1-1-1999

The Influence of Spatial Reference Frames on Imagined Object- and vVewer Rotations

Maryjane Wraga
Harvard University, mwraga@smith.edu

Sarah H. Creem

Dennis R. Proffitt

Recommended Citation

The influence of spatial reference frames on imagined object- and viewer rotations

Maryjane Wraga [a,*], Sarah H. Creem [b], Dennis R. Proffitt [b]

[a] *Department of Psychology, Harvard University, 840 William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA*
[b] *University of Virginia, VA, USA*

Abstract

The human visual system can represent an object's spatial structure with respect to multiple frames of reference. It can also utilize multiple reference frames to mentally transform such representations. Recent studies have shown that performance on some mental transformations is not equivalent: Imagined object rotations tend to be more difficult than imagined viewer rotations. We reviewed several related research domains to understand this discrepancy in terms of the different reference frames associated with each imagined movement. An examination of the mental rotation literature revealed that observers' difficulties in predicting an object's rotational outcome may stem from a general deficit with imagining the cohesive rotation of the object's intrinsic frame. Such judgments are thus more reliant on supplementary information provided by other frames, such as the environmental frame. In contrast, as assessed in motor imagery and other studies, imagined rotations of the viewer's relative frame are performed cohesively and are thus mostly immune to effects of other frames. © 1999 Elsevier Science B.V. All rights reserved.

---

[*] Corresponding author. Tel.: +1-617-496-5921; fax: +1-617-496-3122; e-mail: mjwraga@wjh.harvard.edu

1. Introduction

There are multiple ways in which the human visual system can encode objects. An object can be specified relative to the observer, to the environment, to its own intrinsic structure or to other objects in the environment. Each instance requires the adoption of specific spatial frames of reference. In general, reference frames provide a structure for specifying an object's spatial composition and position.

Spatial reference frames can also be utilized in multiple ways to transform objects in the imagination. For example, if an observer wanted to construe an object at a different orientation without actually performing any actions, she could try at least two mental operations, each of which requires the rotation of a different reference frame with respect to a given stationary frame. She could either picture the object turning to its new orientation (object-relative or intrinsic reference frame) or she could imagine moving herself to the viewpoint corresponding to the new orientation (egocentric or relative reference frame). Both of these processes have been implicated in human beings' ability to update objects across different viewpoints (e.g., Hummel, 1994; Simons & Wang, 1998; Tarr, 1995; Tarr & Pinker, 1989; Wraga, Creem & Proffitt, submitted); the latter also subserves our ability to take the perspectives of others (Piaget & Inhelder, 1967).

Despite the seeming importance of both of these processes, the majority of research has focused primarily on the first type: imagined object rotations. For example, the classic studies of Shepard and colleagues established that observers mentally rotate the axes of one object into congruence with those of another object in deciding whether their shapes are similar (e.g., Cooper, 1975; Shepard & Metzler, 1971). Other studies have examined observers' ability to predict the orientational outcome of single objects rotated about multiple axes (e.g., Pani, 1993,1997; Pani & Dupree, 1994; Parsons, 1995). However, until recently, imagined rotations of the *self* have received less empirical consideration (e.g., Amorim & Stucchi, 1997; Parsons, 1987a,b; Presson, 1982).

The goal of this paper is to provide a more comprehensive account of the role of spatial reference frames in mental rotation. We review studies from several related research domains – mental rotation, object recognition, perspective-taking, and motor imagery – to examine effects of multiple reference frames on imagined transformations of the self and of objects. This approach is specifically intended to shed light on the recent finding of inferior object (versus viewer) rotation performance, as evidenced by longer reaction times and higher error rates (e.g., Amorim & Stucchi, 1997). We find that this discrepancy may be attributable to differences in the way the reference frames corresponding to each imagined rotation are transformed by the human cognitive system.

After a review of reference frames, the first main section of the paper focuses on factors affecting imagined object rotations. A recurring finding of the studies we reviewed is that imagining an object's rotation is problematic when no information other than the object's initial orientation is provided (e.g., Pani, 1993; Pani & Dupree, 1994; Parsons, 1995). This suggests a general deficit with imagining a cohesive rotation of the object's intrinsic frame. For such tasks, observers are likely to depend

on supplementary information from other frames, such as the environmental frame. As evidenced in the object recognition literature, imagined object rotations are further facilitated by view-specific encoding with respect to the relative frame (e.g., Bülthoff & Edelman, 1992; Tarr, 1995).

The second main section of the paper focuses on factors affecting imagined viewer rotations. A review of several motor imagery studies (e.g., Parsons, 1987a,b) indicates that imagined viewer rotations are less susceptible to misalignment with respect to the environmental frame, perhaps due to the inherently cohesive structure of the relative frame itself. We end with a review of research directly comparing imagined object- and viewer rotations (e.g., Presson, 1982; Wraga, Creem & Proffitt, submitted), which provides further evidence of differences in the ways the respective reference frames of each type of rotation are transformed.

2. Spatial frames of reference

We begin our review with a brief discussion of spatial reference frames. As mentioned above, imagining an object rotating to your current viewpoint or imagining yourself rotating around an object to a new viewpoint require the adoption of different reference frames. In the first section, we describe the principle frames involved in such movements. In the second section, we examine how the object and viewer frames move with respect to the environmental frame.

