of a series of points in an arc, symmetry, parallel curves, and vertices indicated by two or more common line terminations at a single point

(Biederman, 1987). Biederman argues that these five non-accidental properties may be mapped directly into three-dimensional primitives by using groupings of binary contrasts (present or not present) to uniquely specify a subset of volumetric primitives drawn from generalized cones (which are generated by moving a cross-section of constant shape but smoothly varying size along an axis of elongation; see Marr and Nishihara, 1978). Specifically, a set of 36 generalized cones, referred to as *geons*, may be generated combining contrasts of the parameters of edge curvature, symmetry, and size variation for the cross section, as well as one parameter of curvature for the axis. Each of these parameters is in turn distinguished by the occurrence of particular non-accidental properties. For example, Biederman (1987; p. 123) suggests that a geon representing an animal horn is parameterized as a curved edge, symmetry, and expanding size for the cross section, as well as a curved axis of elongation.

One of the strengths of RBC is that the constituent primitives of object models, geons, are motivated by image properties that appear to play an important role in the perceptual organization during early human vision (Witkin and Tenenbaum, 1983). Thus, there is some reason to believe that these properties may be recovered, thereby providing the basis for the reconstruction of three-dimensional structure as represented by the relations (such as above or below) between geons in an object model. At this point recognition consists of matching the relations between recovered geons to the stored relations between similar geons in object models.

Another strength of RBC is that by restricting the set of available primitives, Biederman is able to evaluate their adequacy in the representation of objects. Biederman does this by assuming that the 36 geons may be combined into 74,649 two geon objects and about 154 million three geon objects (assuming that on the average any two geons have 57.6 possible spatial relationships). Biederman argues that this number of possible combinations is sufficient for the representation of the 30,000 or so objects that humans are likely to know about. Moreover, since Biederman argues that RBC theory addresses only *primitive access*, recognizing the basic category or class membership of an object, simple geon configurations may be sufficient. For instance, geon models would suffice for the recognition of a chair, but not for the discrimination between two similar chairs. Biederman suggests that this kind of fine differentiation involving essentially identical parts is accomplished by the use alternative representations involving distinctive features and the evaluation of metric spatial relations.

Finally, Biederman (1987) does present evidence from contour deletion studies that are intended to demonstrate that input shapes are parsed into components that correspond to geons. His major finding is that the deletion of contours at inflection points inhibits recognition more than the deletion of equal amounts of contour at non-inflection points. However, these studies are consistent with any theory of shape representation that includes the division of objects into parts, and in particular, any theory that adopts Hoffman and Richards' (1984; also see Bennett and Hoffman, 1987) theory that points of inflection are indicative of part boundaries. Indeed, most theories of recognition, including both viewpoint-dependent and viewpoint-independent theories, adopt this assumption, dividing objects into component parts.

In many ways RBC is the most extensive theory of object recognition available -- however it is still vague about some issues. Most importantly, it does not specify the precise conditions for the recovery of non-accidental properties. Moreover, a more effective means is needed for evaluating the adequacy of geons in the representation of all objects. Simply enumerating the number of possible combinations of geons may not address some important problems in the representation of objects, for instance asymmetries. However these problems may not be insurmountable -- even if the three-dimensional volumetric primitives in RBC theory are replaced with purely two-dimensional parts, it provides one of the best accounts of how the visual system goes from two-dimensional images to part-based representations of objects.

5.5.1. Viewpoint dependency in RBC

While the results of this thesis and Tarr and Pinker (1989a) support the multiple-views-plus-transformation theory and cast doubt on completely viewpoint-independent theories of object recognition, it is not completely inconsistent with Biederman's (1987) theory. Biederman has suggested that geon models are not completely independent of viewpoint. Rather, a particular set of non-accidental properties may suffice for the recovery of a specific configuration of geons over a range of orientations (this is the point of using non-accidental properties -- they are relatively robust over changes in orientation) and that this configuration of geons will be stored as one model of an object. However, for new sets of non-accidental properties within the same object, for instance the same object as seen from the back, different geons may be recovered and stored as a separate model of the object (although there will be some mapping between each of these models into the overall structure of the object). Thus, geon models are to some extent viewpoint dependent -- each model is recoverable from only a limited range of orientations. Further, it is possible that although geons are viewpoint-independent, the spatial relations between geons (such as "on top of") are described with respect to the perspective of the viewer. However, this hypothesis still does not offer an explanation for the systematic effects of orientation on reaction times; for views that contain the same set of non-accidental properties, recognition times should be equivalent. One possible reconciliation for this discrepancy is that geon recognition is used for primitive access (Biederman, 1987), but not complex differentiations among similar objects. Thus, the relations between geons might be described in a one-dimensional ordering of parts similar to that suggested by Tarr and Pinker (1989b), while the more complex relations within input shapes are described in two-dimensional or greater coordinate systems. For primitive access geon descriptions would be sufficient for recognition without invoking mental rotation, but for complex recognition viewpoint-specific descriptions would be required for alignment with similar viewpoint-specific stored representations.

6. Routes to Object Recognition

Evidence for at least two paths to recognition has already been discussed -- along with the viewpoint-dependent theory supported in this thesis, Tarr and Pinker (1989b) argue that when an object need not be described across two dimensions, e.g. when a single directional vector is sufficient for uniquely describing the ordering of an object's parts, recognition is viewpoint-independent. This is true even when more than one particular feature of the object is crucial for recognition; as long as the positions of relevant features (other than their ordering along the description vector) relative to the one-dimensional description vector are irrelevant, recognition remains independent of viewpoint. To some extent, recognition of this sort is an elaboration of the recognition of objects by their unique features, for instance the trunk of an elephant. Certainly, recognition by unique features is uncontroversially one path to recognition (see Eley, 1982). A unique feature is possibly the minimal case of a one-dimensional ordering of parts -- a single ordered part is sufficient for recognition.

One speculation is that viewpoint-independent recognition seems well suited for general category discriminations because many categories of objects seem to be differentiatable by either a single unique part or a small set of parts in a unique order. This is supported by Biederman's (1987) claim that primitive access may be accomplished by the relations between no more than three geons. In contrast, when objects may be discriminated only by complex relationships between parts relative to an ordering of their primary parts, as in the stimuli used in this study (or for instance, how far apart a person's eyes are relative to their nose), recognition shifts to viewpoint-dependent mechanisms, best characterized by the multiple-views-plus-transformation theory.

My speculation is that shaped-based object recognition may be divided between two processes: a viewpoint-independent process best suited to the rapid determination of an object's basic level (a primary goal of recognition according to Bobick, 1987); and a viewpoint-dependent process best suited to the more precise determination of an object's identity. Biederman (1987) has suggested a similar dichotomy, arguing that models composed of two or three geons are sufficient for basic level or primitive access, but that metric information, shading, and other image properties are necessary for within category discriminations. Without yet specifically addressing Biederman's proposal for primitive access, the findings in this study using configurally similar objects suggest that within category discriminations, referred to as complex object recognition, are accomplished by aligning input shapes with stored viewpoint-specific representations. Moreover, as suggested by Tarr and Pinker (1989a), these representations are concrete or pictorial in the sense that they preserve a specific arrangement of an object's parts from a particular viewpoint. In addition to preserving this metric information, such representations appear to be well suited for the preservation of other image properties within objects (for instance shading), many of which are specific to an object as seen from a particular view.

This argument appears to relegate the multiple-views-plus-transformation theory to a circumscribed and uncommon role in object recognition. However, differentiating between members within a class may be far more common than one might think -- for numerous tasks, such as picking out a handtool from a toolbox filled with tools (Figure 23) or recognizing a particular model of Toyota, the simple ordering of parts is insufficient and recognition must rely on the more complex spatial relations described in stored viewpoint-specific representations. Moreover, as discussed previously there are several empirical, physiological, and ecological considerations consistent with there being an important role for a multiple-views-plus-transformation mechanism in recognition. To review briefly, there exists a large body of literature demonstrating that recognition fails for many classes of misoriented objects, for instance faces, maps, and wire forms (see Rock, 1973, 1983; Rock and Di Vita, 1987), and that even familiar objects have a preferred view or set of views (Palmer, et. al., 1981). There is also evidence that neurons in animals are sensitive to faces of the same species only in commonly encountered orientations (Kendrick and Baldwin, 1987; Perrett, et. al. 1987) and that in humans brain lesions due to stroke may impair the recognition of objects in noncanonical orientations (Layman and Greene, 1988).

Thus, it appears that under many circumstances, object recognition is viewpoint-dependent, relying on mechanisms such as multiple-views-plus-transformation.

Insert Figure 23 about here.

6.0.1. Ecological considerations

What advantages are afforded by the availability of two paths for recognition? Consider an observer in the world: although I have argued that viewpoint-dependent recognition is useful, if not common, much of an organism's day-to-day interaction with the environment requires only determining general categories of objects, for instance a chair, a car, or a hungry carnivore (the canonical tiger in the jungle). Thus, the strategy of recognition on which an organism will rely for the majority of its visual information need only constrain itself to general hypotheses about categories (Bobick, 1987). In addition, recognition by this "default" strategy should be fast (it is rarely advantageous to be eaten) and should filter out the majority of visual information irrelevant to the task. This is how I have characterized viewpoint-independent recognition. Not only does it provide only primitive access, the minimal information required for recognition, but it is, by definition, reliable over many viewpoints. Moreover, because of these properties it is relatively fast as well -- a match may be located by accessing only a small number of parts and there is no need to invoke a time consuming normalization process, such as mental rotation.

In contrast, there are also reasons for the existence of a viewpoint-dependent recognition strategy. As mentioned, there are many situations that necessitate differentiating between individuals within a class. In such situations, however, speed of recognition and limiting information content are not at a premium. Rather, the emphasis is on accurate recognition, best accomplished by preserving information so that a precise match may be made. This is how I have characterized viewpoint-dependent recognition. Not only does it preserve information within input shapes by mentally rotating[17] them to stored representations, but by storing representations of objects in multiple-views, it preserves the viewpoint-specific image properties, such as metric distances and shading information. Moreover, because matches are made between these properties within two representations of an object from roughly the same view, viewpoint-dependent recognition provides relatively greater accuracy in determining the specific identity of a particular object. Here however, there is a cost -- recognition is relatively slow due to the utilization of a time-consuming normalization process, mental rotation, and it is viewpoint-dependent, sometimes failing when objects are encountered in unfamiliar orientations.

6.0.2. Adaptive pressures

These characterizations raise the question of what adaptive benefits do two routes to recognition provide? While it is clear that the viewpoint-independent discrimination of general categories is crucial for survival and therefore advantageous, it is interesting to consider those areas in which viewpoint-dependent discrimination of specific identity within class may be advantageous. One possibility is that viewpoint-dependent recognition is useful in two important primate behaviors: tool making and kin identification. Both of these behaviors may be somewhat unique to primates -- complex tool use exists almost solely in primates (and in particular in humans) and many other species identify kin by non-visual means, for instance scent. In both behaviors it is inadequate to determine only the basic category of an object: first, differentiating and consistently duplicating particular tools seems to require metric information; and second, visually identifying individuals within one's own species certainly requires noticing subtle variations among like parts (which is one reason why some researchers have proposed a specialized area of the brain for faces). Thus, while it is possible that many species have viewpoint-independent mechanisms for object recognition, only needs such as visually identifying mates and

---

[17] An important characteristic of mental rotation is that it preserves information -- in particular, the relations between parts over transformation (Shepard and Cooper, 1982).

offspring or accurately replicating particular tools made the existence of a viewpoint-dependent mechanism and second route to object recognition advantageous.

7. Appendix: Construction of Stimuli

All stimuli were drawn with the Cubicomp ModelMaker300 solid modeling system in perspective projection and were designed so that their spatial center coincided with the center of the modeling space. This ensured that all rotations were around the spatial center of the object. The basic set of seven objects, referred to as the *asymmetrical set* (Figure 1), were constructed from cubes connected at the faces. These objects are somewhat similar in appearance to those used by Metzler and Shepard (1974) and Parsons (1987c). Each of the objects shared a main vertical axis seven cubes high with a cube attached to each horizontal side of the bottom cube, thus forming a string of three cubes that clearly marked the bottom of the main axis. Cubes were attached to this axis with the following constraints:

- All objects were asymmetrical across the sagittal, frontal, and horizontal planes.
- Each object contained a string of seven cubes that crossed the main vertical axis through either the sagittal or the frontal plane.
- No other string of cubes on an object was longer than six cubes.
- No cubes were attached to either the top or the bottom cube (other than the cubes marking the bottom of the main axis) of the main vertical axis.
- No cube was attached to the face of a cube when either cube adjacent to that cube along the main axis had a cube attached to the same face.

Standard versions of each objects were determined arbitrarily. Reversed versions of the objects (enantiomorphs) were generated by reflecting the object at upright through the sagittal plane (reversing left and right, but not front and back). Rotations of reversed versions were performed only after the reversal.

Two similar sets of seven objects each were constructed according to the same constraints with additional symmetry constraints. The *left/right symmetrical set* (Figure 16) was designed with symmetry across the sagittal plane (left to right) and with asymmetry across the frontal plane (front to back), while the *left/right-front/back symmetrical set* (Figure 17) was designed with symmetry across both the sagittal and frontal planes, buy not radial symmetry between the sagittal and frontal planes. For both of these sets the left/right symmetry made it impossible to generate enantiomorphs that differed from the original objects.

8. References

Bennett, B. M., and Hoffman, D. D. (1987). Shape decompositions for visual recognition: the role of transversality. In W. Richards and S. Ullman (Eds.), *Image Understanding 1985-86.* Norwood, NJ: Ablex Publishing Corporation.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94* (2), 115-147.

Biederman, I. (In Press). Aspects and extensions of a theory of human image understanding. In Z. W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An Interdisciplinary Perspective.* Norwood, NJ: Ablex Publishing Corporation.

Biederman, I., and Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20,* 38-64.

Bobick, A. F. (1987) Natural object categorization. *MIT A.I. Technical Report, 1001.*

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7,* 20-43.

Cooper, L. A. (1976). Demonstration of a mental analog of an external rotation. *Perception & Psychophysics, 19,* 296-302.

Cooper, L. A., and Podgorny, P. (1976). Mental transformations and visual comparison processes. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 503-514.

Cooper, L. A., and Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual Information Processing.* New York: Academic Press.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review, 95* (1), 115-123.

Corballis, M. C., and McLaren, R. (1984). Winding one's Ps and Qs: mental rotation and mirror-image discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 10* (2), 318-327.

Corballis, M. C., and Nagourney, B. A. (1978). Latency to categorize disoriented alphanumeric characters as letters or digits. *Canadian Journal of Psychology, 32* (3), 186-188.

Corballis, M. C., Zbrodoff, N. J., Shetzer, L. I., and Butler, P. B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory & Cognition, 6*, 98-107.

Eley, M. G. (1982). Identifying rotated letter-like symbols. *Memory & Cognition, 10* (1), 25-32.

Folk, M. D., and Luce, R. D. (1987) Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance, 13* (3), 395-404.

Haldane, J. B. S. (1927/1985). On being the right size. In J. Maynard Smith (Ed.), *On Being the Right Size and Other Essays*. Oxford: Oxford University Press.