2.1. Rotation frames

Rotation of an object predominantly utilizes an object-relative or *intrinsic reference frame*, which is defined with respect to the object's intrinsic top/bottom, front/back, and right/left axes. Rotation of the viewer around the object predominantly utilizes an egocentric or *relative reference frame*, [1] which specifies the location of external objects with respect to the major up/down, front/back, and right/left axes of the observer's body. The egocentric frame is often further broken down to relate objects to specific parts of the body. For example, retinocentric encoding specifies an object with respect to the nodal point of the eye. Headcentric encoding specifies an object with respect to the center of the head. Bodycentric encoding specifies an object with respect to axes of individual body parts, such as the hand.

---

[1] A motion of this type may also be parsed into rotation plus *translation*. However, Shepard (1984, 1994) proposes that such motion is best described by kinematic geometry, which characterizes motions in their simplest form. In particular, Euler's theorem states that an object can move from one point in Euclidean space to another via simple rigid rotation around another (pivot) point. This pivot point can be thought of as the center of any object around which another object (i.e., the viewer) is moving.

The *environmental reference frame* [2] specifies the cardinal directions of north, south, east, and west. However, it can also be thought of more specifically as pertaining to structures and planar surfaces that are usually fixed with respect to the environment, such as the walls, floor, and ceiling of a room. Thus, the turning of a room about an observer or object would constitute an environment rotation.

2.2. Updating with respect to the environmental frame

As pointed out by Hinton and Parsons (1988), a complete description of reference systems requires consideration of their relationship to other frames. Besides the different rotation frames involved, one major difference between imagined object and viewer rotations is the way their effective reference frames change with respect to the environmental frame. In imagined object rotations, the intrinsic frame moves with respect to environment, whereas the observer's relative frame remains fixed (see Fig. 1). In imagined viewer rotations, the intrinsic frame remains fixed, and the relative frame moves with respect to the environment (see Fig. 2). The latter situation has been referred to as the "more radical change in our total experience" because it interferes with our natural inclination to be oriented with respect to the environmental frame (Shepard & Hurwitz, 1984, p. 172). The ensuing conflict between physical and imagined viewpoints has lead some researchers to assert that imagined object rotations must necessarily be easier and more natural to perform than imagined rotations of the self (Cohen et al., 1996; Shepard & Hurwitz, 1984).

We have found this claim to be unsupported in the literature. In fact, it appears that just the opposite is true. It will be shown that updating a spatial array is, in general, easier to achieve after imagined rotation of the viewer rather than of the array itself. Moreover, imagined viewer rotations are less susceptible to manipulations of the environmental frame than are imagined object rotations.

3. Factors affecting imagined object rotations

Two interdependent research domains have contributed to our understanding of the reference frames critical to imagined rotations of objects. A direct assessment of performance comes from the mental rotation literature itself, the initial goal of which was to establish the existence of analog transformations. Knowledge has also been obtained indirectly from the object recognition literature, which has utilized the mental rotation paradigm primarily to discern the nature of object representation.

---

[2] For simplicity's sake, the term "environmental frame" will be used throughout this paper to encompass the more fundamental *gravitational frame*, which determines specific "up" and "down" directions with respect to the environment (e.g., Pani & Dupree, 1994; Parsons, 1995; Shiffrar & Shepard, 1991; cf. Luyat, Ohlmann, & Barraud, 1997; Paillard, 1991).

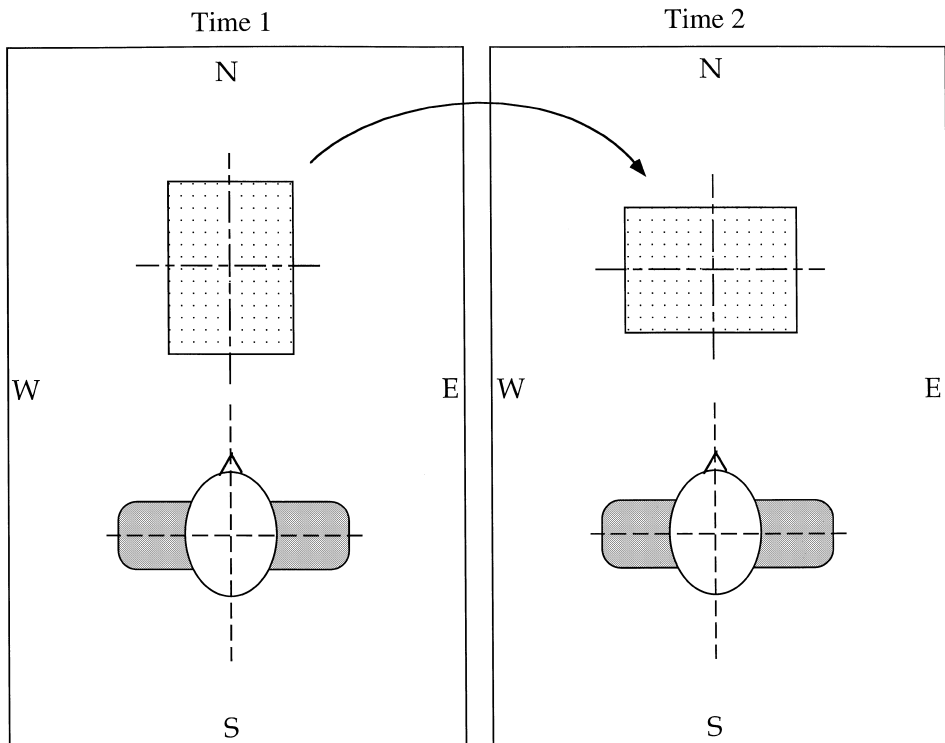Time 1   Time 2

Fig. 1. Imagined 90° clockwise rotation of an object. The object's intrinsic frame of reference changes with respect to the environmental frame from beginning (Time 1) to end (Time 2) of the rotation event. Over the same time course, the observer's egocentric frame remains fixed with respect to the environment.