Hinton, G. E., and Parsons, L. M. (1981). Frames of reference and mental imagery. In J. Long and A. Baddeley (Eds.), *Attention and Performance IX*. Hillsdale, NJ: Lawrence Erlbaum.

Hinton, G. E., and Parsons, L. M. (1988). Scene-based and viewer-centered representations for comparing shapes. *Cognition, 30*, 1-35.

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

Huttenlocher, D. P., and Ullman, S. (1987a). Recognizing rigid objects by aligning them with an image. *M.I.T. A.I. Memo, 937.*

Huttenlocher, D. P., and Ullman, S. (1987b). Object recognition using alignment. *Proceedings of the First International Conference on Computer Vision*. IEEE Computer Society Press, 102-111.

Huttenlocher, D. P., and Ullman, S. (1988). Recognizing solid objects by alignment. *Manuscript submitted for publication.*

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13* (4), 289-303.

Jolicoeur, P., and Landau, M. J. (1984). Effects of orientation on the identification of simple visual patterns. *Canadian Journal of Psychology, 38* (1), 80-93.

Kaushall, P. and Parsons, L. M. (1981). Optical information and practice in the discrimination of 3-D mirror-reflected objects. *Perception, 10,* 545-562.

Kendrick, K. M., and Baldwin, B. A. (1987). Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. *Science, 236,* 448-450.

Koenderink, J. J. (1987). An internal representation for solid shape based on the topological properties of the apparent contour. In W. Richards and S. Ullman (Eds.), *Image Understanding 1985-86.* Norwood, NJ: Ablex Publishing Corporation.

Koenderink, J. J., and van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics, 32,* 211-216.

Koriat, A., and Norman, J. (1985). Mental rotation and visual familiarity. *Perception & Psychophysics, 37,* 429-439.

Layman, S., and Greene, E. (1988). The effect of stroke on object recognition. *Brain and Cognition, 7,* 87-114.

Lindsay, P. H., and Norman, D. A. (1977). *Human Information Processing: An Introduction to Psychology,* 2nd ed. New York: Academic Press.

Lowe, D. G. (1985). *Perceptual Organization and Visual Recognition.* Boston, MA: Kluwer Academic Publishers.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31,* 355-395.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: Freeman.

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

Metzler, J., and Shepard, R. N. (1974). Transformational studies of the internal representation of three-dimensional objects. In R. L. Solso (Ed.), *Theories of Cognitive Psychology: The Loyola Symposium.* Potomac, MD: LEA.

Palmer, S., Rosch, E., and Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long and A. Baddeley (Eds.), *Attention and Performance IX.* Hillsdale, NJ: Lawrence Erlbaum.

Parsons, L. M. (1987a). Imagined spatial transformation of one's body. *Journal of Experimental Psychology: General, 116* 2, 172-191.

Parsons, L. M. (1987b). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology, 19,* 178-241.

Parsons, L. M. (1987c). Visual discrimination of abstract mirror-reflected three-dimensional objects at many orientations. *Perception & Psychophysics, 42* (1), 49-59.

Pentland, A. P. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence, 28,* 293-331.

Perrett, D. I., Mistlin, A. J., and Chitty, A. J. (1987). Visual neurones responsive to faces. *Trends in Neuroscience, 10* (9), 358-364.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition, 18,* 1-63.

Robertson, L. C., Palmer, S. E., and Gomez, L. M. (1987). Reference frames in mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13* (3), 368-379.

Rock, I. (1973). *Orientation and Form.* New York: Academic Press.

Rock, I. (1983). *The Logic of Perception.* Cambridge, MA: The MIT Press.

Rock, I., and Di Vita, J. (1987) A case of viewer-centered object perception. *Cognitive Psychology, 19,* 280-293.

Rock, I., Di Vita, J., and Barbeito, R. (1981). The effect on form perception of change of orientation in the third dimension. *Journal of Experimental Psychology: Human Perception and Performance, 7* (4), 719-732.

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

Shepard, S., and Metzler, D. (1988). Mental rotation: effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance, 14* (1), 3-11.

Shinar, D., and Owen, D. H. (1973). Effects of form rotation on the speed of classification: the development of shape constancy. *Perception & Psychophysics, 14* (1), 149-154.

Simion, F., Bagnara, S., Roncato, S., and Umilta, C. (1982). Transformation processes upon the visual code. *Perception & Psychophysics, 31* (1), 13-25.

Tarr, M. J., and Pinker, S. (1989a). Mental rotation and orientation- dependence in shape recognition. *Cognitive Psychology, 21* (2), 233-282.

Tarr, M. J., and Pinker, S. (1989b) Human object recognition uses al viewer-centered reference frame. *Manuscript submitted for publication.*

Treisman, A., and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Ullman, S. (1979). *The Interpretation of Visual Motion.* Cambridge, MA: The MIT Press.

Ullman, S. (1986). An approach to object recognition: aligning pictorial descriptions. *M.I.T. A.I. Memo, 931.*

White, M. J. (1980). Naming and categorization of tilted alphanumeric characters do not require mental rotation. *Bulletin of the Psychonomic Society, 15* (3), 153-156.

Witkin, A. P., and Tenenbaum, J. M. (1983) On the role of structure in vision. In A. Rosenfeld, B. Hope, and J. Beck (Eds.), *Human and Machine Vision.* New York: Academic Press.

9. Table and Figure Captions

9.0.1. Table captions

Table 1. Rates of rotation (in degrees/second) from mental rotation studies using three-dimensional stimuli and from Experiments 1-5 broken down by axis of rotation. Axis yielding the *slowest* rate of rotation is underlined for each experiment. Data for Metzler and Shepard (1974) taken from Shepard and Metzler (1988).

Table 2. Mean percent errors for Blocks 1, 12, and 13 of Experiment 1.

Table 3. Mean percent errors for Blocks 1, 12, and 13 of Experiment 2.

Table 4. Mean percent errors for Blocks 1, 12, and 13 of the Both-Versions Condition of Experiment 3.

Table 5. Mean percent errors for Blocks 1, 12, and 13 of the One-Version Condition of Experiment 3.

Table 6. Mean percent errors for Blocks 1, 12, and 13 of Experiment 4.

Table 7. Mean percent errors for Blocks 1, 12, and 13 of Condition AS of Experiment 5.

Table 8. Mean percent errors for Blocks 1, 12, and 13 of Condition SS of Experiment 5.

9.0.2. Figure captions

Figure 1. Standard versions of asymmetrical objects in their near-upright training orientation. In each of these objects the bottom of the object is marked by a "foot" of three cubes that terminates the main axis, marking it as well.

Figure 2. Axes of rotation with arrows indicating the direction of rotation starting at 0°. A rotation around one axis is accompanied by constant rotations of 10° around the other two axes.

Figure 3. (a) Angular layout of practice orientations in all experiments with the exception of the One-Version Condition of Experiment 3 where the practice orientations were the training orientation and 40°, 70°, 100°, and 190° around each axis. (b) Standard version of Object 1 in the four practice orientations.

Figure 4. Mean handedness discrimination times collapsed over version for early (Block 1) and late (Block 12) trials in Experiment 1. Each axis of rotation is displayed separately.

Figure 5. Changes in slopes with practice as measured by block number in Experiment 1 (smaller slopes reflect faster putative rates of rotation). Each axis of rotation is displayed separately.

Figure 6. Mean handedness discrimination times collapsed over version for new, never-before-seen orientations (Block 13) in Experiment 1. Each axis of rotation is displayed separately and familiar orientations (0° and 130°) are plotted as separate points.

Figure 7. Slopes for Blocks 1, 12, 13 (familiar orientations), 13 (new orientations), and 13 (new orientations in the 160° to 340° range) of Experiment 1. Each graph indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 8. Mean recognition times for new, never-before-seen orientations (Block 13) in Experiment 2. Each axis of rotation is displayed separately and familiar orientations (0° and 130°) are plotted as separate points.

Figure 9. Slopes for Blocks 1, 12, 13 (familiar orientations), 13 (new orientations), and 13 (new orientations in the 160° to 340° range) of Experiment 2. Each graph indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 10. Mean recognition times collapsed over axis of rotation for early trials (Block 1) in (a) the Both-Versions and (b) the One-Version Conditions of Experiment 3. Each handedness version is displayed separately.

Figure 11. Mean recognition times collapsed over version for new, never-before-seen orientations (Block 13) in the Both-Versions Condition of Experiment 3. Each axis of rotation is displayed separately and familiar orientations (0° and 130°) are plotted as separate points.

Figure 12. Slopes for Blocks 1, 12, 13 (familiar orientations), 13 (new orientations), and 13 (new orientations in the 160° to 340° range) of the Both-Versions Condition of Experiment 3. Each graph is collapsed over version and indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 13. Slopes for Blocks 1, 12, 13 (familiar orientations), 13 (new orientations), and 13 (new orientations in the 220° to 340° range) of the One-Version Condition of Experiment 3. Each graph is collapsed over version and indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 14. Mean recognition times broken down by (a) standard and (b) reversed handedness versions for new, never-before-seen orientations (Block 13) in Experiment 4. Each axis of rotation is displayed separately and familiar orientations (0° and 130°) are plotted as separate points.

Figure 15. Slopes for Blocks 1, 12, 13 (familiar orientations), 13 (new orientations), and 13 (new orientations in the 160° to 340° range) of Experiment 4. In Block 13, the first block to contain reversed versions, slopes are broken down by handedness version. Each graph indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 16. Left/Right symmetrical, Front/Back asymmetrical objects in their near-upright training orientation. In each of these objects the bottom of the object is marked by a "foot" of three cubes that terminates the main axis, marking it as well.

Figure 17. Left/Right symmetrical, Front/Back symmetrical objects in their near-upright training orientation. In each of these objects the bottom of the object is marked by a "foot" of three cubes that terminates the main axis, marking it as well.

Figure 18. Mean recognition times for Block 1 in Condition AS and Condition SS of Experiment 5. Each graph indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 19. Mean recognition times for new, never-before-seen orientations (Block 13) in Condition AS of Experiment 5. Each axis of rotation is displayed separately and familiar orientations (0° and 130°) are plotted as separate points.

Figure 20. Slopes for Blocks 1, 12, 13 (familiar orientations), 13 (new orientations), and 13 (new orientations in the 160° to 340° range) of Condition AS of Experiment 5. Each graph indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 21. Mean recognition times for new, never-before-seen orientations (Block 13) in Condition SS of Experiment 5. Each axis of rotation is displayed separately and familiar orientations (0° and 130°) are plotted as separate points.

Figure 22. Slopes for Blocks 1, 12, 13 (familiar orientations), 13 (new orientations), and 13 (new orientations in the 160° to 340° range) of Condition SS of Experiment 5. Each graph indicates the axis of rotation with (a) being around the x axis, (b) around the y axis, and (c) around the z axis.

Figure 23. Finding the right tool in a toolbox.

Experiment	**X Axis**	**Y Axis** (In Depth)	**Z Axis** (In Plane)
	Slope (deg/sec)		
Metzler & Shepard (1974)			
Experiment I		64	46
Experiment II (mixed)		40	50
Experiment II (pure)		38	47
Shepard & Metzler (1988)			
One-Stimulus		343	
Two-Stimulus		129	
Parsons (1987c)	67	42	35
This Study			
Experiment 1 -- Handedness			
Block 1	61	82	94
Block 13 (160°-340°)	74	85	556
Experiment 2 -- Recognition			
Block 1	76	167	125
Block 13 (160°-340°)	76	119	169
Experiment 3 -- Both-Versions			
Block 1	104	175	139
Block 13 (160°-340°)	149	156	208
Experiment 3 -- One-Version			
Block 1	414	469	268
Block 13 (220°-340°)	149	154	167
Experiment 4			
Block 1	79	233	141
Block 13 Standard (160°-340°)	108	233	123
Block 13 Reversed (160°-340°)	103	244	286
Experiment 5 -- AS Condition			
Block 1	64	154	114
Block 13 (160°-340°)	106	500	196
Experiment 5 -- SS Condition			
Block 1	137	263	154
Block 13 (160°-340°)	64	127	345

Table 1

Experiment 1	**10°**	40°	70°	100°	**130°**	160°	190°	220°	250°	280°	310°	340°
Block 1												
x	9.0				34.0							
y	9.0				22.9							
z	9.0				18.8							
Block 12												
x	2.1				1.4							
y	2.1				3.5							
z	2.1				1.4							
Block 13												
x	0.7	4.2	4.9	2.8	1.4	6.3	14.6	22.9	18.8	6.9	5.6	2.1
y	0.7	4.2	4.9	11.1	6.3	5.6	15.3	7.6	9.7	6.3	4.2	4.9
z	0.7	6.3	4.2	1.4	2.8	2.8	8.3	12.5	5.6	0.7	2.8	2.1

Table 2

Experiment 2	**10°**	40°	70°	100°	**130°**	160°	190°	220°	250°	280°	310°	340°
Block 1												
x	4.2				30.6							
y	4.2				18.1							
z	4.2				22.2							
Block 12												
x	0.2				2.8							
y	0.2				1.4							
z	0.2				2.1							
Block 13												
x	0.7	23.6	15.3	12.5	2.8	4.2	10.4	18.1	26.4	11.1	12.5	2.8
y	0.7	1.4	8.3	22.2	2.8	5.6	3.5	4.9	14.6	4.9	4.2	2.1
z	0.7	4.2	2.1	2.8	0.7	5.6	6.3	13.9	4.9	4.2	4.9	0.7

Table 3

Experiment 3 Both-Versions	**10°**	40°	70°	100°	**130°**	160°	190°	220°	250°	280°	310°	340°
Block 1												
Standard												
x	8.3				18.1							
y	8.3				11.1							
z	8.3				15.3							
Reversed												
x	4.6				19.4							
y	4.6				9.7							
z	4.6				23.6							
Block 12												
x	1.4				4.2							
y	1.4				0.7							
z	1.4				2.1							
Block 13												
x	0.5	15.3	21.5	6.3	2.1	4.2	4.2	13.2	7.6	14.6	13.2	5.6
y	0.5	0.0	6.3	9.7	0.0	3.5	2.8	4.9	5.6	11.8	7.6	2.1
z	0.5	11.8	2.1	1.4	0.7	3.5	3.5	15.3	8.3	3.5	1.4	1.4

Table 4

Experiment 3 One-Version	**10°**	**40°**	**70°**	100°	**130°**	160°	**190°**	220°	250°	280°	310°	340°
Block 1												
Standard												
x	4.6	23.6	16.7		20.8		18.1					
y	4.6	8.3	22.2		5.6		2.8					
z	4.6	23.6	5.6		9.7		13.9					
Reversed												
x	6.0	29.2	26.4		22.2		18.1					
y	6.0	11.1	12.5		11.1		12.5					
z	6.0	12.5	18.1		12.5		11.1					
Block 12												
x	1.9	4.9	2.8		4.2		0.7					
y	1.9	2.1	4.9		1.4		0.7					
z	1.9	0.7	2.1		2.8		2.1					
Block 13												
x	1.9	4.2	1.4	9.7	1.4	4.2	1.4	16.0	15.3	19.4	15.3	4.9
y	1.9	0.7	0.7	14.6	2.1	3.5	2.1	2.8	12.5	11.1	6.9	6.9
z	1.9	4.9	1.4	4.9	1.4	3.5	1.4	15.3	5.6	7.6	7.6	1.4