3.1. Studies of mental rotation

The mental rotation of objects was initially thought to be a natural ability requiring little effort. The classic studies of Shepard and colleagues established that observers are rather good at such tasks (e.g., Cooper, 1975; Shepard & Metzler, 1971; for a more extensive collection of papers, see Shepard & Cooper, 1982). In their studies, participants were typically presented with depictions of pairs of three-dimensional novel objects, one of which was at a different orientation with respect to the other. Their task was to decide if the objects were the same or different. Response latency was found to increase linearly with the angle of displacement between the two objects. Shepard and colleagues interpreted this finding as evidence that participants had mentally rotated one object into congruence with the other. Moreover, the internal process of mental rotation was found to approximate physical constraints associated with rotations of actual objects. These findings held up over a wide range of rotation axes and degrees of rotation (Shepard & Metzler, 1971).

However, more recent studies have revealed limits to mental rotation ability (e.g., Just & Carpenter, 1985; Massironi & Luccio, 1989; Parsons, 1995). One factor that
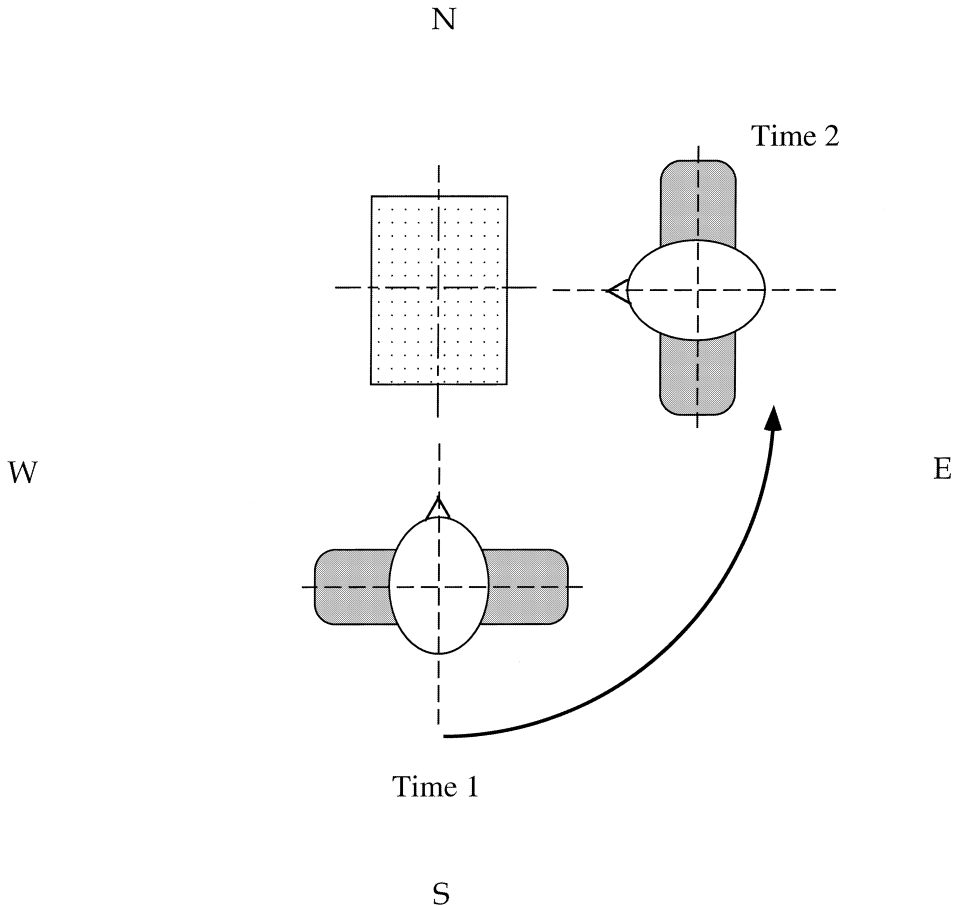
N

Time 2

W

E

Time 1

S

Fig. 2. Imagined 90° counterclockwise rotation of the observer. The viewer's relative frame of reference varies with respect to the environmental frame from beginning (Time 1) to end (Time 2) of the rotation event. Over the same time course, the object's intrinsic frame remains fixed with respect to the environment.

appears to affect performance is type of task. For example, using a novel yet complementary mental rotation task to the one previously mentioned, Parsons (1995) found dramatically impoverished results. Participants were required to predict the outcome of a Shepard–Metzler object rotated along a specific axis, through a range of angles. Once they had imagined rotating the object a given number of degrees, they were presented with a picture of the object at another orientation and had to decide whether it correctly depicted the specified rotation. Parsons found that reaction times for this predicted-outcome task were on average five times longer than those previously found with the congruence task, and participants made many more errors. Moreover, attempts to simplify the task, such as the addition of a physical axis of rotation to the objects (depicted by a rod), improved performance only slightly.