Table 5

Experiment 4	**10°**	40°	70°	100°	**130°**	160°	190°	220°	250°	280°	310°	340°
Block 1												
x	12.7				45.1							
y	12.7				20.1							
z	12.7				25.0							
Block 12												
x	0.2				1.4							
y	0.2				2.1							
z	0.2				1.4							
Block 13												
Standard												
x	0.0	22.2	20.8	4.2	1.4	4.2	2.8	19.4	15.3	23.6	19.4	4.2
y	0.0	2.8	11.1	18.1	0.0	2.8	2.8	1.4	12.5	9.7	1.4	1.4
z	0.0	6.9	2.8	0.0	1.4	2.8	8.3	9.7	8.3	4.2	5.6	0.0
Reversed												
x	1.4	19.4	13.9	13.9	5.6	6.9	8.3	33.3	16.7	19.4	9.7	5.6
y	1.4	2.8	1.4	9.7	0.0	1.4	4.2	0.0	9.7	6.9	8.3	5.6
z	1.4	5.6	1.4	0.0	2.8	1.4	5.6	22.2	5.6	0.0	2.8	1.4

Table 6

Experiment 5 Condition AS	**10°**	40°	70°	100°	**130°**	160°	190°	220°	250°	280°	310°	340°
Block 1												
x	5.3				41.7							
y	5.3				9.7							
z	5.3				16.0							
Block 12												
x	0.9				1.4							
y	0.9				0.0							
z	0.9				1.4							
Block 13												
x	0.5	32.7	16.7	6.9	0.7	4.2	4.9	31.9	13.2	6.9	4.2	2.8
y	0.5	0.7	0.7	2.1	0.7	0.7	5.6	0.0	9.7	5.6	2.8	0.0
z	0.5	16.0	1.4	1.4	3.5	0.7	0.7	15.3	2.8	2.8	2.8	0.7

Table 7

Experiment 5 Condition SS	**10°**	40°	70°	100°	**130°**	160°	190°	220°	250°	280°	310°	340°
Block 1												
x	4.9				16.0							
y	4.9				11.1							
z	4.9				17.4							
Block 12												
x	1.2				2.1							
y	1.2				1.4							
z	1.2				3.5							
Block 13												
x	0.5	42.4	20.8	9.0	4.9	1.4	2.1	42.4	17.4	13.2	3.5	2.8
y	0.5	1.4	0.7	6.9	0.7	0.0	1.4	1.4	4.2	3.5	1.4	2.1
z	0.5	7.6	0.7	2.1	0.7	2.1	2.8	13.9	2.1	2.1	2.1	2.1

Table 8

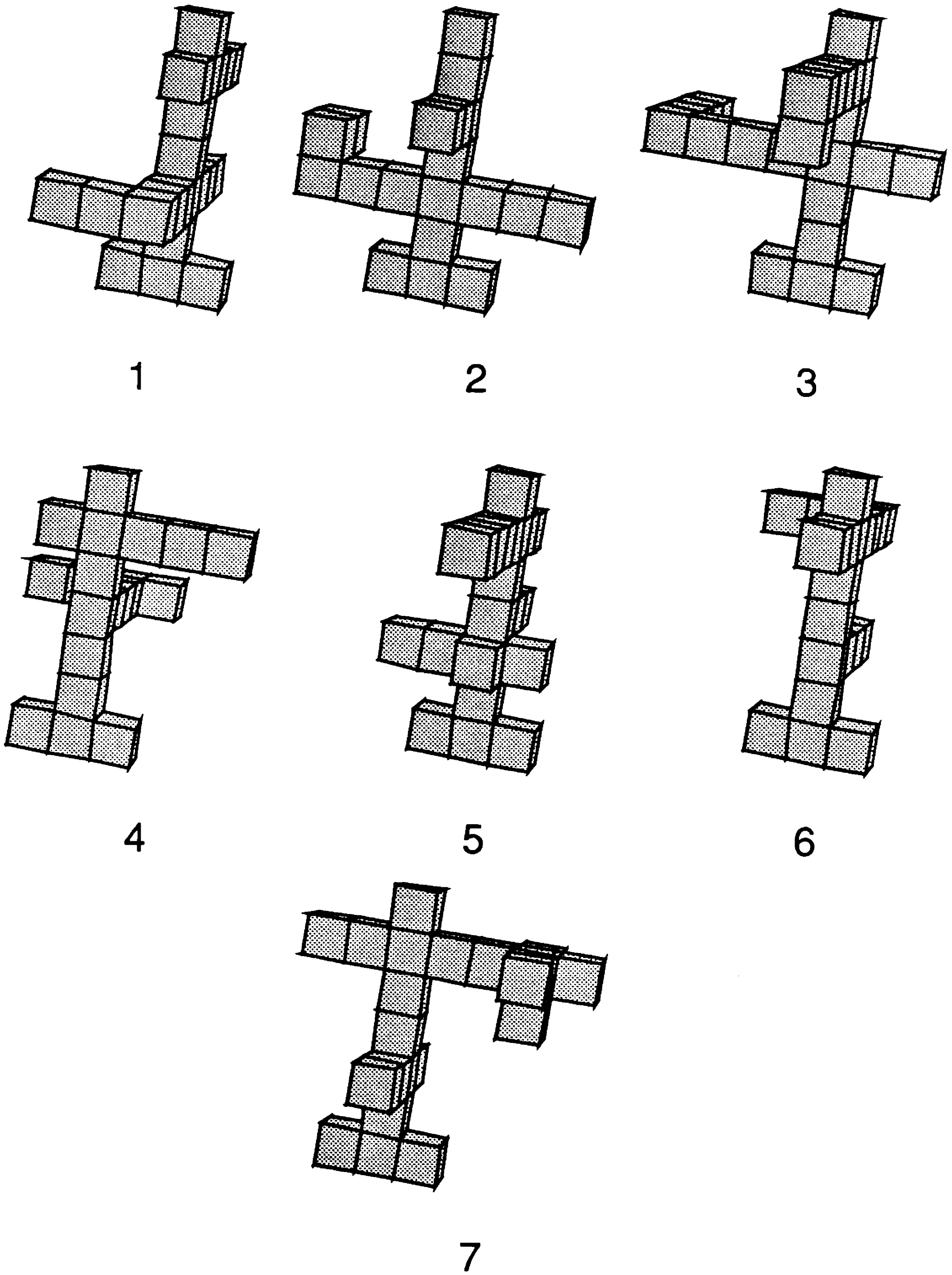

1 2 3 4 5 6 7

Figure 1

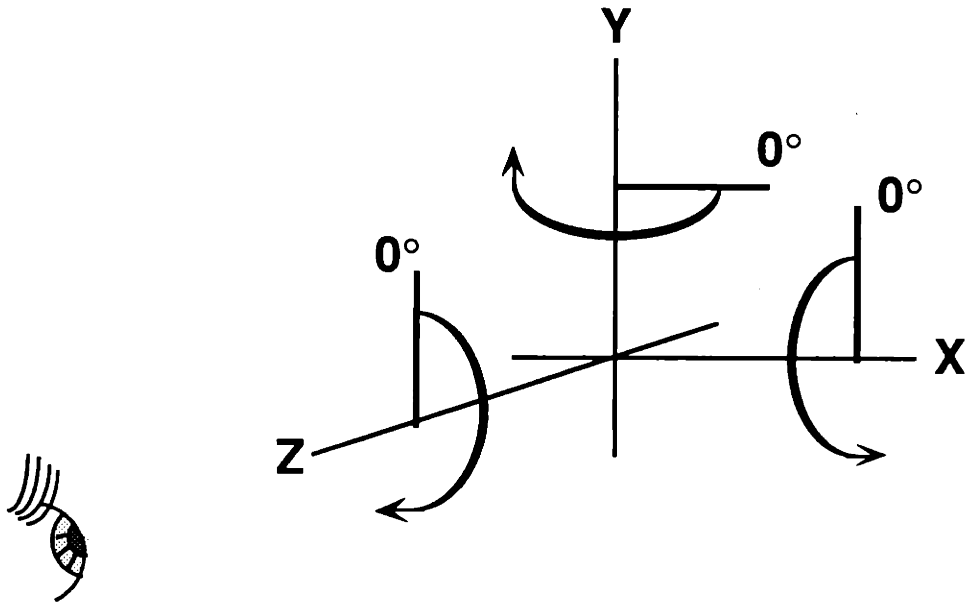
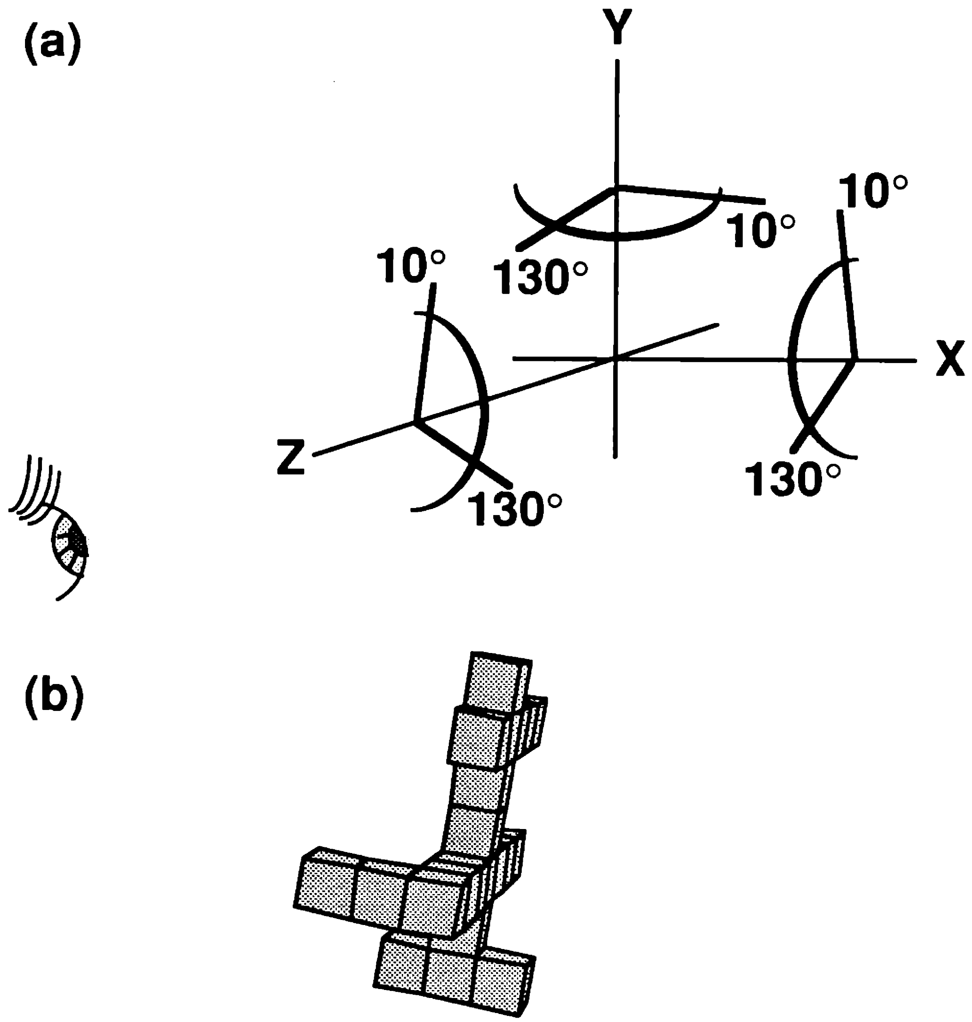
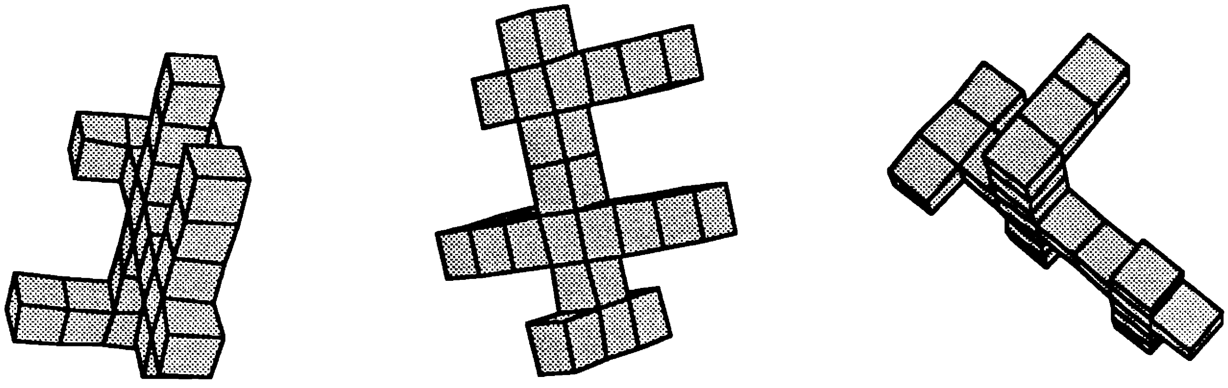


Figure 2



Training (10°,10°,10°)



X
130°

Y
130°

Z
130°

Figure 3

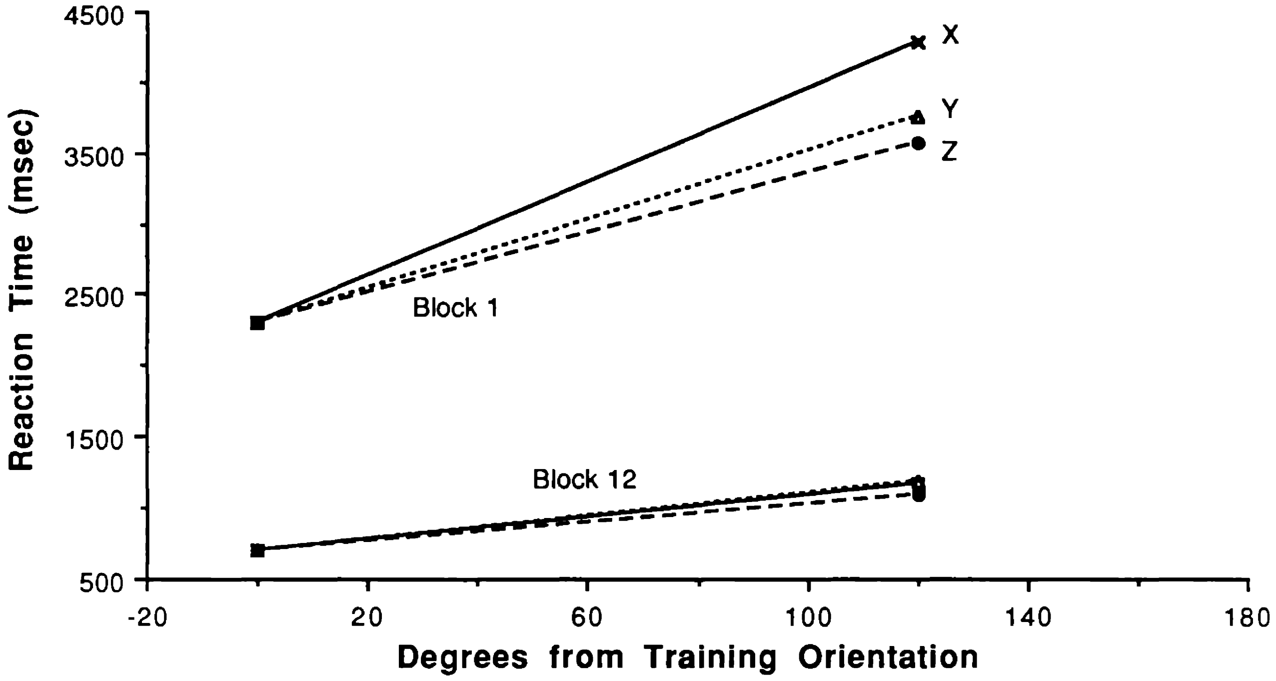

Reaction Time (msec)
4500
3500
2500
1500
500
-20
20
60
100
140
180
Degrees from Training Orientation
X
Y
Z
Block 1
Block 12

Figure 4

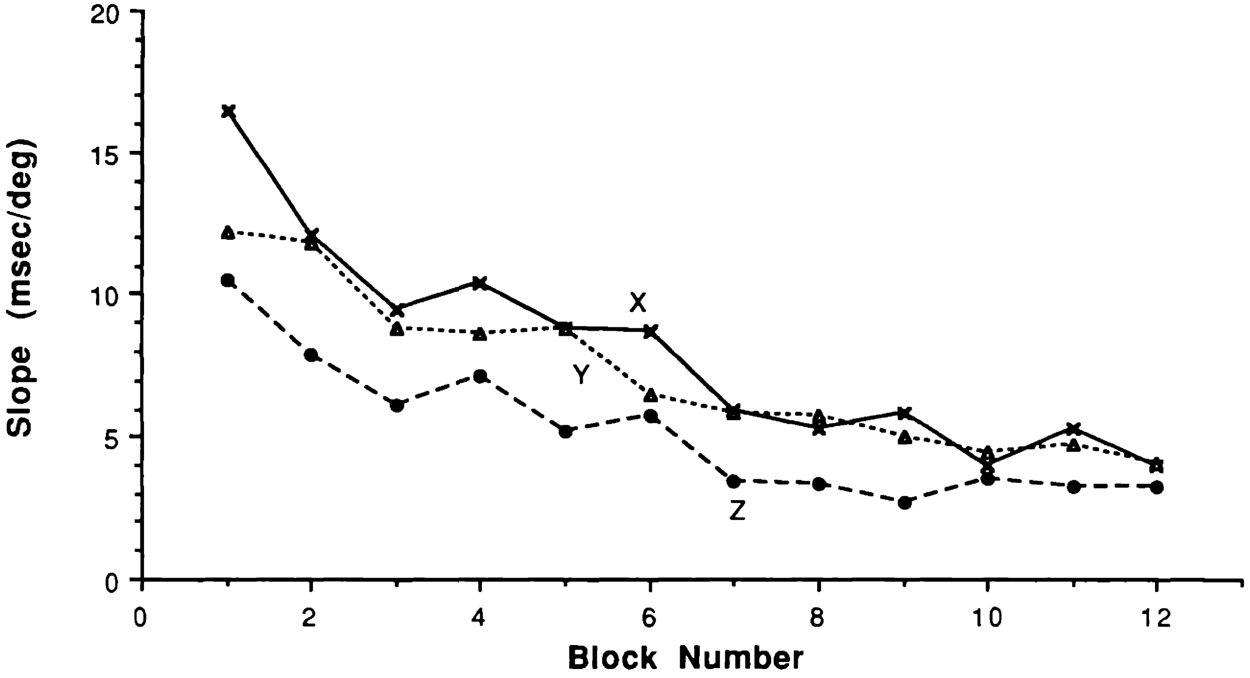


Figure 5

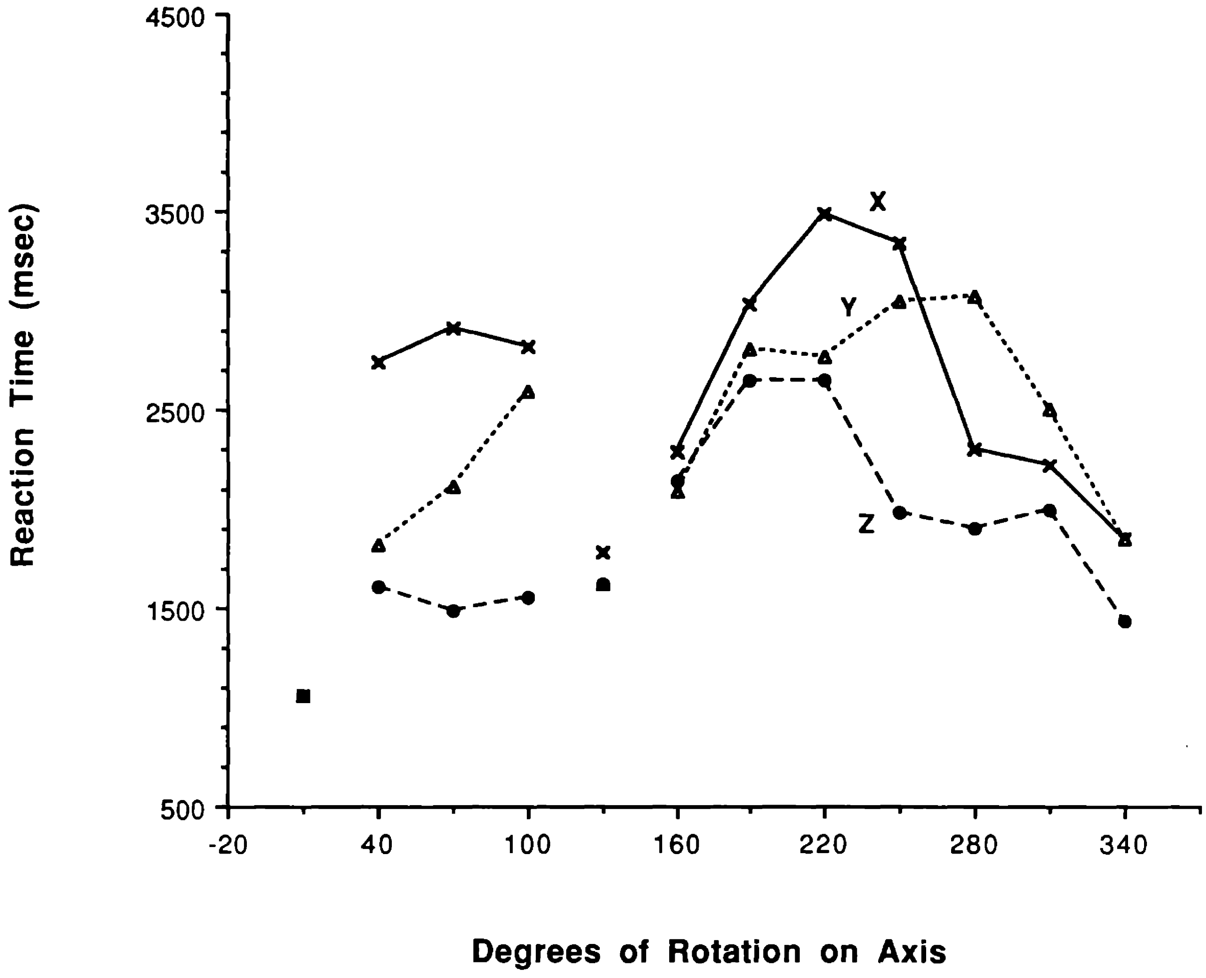

4500
3500
2500
1500
500
Reaction Time (msec)
-20
40
100
160
220
280
340
Degrees of Rotation on Axis
X
Y
Z

Figure 6

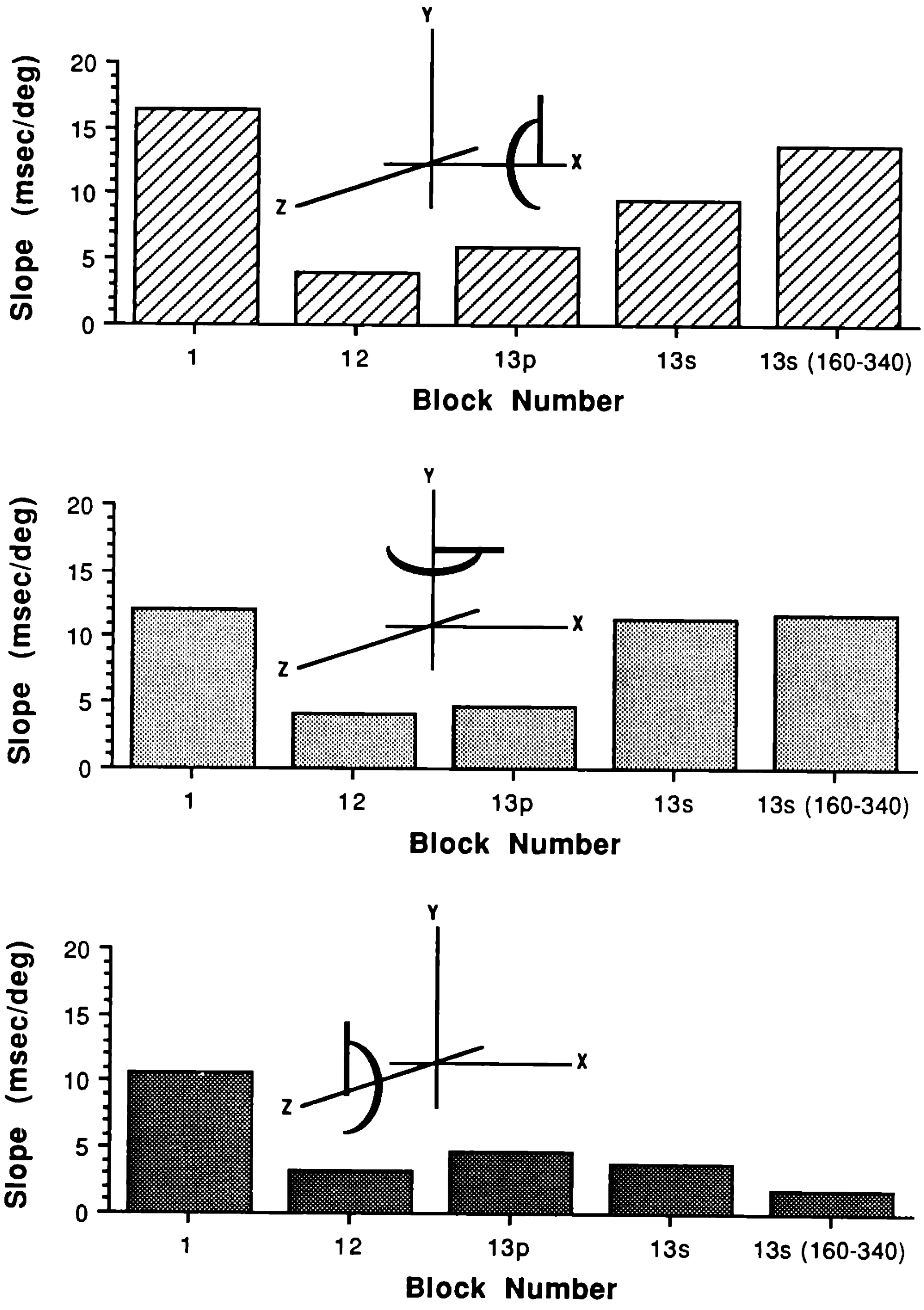
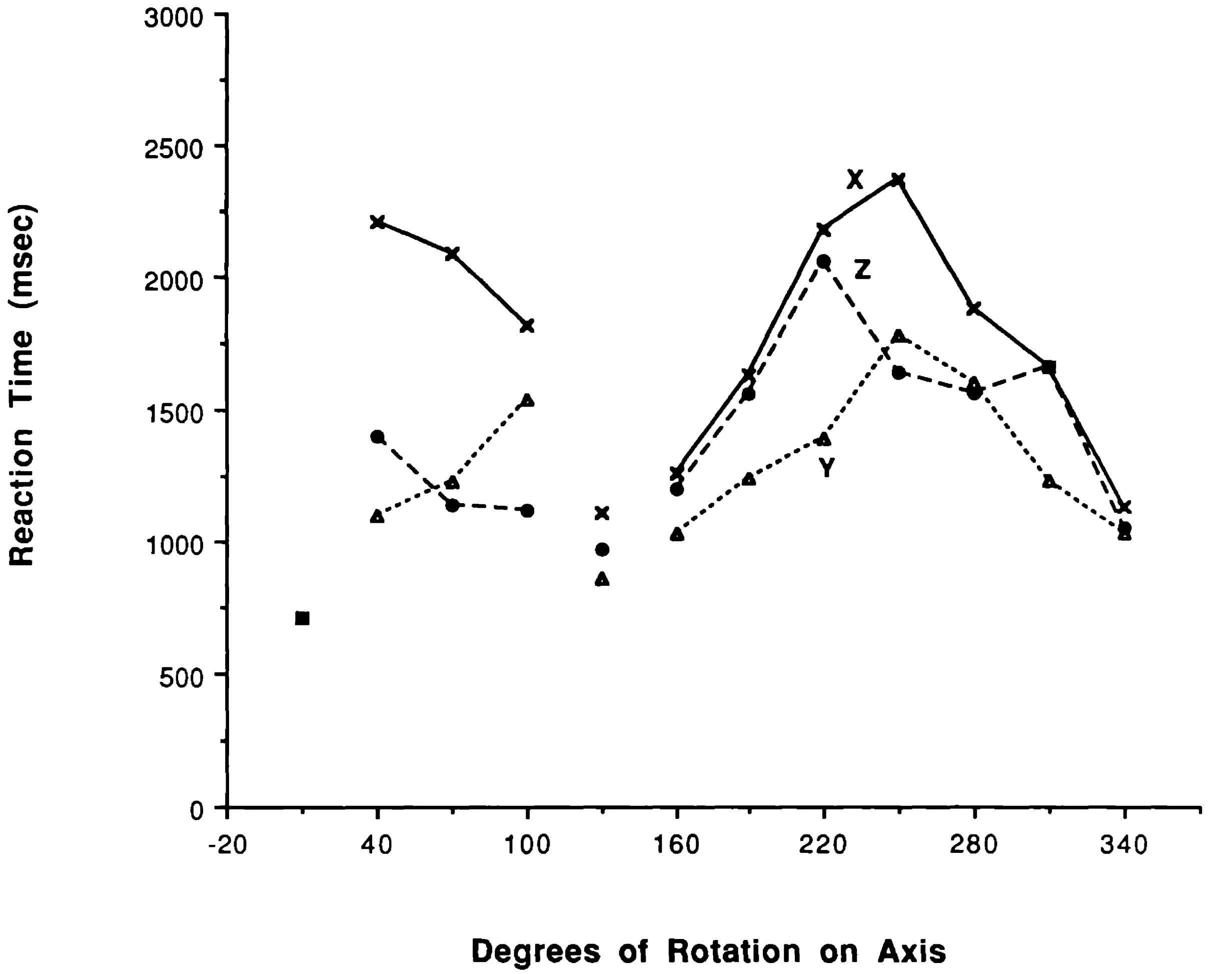