Another revealing finding of Parsons' study was that performance was significantly affected by the spatial relationship of the major reference axes. Reaction times were fastest when the axes of the object, rotation, and gravitational vertical were coincident. Performance was poorer when the object and rotation axes were not aligned with the gravitational vertical. Similar results have been reported elsewhere for mental rotations of a square (Pani, 1993), and for same-different judgments of successive *physical* rotations of cubes (Shiffrar & Shepard, 1991). Parsons concluded that poor performance resulted from participants' intrinsic inability to keep track of the object's major axis with respect to rotation and environmental frames during its imagined movement.

Pani and Dupree (1994) conducted a more systematic examination of reference frame effects. They tested the effectiveness of different reference frames on observers' ability to predict the outcome of a rotated three-dimensional square-and-rod arrangement. The arrangement consisted of a square surface whose midpoint was intersected by a rod so that the square was oriented obliquely (i.e., 45°) with respect to the rod. The arrangement was presented so that the rod was either aligned with or oblique to the gravitational frame of reference. Observer viewpoint was also manipulated so that line of gaze (i.e., retinocentric reference frame) was aligned with or oblique to the rod. Participants indicated a 180° rotation of the square by orienting another three-dimensional square in space. Similar to the Parsons study, performance in this predicted-outcome task was again much poorer compared to the speed of previous congruence tasks. Pani and Dupree also found consistent reference frame effects. The alignment of the square-and-rod with the gravitational frame provided the best conditions for accurate predictions of orientation. In contrast, manipulations of the retinocentric frame neither helped nor hindered judgments (see also Hinton & Parsons, 1988).

A general criticism of these studies might be that the test objects were too complex. However, further support for reliance on environmental referents has been obtained from studies using simpler, two-dimensional stimuli (e.g., Corballis, Nagourney, Shetzer & Stefanatos, 1978; Rock, 1973). For example, Corballis, et al. (1978) tested normal-mirror discriminations of rotated alphanumeric characters under conditions where participants' heads or bodies were either aligned with the gravitational vertical or misoriented from it up to 60°. An analysis of response latencies revealed that participants made their judgments by rotating the characters to the gravitational vertical rather than the headcentric or bodycentric reference frames. Moreover, this result was apparently not due to the overfamiliarity of the stimuli: The preference for environmental alignment also held for rotations of novel letterlike symbols.

Recent findings of McMullen and Jolicoeur (1990) indicate that type of task may determine whether environmental frames are elicited in imagined object rotations. In their study, participants performed two tasks, each of which included head upright and tilted viewing conditions. In the naming task, subjects named line drawings of objects presented at different orientations. They also made normal-mirror discriminations of the same stimuli. Judgments in the naming task were found to be more closely aligned with the egocentric frame, whereas normal-mirror judgments

replicated the environmental frame alignment of previous studies. Thus, with stimulus properties held constant, elicitation of the environmental frame was dependent on type of task. McMullen and Jolicoeur concluded that the different spatial referents involved in each task dictated which frame was utilized. Normal-mirror discriminations require interobject comparisons in external space; hence, reliance on the environmental frame. Object naming, on the other hand, is not contingent on environmental referents.

This notion is supported by a study by Rock and Heimer (1975). They initially found that participants relied mostly on the environmental frame for a recognition task of ambiguously oriented object fragments. However, when prior information of a stimulus's orientation with respect to participants' retinocentric frame was given, recognition judgments were more in line with the retinocentric frame. An implication of these findings is that the human visual system exhibits a degree of flexibility as to which reference frame is selected.

The mental rotation studies suggest several important issues. First, there are limits to our ability to mentally transform objects. Performance appears to depend on the type of task carried out. Congruence tasks are performed with ease, whereas tasks in which observers must predict the prescribed orientation of an object are more difficult. Moreover, the difficulties of predicted-outcome tasks can be compounded by the object's orientation with respect to other reference frames. Alignment with the gravitational frame appears to be a critical factor for facilitating performance. On the other hand, the relative reference frame appears to play a more substantive role in recognition tasks. This will be discussed further in Section 3.2.

3.2. Studies of object recognition

A current debate in the domain of object recognition centers on whether objects are encoded independently of observer viewpoint (e.g., Biederman & Gerhardstein, 1993, 1995; Bülthoff & Edelman, 1992; Tarr, 1995; Tarr & Bülthoff, 1995; Tarr & Pinker, 1989; see also Lawson, 1999). The mental rotation paradigm has played a primary role in this discussion. Given that object recognition involves matching images to stored representations, inferences about how such representations are encoded can be gathered by asking observers to name objects presented at multiple orientations. If response latencies exhibit the classic mental rotation function (increasing RT with degree of stimulus misorientation), viewpoint-dependent encoding is inferred (c.f. Willems & Wagemans, submitted); the absence of such a function reflects viewpoint-independent encoding. In terms of spatial referents, these approaches implicate relative and intrinsic frames, respectively.

Biederman and colleagues fall into the viewpoint-independent camp (e.g., Biederman & Cooper, 1991, 1992; Biederman & Gerhardstein, 1993). They have argued that an object is encoded as a collection of three-dimensional subcomponents, called geons, that are structurally related to one another. The resulting geon structural description (GSD) allows invariance over a range of viewpoints. The appeal of this account is that it provides a mechanism for the recognition of two similar objects containing completely different contours. However, one major

weakness of GSD theory is its limited application. According to Biederman and Gerhardstein (1993), three stringent criteria must be met in order for viewpoint invariance to be achieved: (1) a given object must be decomposable into single geons; (2) each object's GSD must be unique; and (3) an object's GSD must not appear to change over different viewpoints (e.g., via accretion and/or deletion of specific geons resulting from movement).

There is some empirical evidence in support of GSD theory. For example, Biederman and Gerhardstein (1993) presented participants with line drawings of common three-dimensional objects such as lamps and airplanes, which appeared at rotations of up to 135° with respect to the vertical axis. Participants' task was to name the objects as quickly as possible. In line with GSD theory, response latencies were found to be unaffected by stimulus orientation. Subsequent experiments meeting the GSD criteria tested unfamiliar volumetric stimuli, and also yielded RT functions that were more or less flat. Unfortunately, there are other ways to account for these results. If the studies are evaluated as straightforward tests of mental rotation (rather than object recognition), several methodological shortcomings emerge. One is that the range of rotation angles Biederman and Gerhardstein used was severely restrictive: Three out of five experiments used ranges of 90° or less (c.f. McMullen & Jolicoeur, 1990, who used a range of 300°). A second problem is that the objects were highly symmetrical and much simpler than most mental-rotation stimuli previously tested. For example, in one study (Experiment 4) participants identified simple geometric shapes such as cubes and cylinders (single volumes of geon classes) that were rotated along the y-axis in 45° increments from 0° to 90°. When evaluated from a mental-rotation perspective, it is no surprise that effects of orientation were not found: Rotations of such simple stimuli through such a restricted range of rotation would be so fast as to appear negligibly different from zero-rotation cases! Indeed, for all of the experiments in this study, RT rates were dramatically faster than those of previous mental rotation studies. Biederman and Gerhardstein interpreted the finding as evidence for viewpoint-invariant encoding, but it is certainly plausible that rapid rotation from initial viewpoint also occurred. (For additional comments, see Tarr & Bülthoff, 1995).

In fact, when more complex stimuli are tested over a large range of rotation angles and axes, there is mounting evidence suggesting that encoding occurs with respect to initial viewpoint, for both single objects (Bülthoff & Edelman, 1992; Tarr, 1995; Tarr & Bülthoff, 1995; Tarr & Pinker, 1989; Tarr, Williams, Hayward & Gauthier, 1998) and arrays of multiple objects (Diwadkar & McNamara, 1997). For example, Tarr (1995) had participants learn a series of three-dimensional Shepard–Metzler-like objects from one viewpoint. In the test phase, the objects either appeared at the same initial orientation or at new orientations, with respect to the x, y, and z axes. Tarr found that time to recognize a given object increased linearly as a function of angular disparity between its initial and tested views. When participants were then allowed to view the objects from different perspectives, later performance at these now-familiar perspectives showed few orientation effects. However, when the familiar perspectives were subsequently presented at new orientations, performance was again found to be related to the angular disparity from the nearest familiar perspective.

Viewpoint-dependency is also evidenced in a study by Rock and DiVita (1987). Participants first viewed nonsense wire figures from one perspective and then performed each of three critical test conditions. In the first, the object appeared in the same position as initial viewpoint, yielding an identical retinal image; in the second, the object appeared in a different position, yielding a different retinal image; in the third, the object appeared in a different position, but was adjusted so that its retinal image was identical to initial viewpoint. Rock and DiVita found that accuracy was highest for both conditions in which the object's retinal image was identical to initial viewpoint; performance was degraded for the condition in which the objects' retinal image varied. It would be interesting to observe whether similar dissociations would be found with RT data (which was not recorded in this study).

The object recognition studies suggest that some classes of objects are initially encoded with respect to the relative reference frame, whereas others may be encoded with respect to the intrinsic frame. Tarr and Bülthoff (1995) have proposed that viewpoint-dependence increases progressively with greater visual similarity between the objects being discriminated. This is line with the mental rotation studies in which object recognition tasks were found to elicit the relative (specifically, retinocentric) frame (e.g., McMullen & Jolicoeur, 1990). On the other hand, in line with the results of Biederman and colleagues, more visually distinct objects may be encoded with respect to the intrinsic frame.

4. Factors affecting imagined viewer rotations

Several research domains have contributed to our understanding of performance in imagined self-movement. Early developmental work focused on the problem of whether young children could take the perspective of others. More recent work in motor imagery has emphasized whether imagined movements of the self are analogous to corresponding physical movements. Finally, research comparing performance of multiple imagined rotations has focused on whether the human visual system treats all imagined transformations similarly.