Y
X
Z
Slope (msec/deg)
0
5
10
15
20
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
0
5
10
15
20
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
0
5
10
15
20
1
12
13p
13s
13s (160-340)
Block Number

Figure 7



Reaction Time (msec)
3000
2500
2000
1500
1000
500
0
-20
40
100
160
220
280
340
Degrees of Rotation on Axis
X
Z
Y

Figure 8

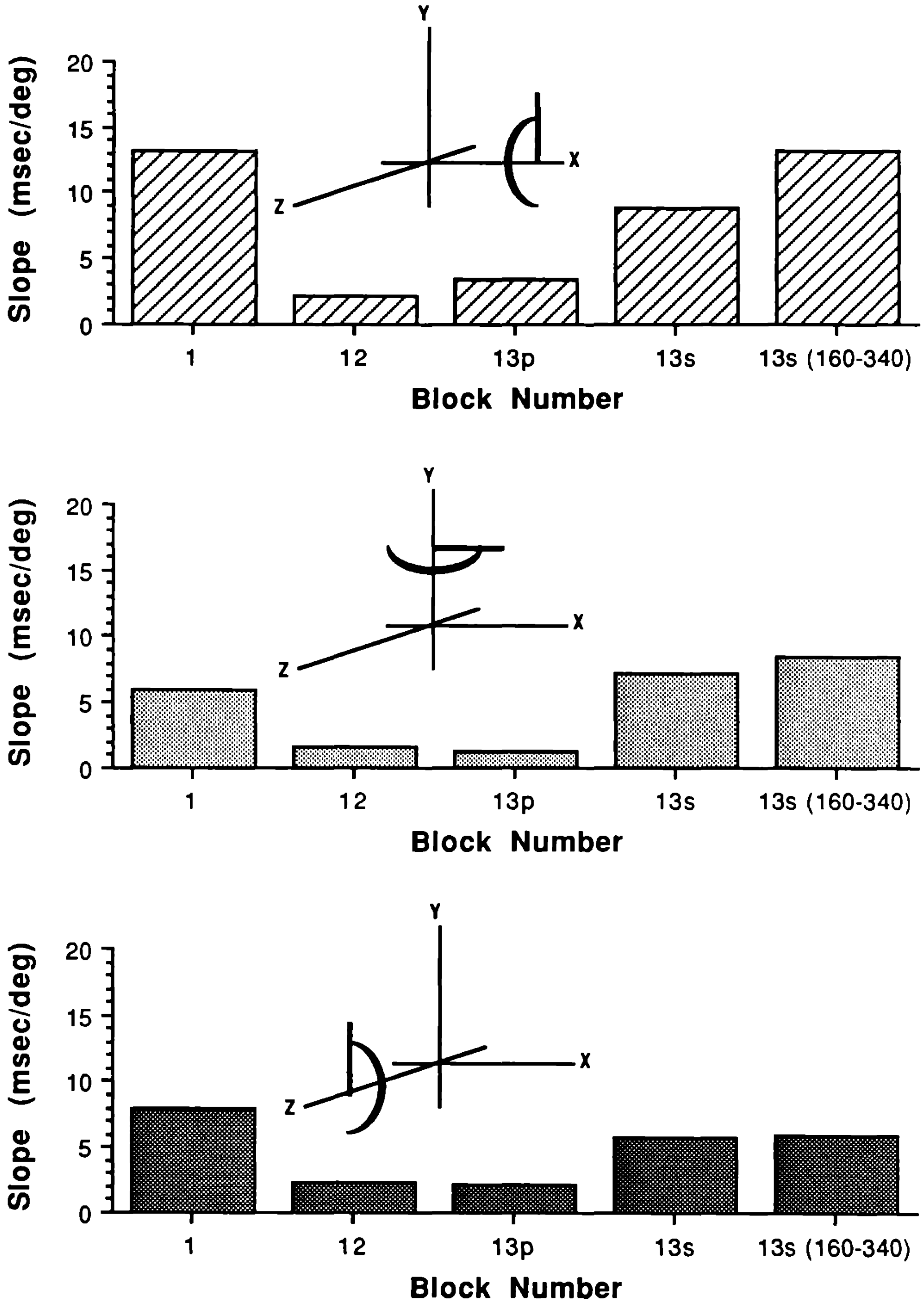
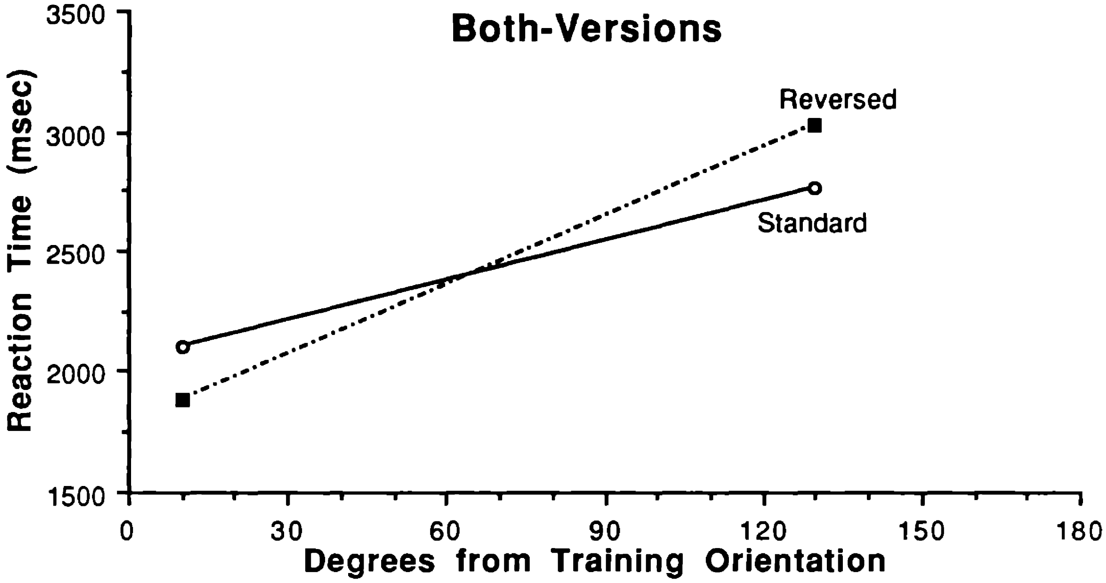

Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number

Figure 9


Both-Versions
Reversed
Standard
Reaction Time (msec)
3500
3000
2500
2000
1500
0
30
60
90
120
150
180
Degrees from Training Orientation

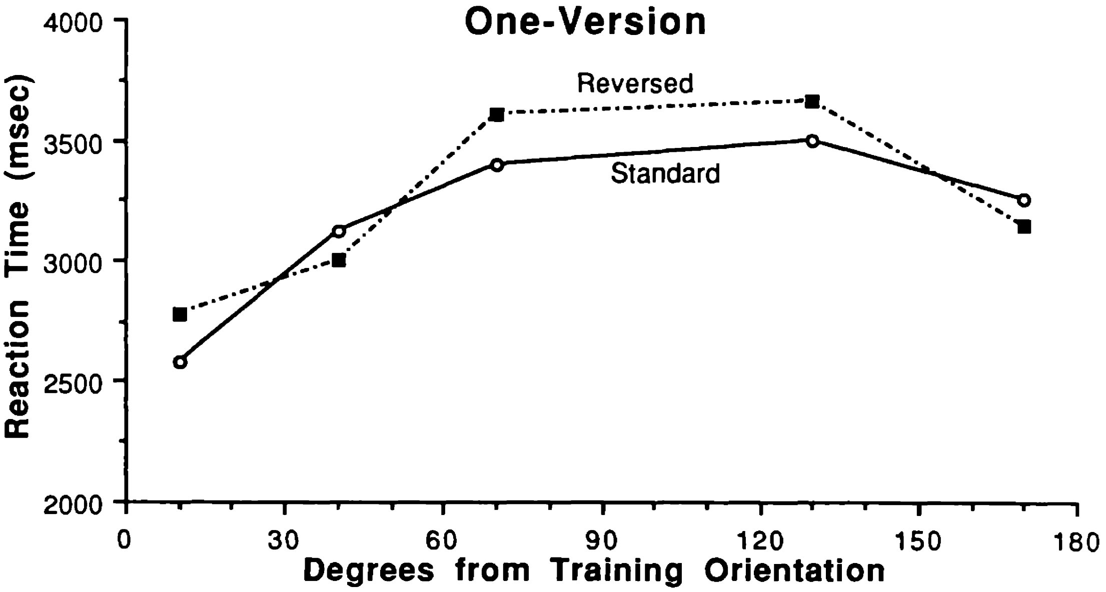

One-Version
Reversed
Standard
Reaction Time (msec)
4000
3500
3000
2500
2000
0
30
60
90
120
150
180
Degrees from Training Orientation

Figure 10

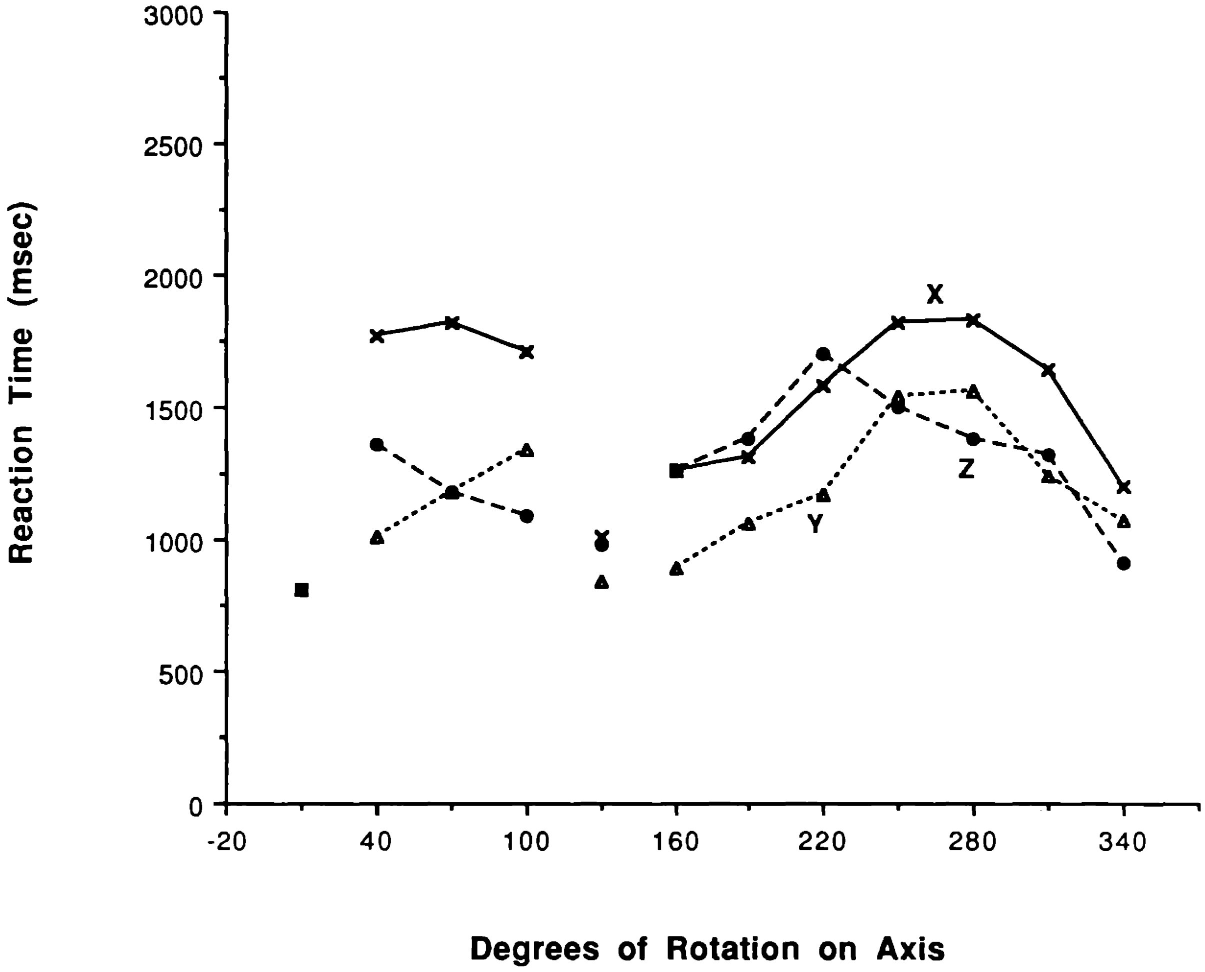

3000
2500
2000
1500
1000
500
0
-20
40
100
160
220
280
340
Reaction Time (msec)
Degrees of Rotation on Axis
X
Y
Z

Figure 11

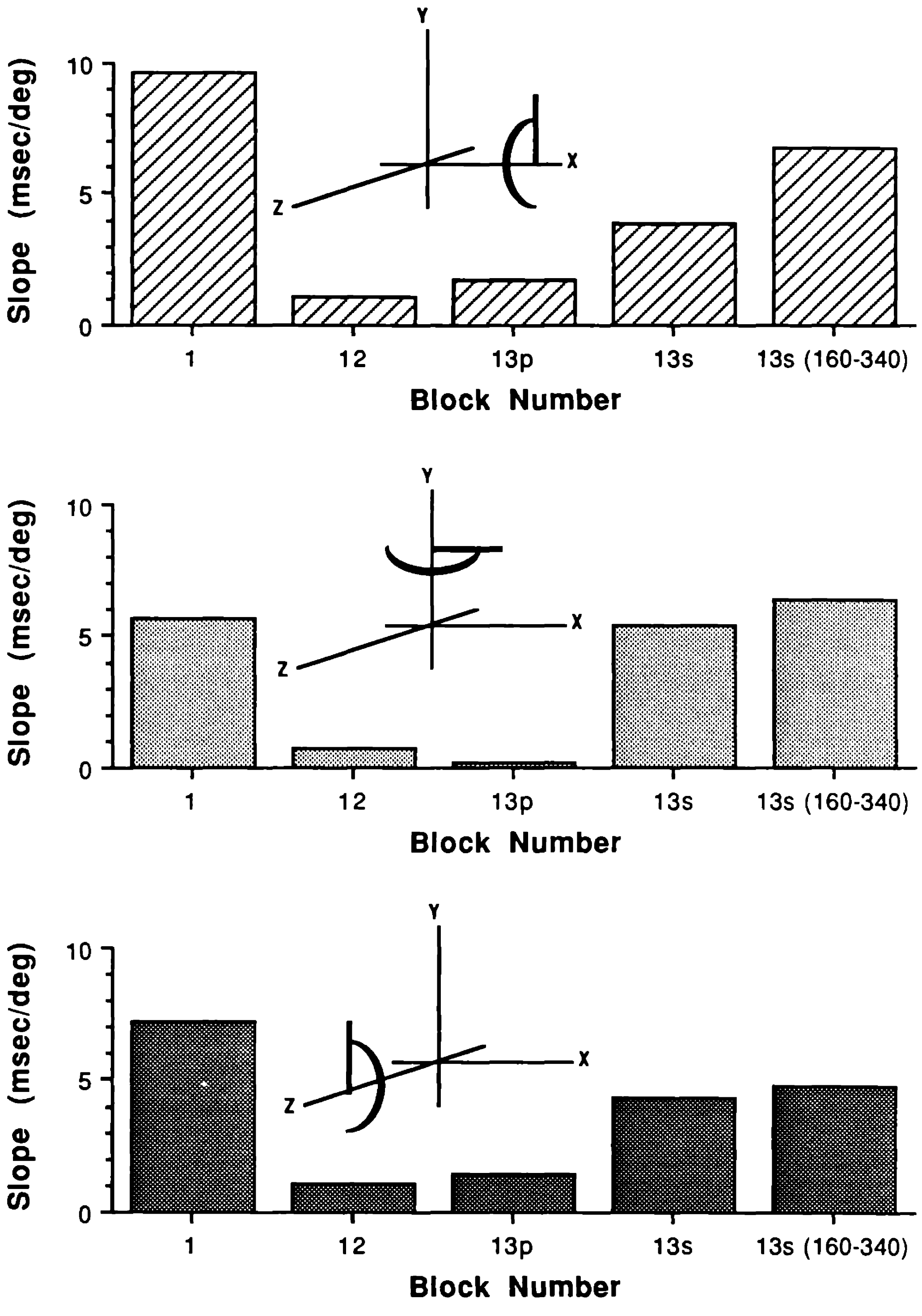

Y
X
Z
Slope (msec/deg)
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
10
5
0
1
12
13p
13s
13s (160-340)
Block Number