4.1. Studies of perspective taking

In contrast to the literature on mental transformations of objects, early studies on mental transformations of the self suggested that observers generally have difficulties imagining perspectives other than their own. This work was pioneered by Piaget (Piaget & Inhelder, 1967). He used the "three mountains" perspective problem to demonstrate that children have trouble imagining how a group of mountains would look from another's viewpoint, a finding that has been widely replicated (e.g., Flavell, 1968; Laurendeau & Pinard, 1970; c.f. Newcombe & Huttenlocher, 1992).

Similar results have also been found with adults. Rock, Wheeler and Tudor (1989) presented observers with novel twisted-wire objects and asked them to predict what the objects would look like from a perspective shifted 90° from their own. Participants exhibited difficulties in updating the objects from the new perspective over a

range of response measures. However, they also had equal difficulty predicting the outcome of a 90° rotation of the objects themselves. Thus, it is unclear as to whether participants' problems stemmed from a lack of perspective-taking ability or from a more general inability with representing these stimuli. A subsequent experiment by Farah, Rochlin and Klein (1994) revealed that the addition of surfaces to the wire forms facilitated object rotation performance. Unfortunately, viewer rotations were not tested with these stimuli.

4.2. Studies of motor imagery

Another paradigm for studying imagined viewer transformations has produced more definitive results. This approach examines how well observers can determine whether depictions of human body parts (such as hands and feet) belong to the right or left side of a human body (Ashton, McFarland, Walsh & White, 1978; Cooper & Shepard, 1975; Parsons, 1987a,b, 1994; Sekiyama, 1982). For example, Parsons (1987a) presented observers with drawings of right and left hands and feet, in various orientations with respect to the vertical. He found that participants judged quite easily whether the body parts were right or left by performing egocentric transformations: They imagined their own hands (and feet) rotating into the orientation of each corresponding stimulus, rather than imagining the stimulus rotating to the vertical. Moreover, the starting orientations of internally represented body parts were not arbitrary. Response latencies for most depicted stimulus orientations corresponded to the angular displacement between the physical orientation of the participant's relevant body part and the orientation of the stimulus. [3] Orientation effects were also found, but these generally corresponded to the degree of awkwardness of movement implied by the oriented stimulus. Participants responded faster and with fewer errors for imagined rotations corresponding to natural movements than for those corresponding to awkward or uncomfortable movements.

Similar results were obtained for judgments involving full-body rotations (Parsons, 1987b). Participants viewed drawings of human bodies with one arm (right or left) outstretched, at various orientations and axes, and made right/left arm judgments as above. Participants had little trouble solving the task by imagining rotating their bodies into the misoriented stimuli. Response time to some depictions took longer, such as those of bodies appearing upside-down or with heads facing away from the observer. However, these effects were much smaller than the gravitational misalignment effects reported in object-rotation studies (e.g., Pani, 1993; Pani & Dupree, 1994; Parsons, 1995). Interestingly, a manipulation designed to make the environmental frame more salient did not affect judgments. The addition of a ground plane into the scene did not inhibit participants' ability to make judgments for depictions of body orientations violating physical laws (e.g., body perpendicular to

---

[3] This was confirmed in a later study where manipulations of the position of participants' hands affected response latencies of the same hand stimuli in predictable ways (Parsons, 1994).

gravitational vertical, appearing to balance on the ground by outstretched right arm).

Recent neuroimaging studies confirm that these types of tasks require egocentric encoding. Kosslyn, DiGirolamo, Thompson and Alpert (1998) conducted a positron emission tomography (PET) study that compared participants' performance on same-different discriminations of objects and body parts. Right-handed participants viewed depictions of either pairs of Shepard–Metzler objects or pairs of hands. In each pair, one stimulus was at a different orientation with respect to the other. Kosslyn et al. found that imagined rotations of hands activated primary motor areas of the brain associated with executing hand movements, whereas imagined rotations of the objects did not (see also Parsons et al., 1995). Moreover, the activated areas for the hand task were lateralized to the left hemisphere only, which suggests that participants had mentally rotated their dominant right hands to solve the task.

These experiments collectively suggest that, similar to intrinsic object rotations, performance on egocentric viewer rotation tasks may be task-dependent. However, egocentric rotations appear to be governed by different factors than object rotations. Imagined rotations of the self adhere to the physiological and kinematic constraints of corresponding physical actions rather than constraints of external space. Movements that are awkward to perform take longer to imagine. Direct manipulations of the observers' own egocentric frame also affect performance. In contrast, manipulations with respect to the environmental frame produce somewhat weaker effects.

4.3. Studies comparing object and viewer rotations

Another approach to determining how reference frames affect mental transformations is to contrast performance of tasks in which different reference frames are known to be activated, such as object versus viewer rotations. Direct comparisons between these two types of rotation have yielded mixed results (Amorim & Stucchi, 1997; Huttenlocher & Presson, 1979; Presson, 1982). Amorim and Stucchi (1997) found a viewer advantage in a task where participants imagined standing at the periphery of a large clock that was lying on the horizontal plane. In the center of the clock they imagined a three-dimensional block letter. For the viewer task, participants imagined moving themselves to a given location around the clock, and updated the orientation of the letter by indicating what number it pointed to on the clock from the new perspective. In the object task, they imagined pointing the letter to a given location on the clock, and then updated their own position on the clock with respect to the new letter orientation. Participants performed better in the viewer task than the object task, as evidenced by faster response times and fewer errors. A similar viewer advantage has been found for updating after physical movement (Simons & Wang, 1998).