Figure 12

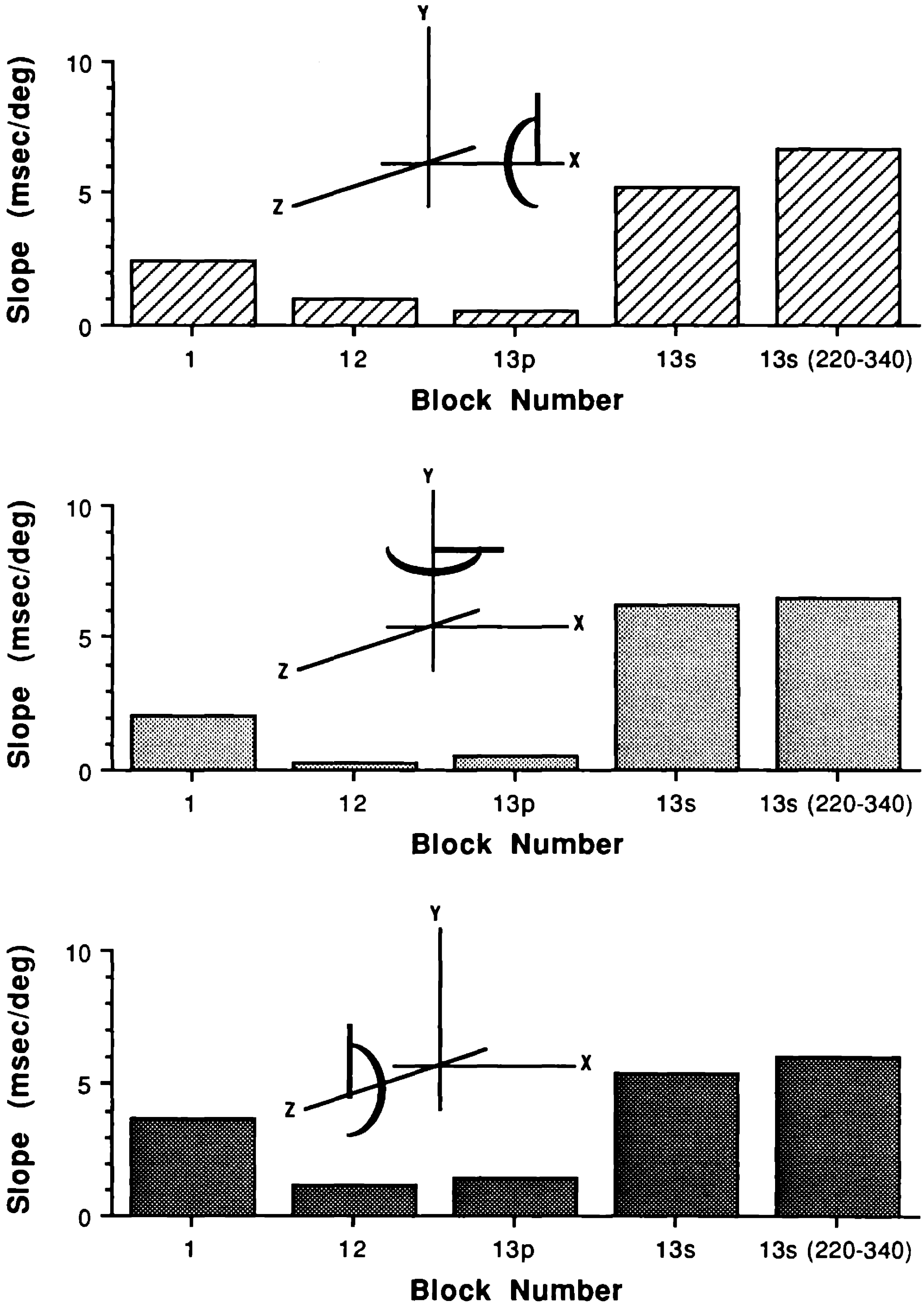

Y
X
Z
Slope (msec/deg)
10
5
0
1
12
13p
13s
13s (220-340)
Block Number
Y
X
Z
Slope (msec/deg)
10
5
0
1
12
13p
13s
13s (220-340)
Block Number
Y
X
Z
Slope (msec/deg)
10
5
0
1
12
13p
13s
13s (220-340)
Block Number

Figure 13

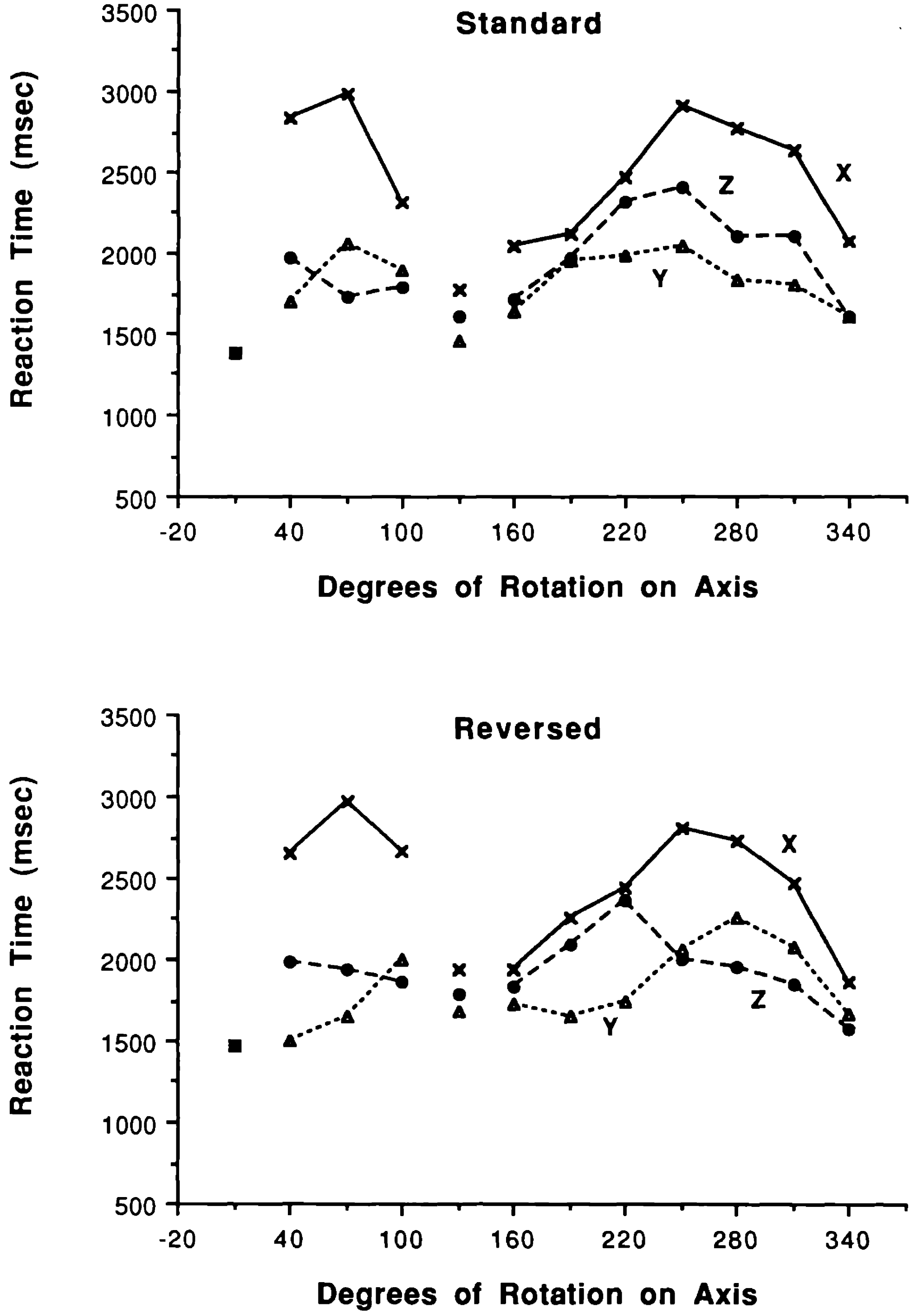

Standard
Reaction Time (msec)
3500
3000
2500
2000
1500
1000
500
-20
40
100
160
220
280
340
Degrees of Rotation on Axis
X
Z
Y
Reversed
Reaction Time (msec)
3500
3000
2500
2000
1500
1000
500
-20
40
100
160
220
280
340
Degrees of Rotation on Axis
X
Z
Y

Figure 14

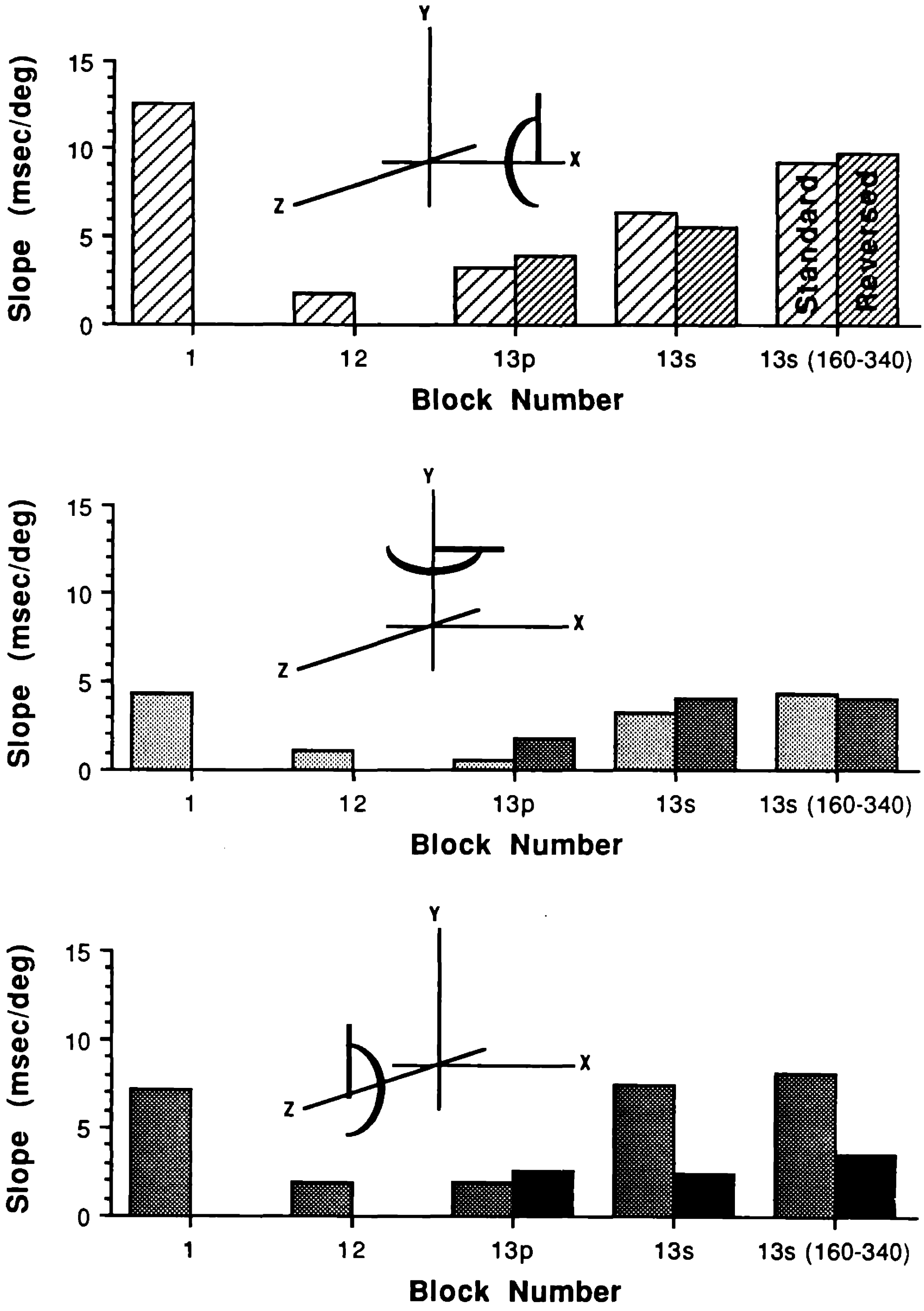

Y
X
Z
Slope (msec/deg)
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Standard
Reversed
Y
X
Z
Slope (msec/deg)
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number

Figure 15

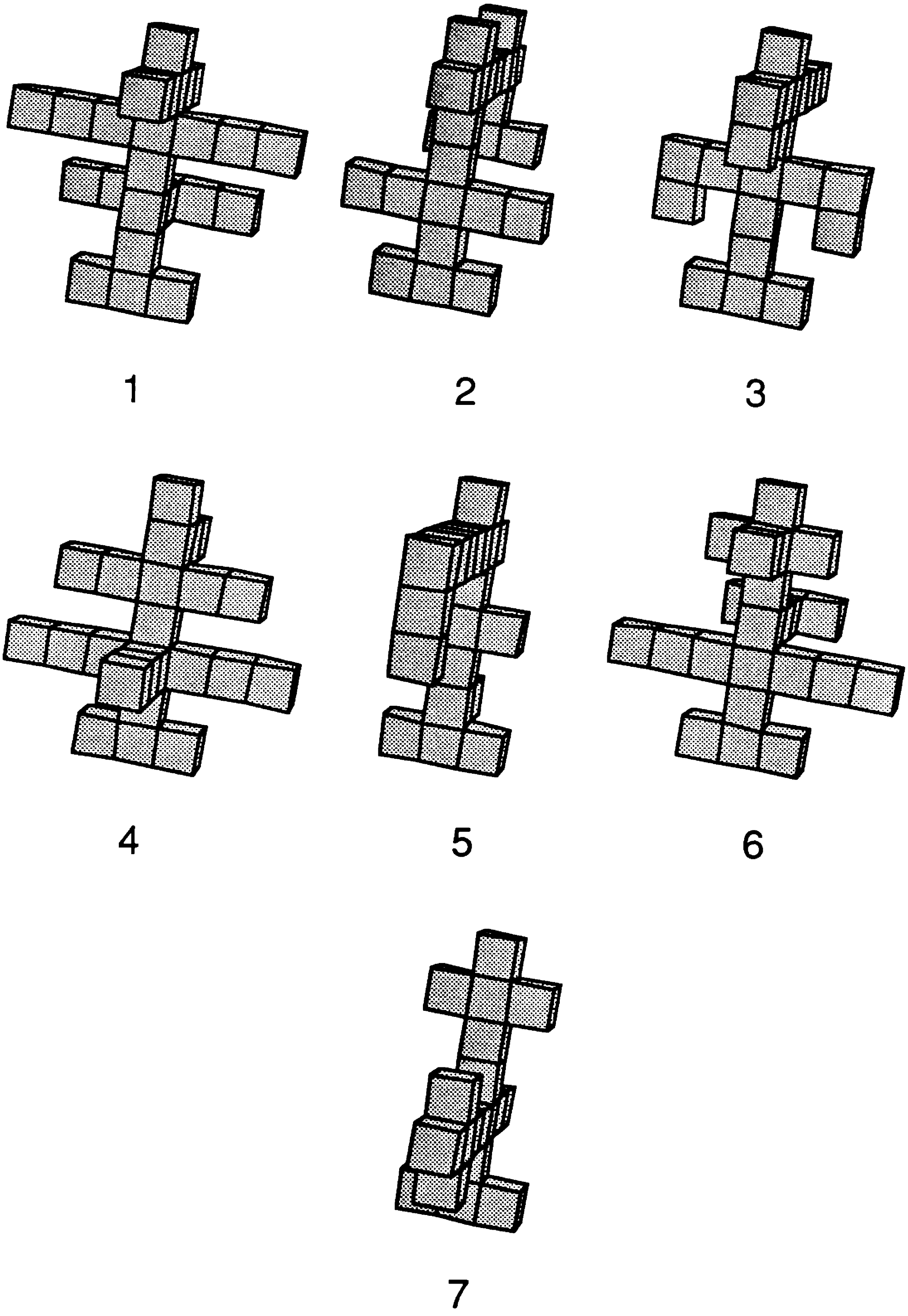
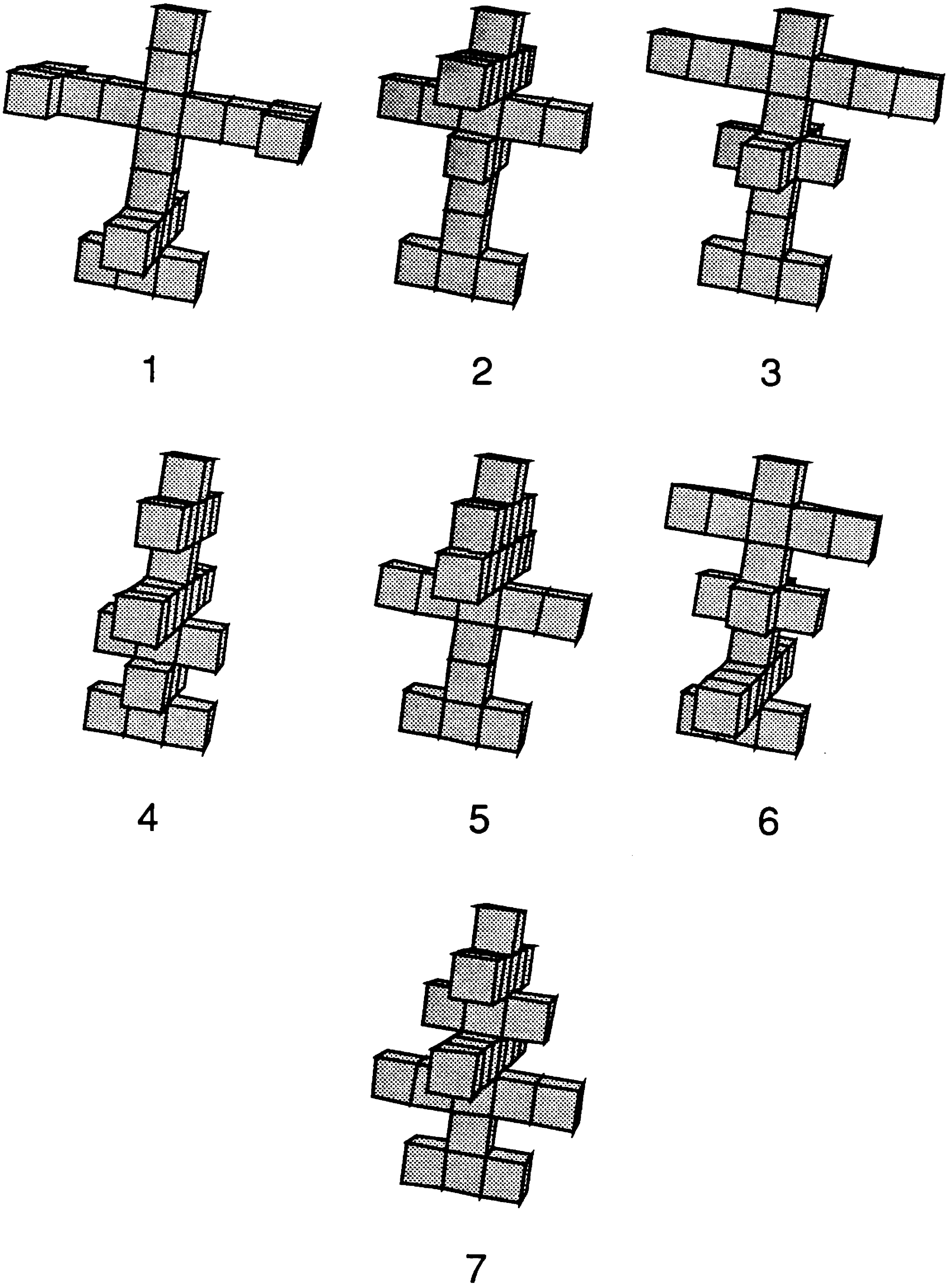


Figure 16



1 2 3

4 5 6

7

Figure 17

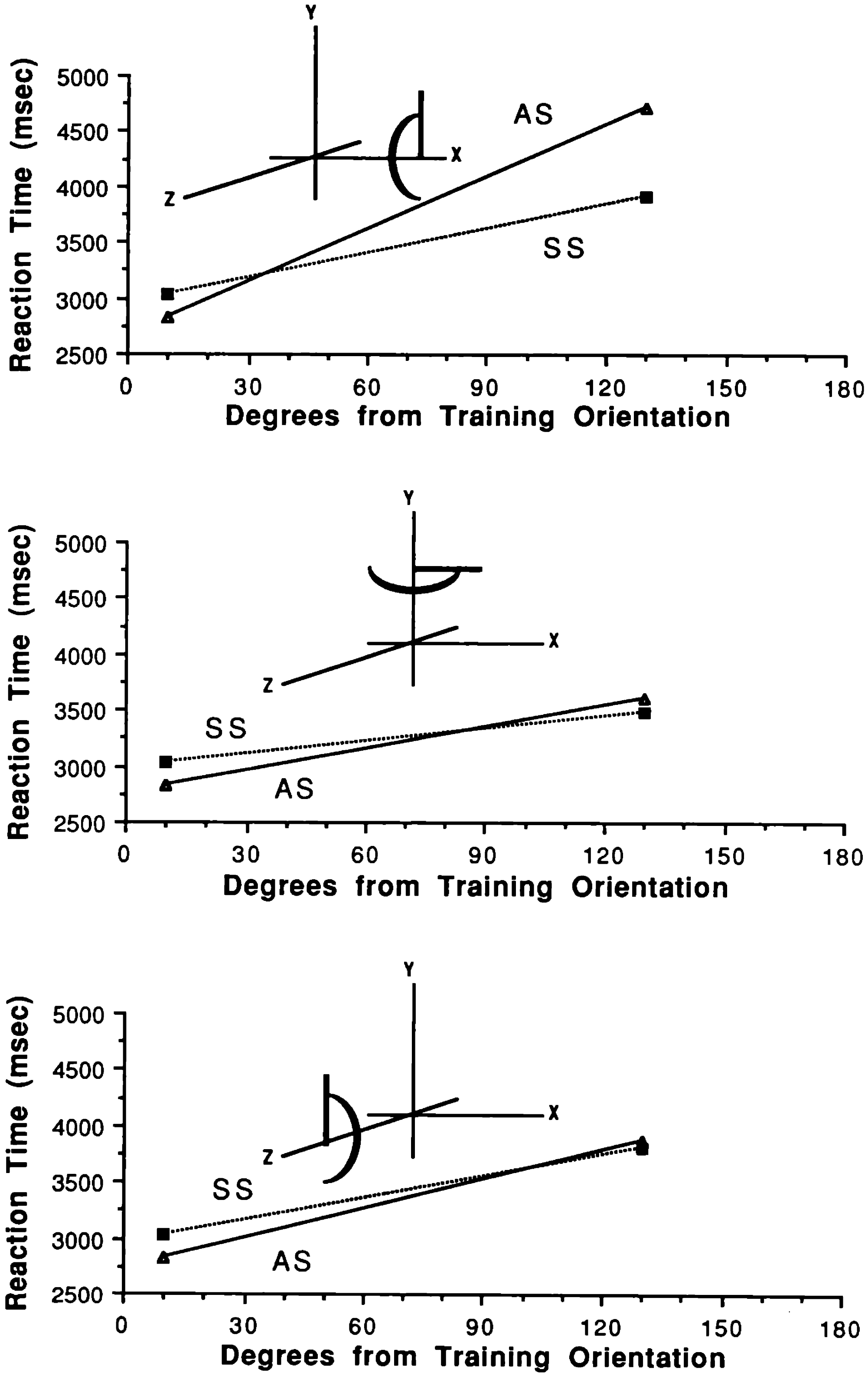

Y
X
Z
AS
SS
Reaction Time (msec)
5000
4500
4000
3500
3000
2500
0
30
60
90
120
150
180
Degrees from Training Orientation
Y
X
Z
SS
AS
Reaction Time (msec)
5000
4500
4000
3500
3000
2500
0
30
60
90
120
150
180
Degrees from Training Orientation
Y
X
Z
SS
AS
Reaction Time (msec)
5000
4500
4000
3500
3000
2500
0
30
60
90
120
150
180
Degrees from Training Orientation

Figure 18

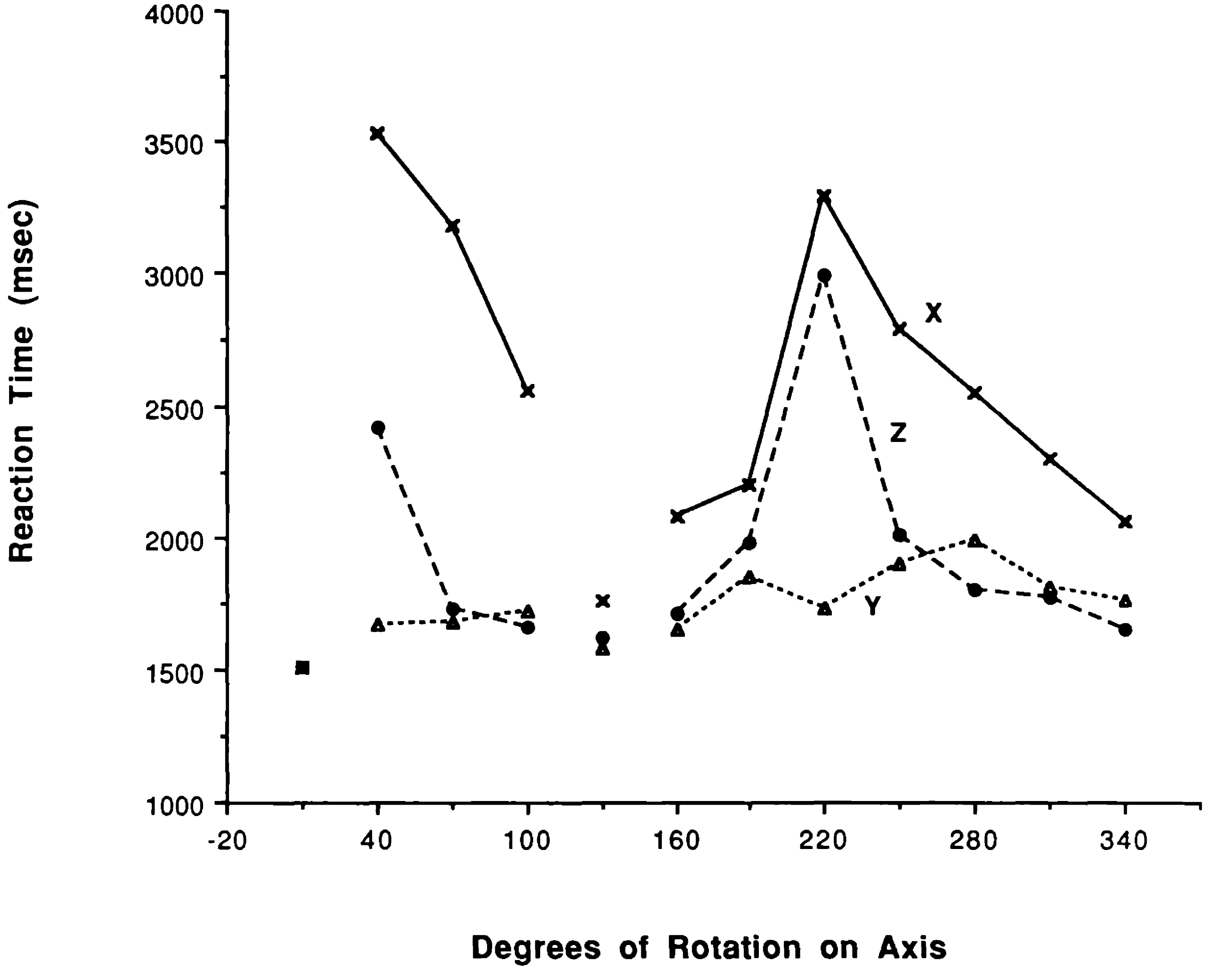

Reaction Time (msec)
4000
3500
3000
2500
2000
1500
1000
-20
40
100
160
220
280
340
X
Z
Y
Degrees of Rotation on Axis