Presson (1982) found relatively poor imagined-viewer performance in a study involving transformations of arrays of objects. In one experiment, participants were shown an array of blocks and were asked to construct a new array from a perspective corresponding to either a rotation of themselves around the array or a rotation of the array itself. Presson reported an advantage for the imagined array rotations, but

noted that participants solved the task by transforming objects piecemeal, rather than performing holistic rotations of the array. When the dependent measure was changed to one of selecting the correct depiction of the array, the array-rotation advantage disappeared (Presson, 1982, p. 244). In two additional experiments, participants answered questions about the location of objects in a miniature array after imagining either the array rotating (Array task) or themselves rotating around the array (Viewer task). Presson found that performance of either task depended on the type of question asked. When participants were asked to name an object that would be present at a given location in the array after performing either type of imagined rotation ("item" questions, e.g., "If you/the array were rotated 90°, what object would be on the right?"), an advantage for imagined viewer rotations was found. However, when the task involved stating the location of a named object in the array after rotation ("position" questions, e.g., "If you/the array were rotated 90°, where would the drum be?"), the opposite result was found: Imagined array rotations were superior to viewer rotations. Similar array advantages using position questions have been reported elsewhere (Hardwick, McIntyre & Pick, 1976; Huttenlocher & Presson, 1979).

Wraga, Creem and Proffitt (submitted) proposed that the array advantage for position questions was likely spurious, due to the fact that participants could resort to strategies other than holistic rotations in the Array task. For example, in solving the above drum question, participants could simply imagine moving a single object – the drum – while ignoring the other items in the array. The imagined displacement of a single object could be achieved via translation rather than rotation, much like a single car of a Ferris wheel traverses a circular path without rotating. In contrast, the Viewer task required rotating the prescribed number of degrees and updating the entire array to find the drum. Imagined translations are generally thought to be easier to perform than imagined rotations (Easton & Sholl, 1995; Rieser, 1989).

Wraga et al. conducted a variation of Presson's position question experiment to test this hypothesis (Wraga, Creem and Proffitt, submitted, Experiment 3). Participants performed imagined Array and Viewer rotations to position questions. In addition, "catch trials" were added to both tasks to determine whether participants were performing rotations. For each catch trial, participants updated the position of any object in the array relative to the perspective arrived at from a given position question. For example, if the position question were "Rotate 90°, where is the drum?", the catch trial would be "Now, what's on your right?" The results revealed that performance on the position questions was relatively easy in the Array task, but performance became much poorer on catch trials, indicating that participants had not holistically rotated the array initially during the position question. In contrast, performance in the Viewer task was consistent across both question types. Thus, it is likely that all experiments showing an array advantage using position questions have inadvertently tapped performance of imagined object translations rather than rotations.

Wraga et al. went on to clarify the conditions under which imagined object- and viewer rotations are advantageous. In one experiment (Experiment 4), the array was collapsed into one object: a rectangular block whose four edges were each painted a

different color. In this way, the four components to be learned (i.e., colors) were spatially adjoined and thus encodeable as a single unit, much like the four axes of the body utilized in the Viewer task. Participants either imagined the block rotating around its own axis (Object task) or imagined themselves rotating around the block (Viewer task). In another experiment (Experiment 5), an object with a more familiar configuration (toy car) was used. The car's components (hood, driver's side, passenger's side, trunk) were thus both spatially connected and highly familiar. Despite either manipulation, participants responded more quickly and made fewer errors in the Viewer task than the Object task. However, performance in the Car-Object task improved over the Block-Object task, whereas Viewer performance was equivalent across both. Wraga et al. attributed the improved performance to the fact that participants were more successful in performing cohesive transformations of the car's components.

A subsequent experiment (Experiment 6) attempted to better facilitate object rotations, using a somewhat different manipulation. Recent studies have shown that the ability to localize objects after imagined viewer rotations improves if the observer is physically rotated while the imagined transformation is being performed (Farrell & Robertson, 1998; Presson & Montello, 1994; Rieser, 1989; Rieser, Guth & Hill, 1986). Wraga, Creem and Proffitt extended this manipulation to passive rotations of the *object*. Participants again performed both Object and Viewer tasks. However, for each rotation trial of the Object task, the experimenter rotated the block in the blindfolded participants' upturned palm while the corresponding degrees were given verbally. This manipulation led to a significant decrease in response latency and errors, comparable to Viewer levels. An additional condition with a different group of participants in which the block remained stationary in participants' upturned palm resulted in the typical Viewer advantage. Wraga, Creem and Proffitt concluded that the turning block had provided participants with haptic information specifying the object's orientation with respect to the environment or themselves at every phase of the transformation, which enhanced their visual imagery of a rigid, cohesive transformation.