Figure 19

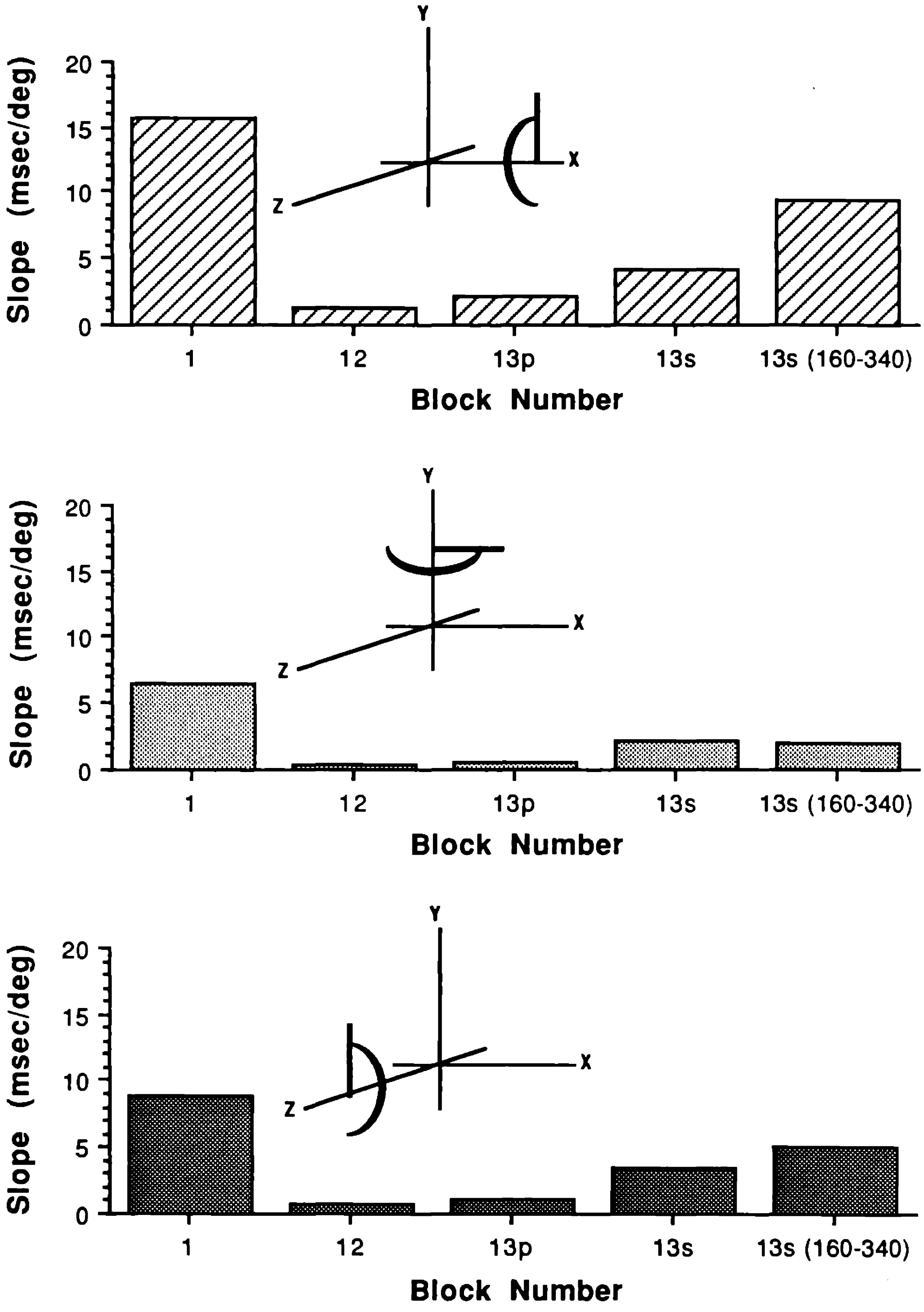

Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number

Figure 20

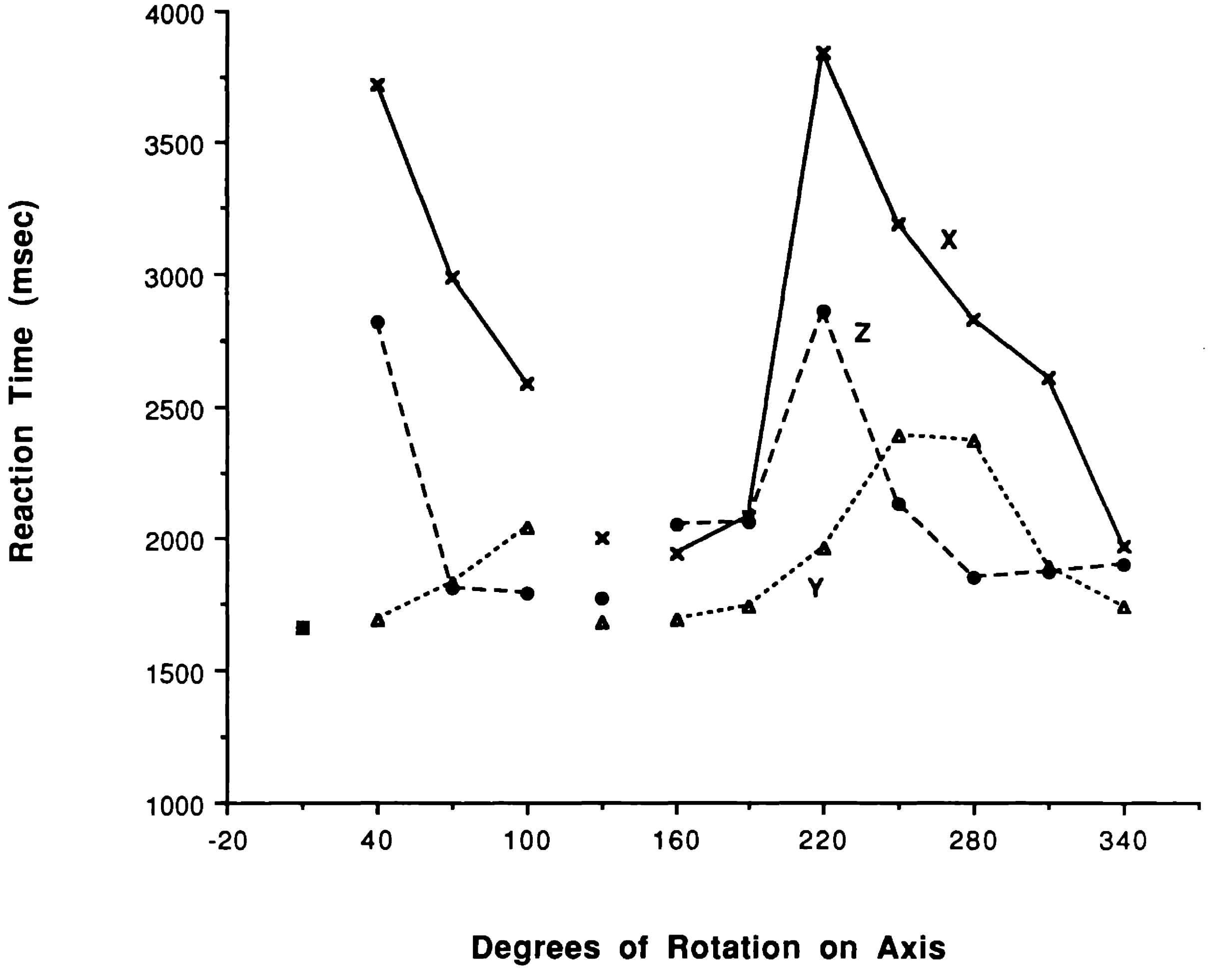

4000
3500
3000
2500
2000
1500
1000
Reaction Time (msec)
-20
40
100
160
220
280
340
Degrees of Rotation on Axis
X
Z
Y

Figure 21

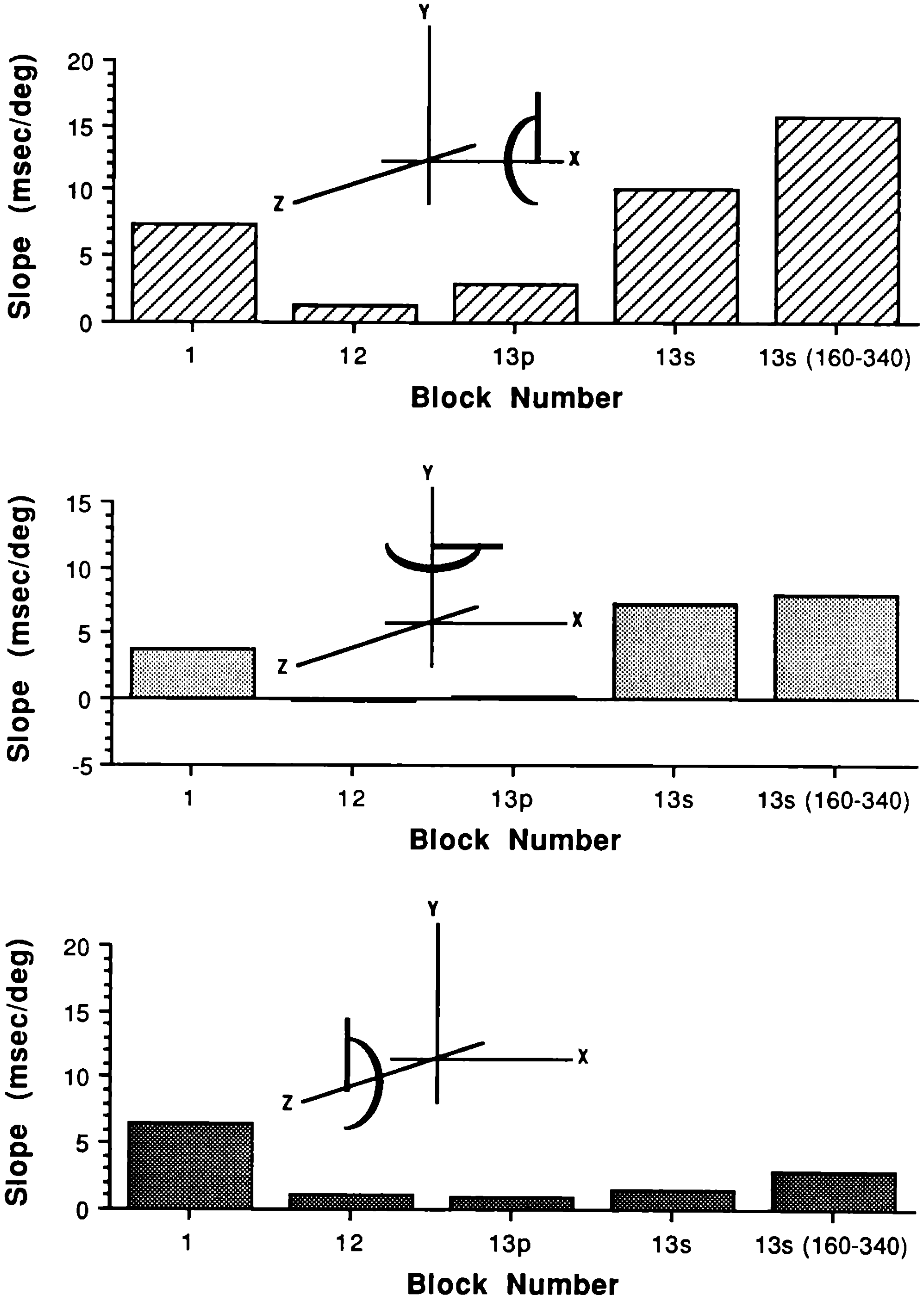
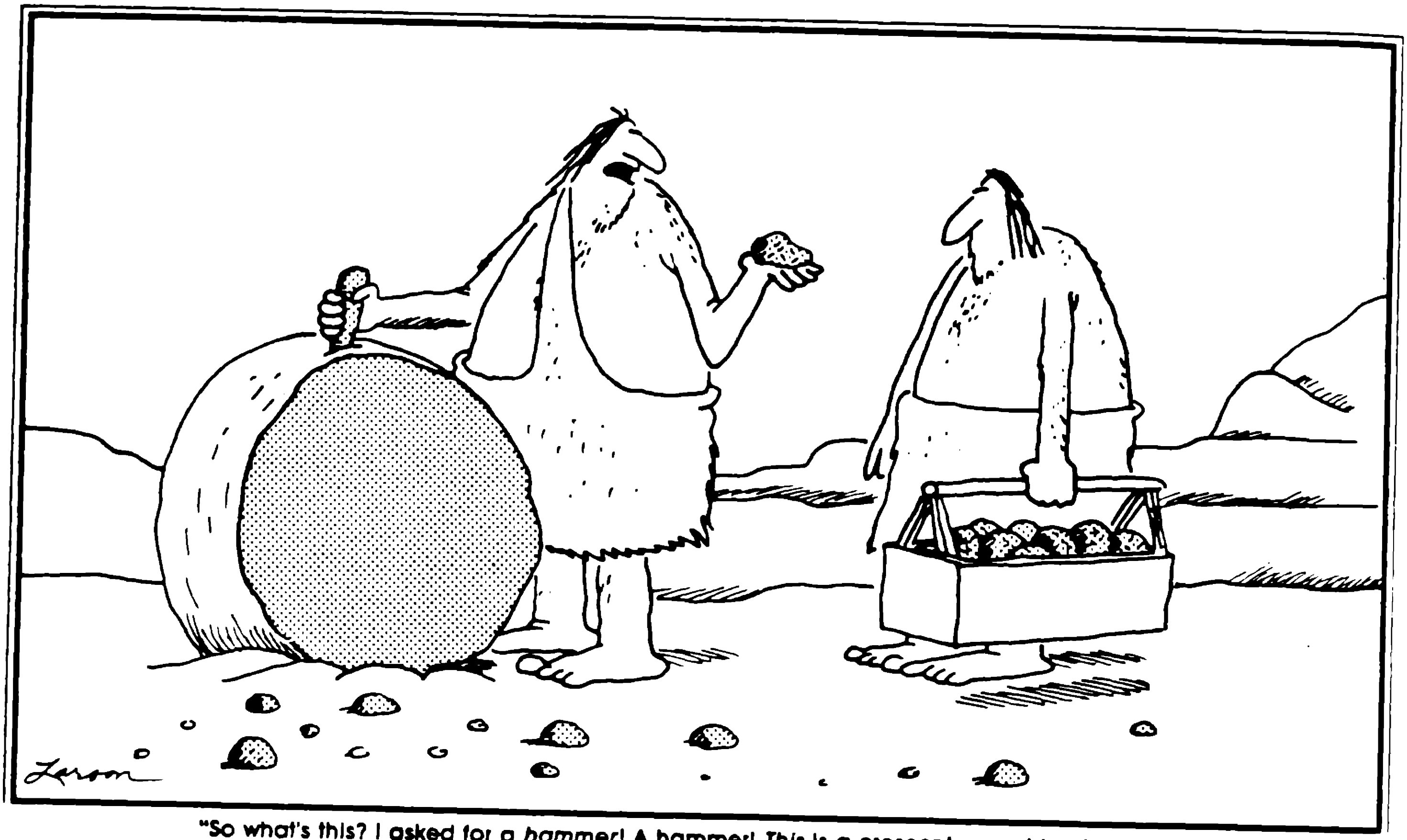

Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
15
10
5
0
-5
1
12
13p
13s
13s (160-340)
Block Number
Y
X
Z
Slope (msec/deg)
20
15
10
5
0
1
12
13p
13s
13s (160-340)
Block Number

Figure 22



"So what's this? I asked for a *hammer*! A hammer! *This* is a crescent wrench! ... Well, maybe it's a hammer. ... Damn these stone tools."

Figure 23