5. Conclusions

The goal of this paper was to examine the spatial reference frames relevant to imagined object and viewer rotations, in order to understand the relatively longer RTs and higher error rates found with object performance in recent studies. A critical factor to this result may be the type of task utilized in rotation comparison studies. Most studies have used tasks in which observers predict a rotational outcome, either of themselves or of an object (e.g., Presson, 1982; Wraga, Creem and Proffitt, submitted). As reviewed in this paper, a common finding of the mental rotation literature was that these types of predicted-rotation tasks are difficult to perform on objects (e.g., Pani, 1993; Pani & Dupree, 1994; Parsons, 1995). In general, the accuracy of such judgments was found to be heavily dependent on

alignment of the object with respect to the environmental frame. When the object was misaligned, performance deteriorated. Moreover, poor performance was not due to memory load: In some of the experiments (e.g., Pani & Dupree, 1994), the object to be rotated was physically present during the task. On the other hand, recognition tasks such as normal-mirror discriminations, same-different judgments, and object classifications were found to be relatively easier.

What might account for this marked discrepancy in object-rotation performance? The answer may lie in the representational requirements of the two respective task types. Recognition tasks require a match between two existing representations, or in the case of sequential stimulus presentation, a match of an existing representation to a stored one. Thus, the beginning and end positions of a given rotation event are predetermined and nonarbitrary. To solve such a task, all of an object's components are collectively rotated into congruence with the end-position representation. A decision is then made as to whether all of the components match up to those of the end-position representation. In contrast, a predicted-outcome task requires the building of a new representation from an existing one: Its rotation event has an indeterminate end position. Difficulties in this task might thus stem from inherent problems with performing cohesive rotations of all components of the intrinsic representation. Indeed, several studies have indicated that, when given the opportunity, observers opt to transform single components of an object rather than complete a holistic rotation (Presson, 1982; Wraga, Creem and Proffitt, submitted). Moreover, Wraga, Creem and Proffitt's block-turning experiment demonstrated that such performance could be improved when on-line haptic information for an object's rigid, cohesive rotation was given.

In the context of this difficulty, the evidence for viewpoint-specific encoding reported in the object-recognition literature makes sense (e.g., Bülthoff & Edelman, 1992; Tarr, 1995; Tarr & Pinker, 1989). Perhaps the nonarbitrary starting points that result from viewpoint-specific encoding help facilitate rotation of the intrinsic frame.

Object difficulties aside, it is of equal importance to ascertain why predicted-outcome tasks of imagined viewer rotations are found to be less problematic. The answer may be revealed in the structure of the relative reference frame itself. The front-back and right-left axes of the relative frame "belong" to the observer. When the observer moves, the entire relative frame moves with her: It is biologically impossible to move the relative frame in a piecemeal fashion. In contrast, most objects can be easily separated into parts. An evolutionary argument can also be made for this difference in abilities. We have evolved as moving organisms in a relatively stable environment, in which objects themselves rarely rotate. Thus, updating the world with respect to the relative frame may be a more natural ability (for similar comments, see also Farrell & Robertson, 1998; Simons & Wang, 1998; Wraga, Creem and Proffitt, submitted).

To summarize, the discrepancy between performance of imagined object and viewer rotations may be attributable to differences in the way their corresponding reference frames are transformed by the human cognitive system. This review showed that difficulties in performing cohesive rotations of an object's intrinsic frame result in reliance on information from other frames, such as the environmental

frame. In contrast, the stability of the observer's relative frame renders its rotations to be less susceptible to multiple-frame influences.

Acknowledgements

References

Amorim, M., & Stucchi, N. (1997). Viewer- and object-centered mental explorations of an imagined environment are not equivalent. *Cognitive Brain Research*, *5*, 229–239.

Ashton, R., McFarland, K., Walsh, F., & White, K. (1978). Imagery ability and the identification of hands: A chronometric analysis. *Acta Psychologica*, *42*, 253–262.

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual priming. *Perception*, *20*, 585–593.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 121–133.

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

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1506–1514.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, *89*, 60–64.

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., Bookheimer, S. Y., Rosen, B. R., & Belliveau, J. W. (1996). Changes in cortical activity during mental rotation: A mapping study using functional MRI. *Brain*, *119*, 89–100.

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

Cooper, L. A., & Shepard, R. N. (1975). Mental transformations in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance*, *104*, 48–56.

Corballis, M. C., Nagourney, B. A., Shetzer, L. I., & Stefanatos, G. (1978). Mental rotation under head tilt: Factors influencing the location of the subjective reference frame. *Perception & Psychophysics*, *24*, 263–273.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, *8*, 302–307.

Easton, R. D., & Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 483–500.

Farah, M. J., Rochlin, R., & Klein, K. L. (1994). Orientation invariance and geometric primitives in shape recognition. *Cognitive Science*, *18*, 325–344.

Farrell, M. J., & Robertson, I. H. (1998). Mental rotation and the automatic updating of body-centered spatial relationships. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 227–233.

Flavell, J. (1968). *The development of role-taking and communication skills in children*. New York: Wiley.

Hardwick, D., McIntyre, C., & Pick, H. (1976). The content and manipulation of cognitive maps in children and adults. *Monographs of the society for research in child development*, 41.

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

Hummel, J. E. (1994). Reference frames and relations in computational models of object recognition. *Current Directions in Psychological Science*, *3*, 111–116.

Huttenlocher, J., & Presson, C. C. (1979). The coding and transformation of spatial information. *Cognitive Psychology*, *11*, 375–394.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*, 137–172.

Kosslyn, S. M., DiGirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. *Psychophysiology*, *35*, 151–161.

Laurendeau, M., Pinard, A. (1970). *The development of the concept of space in the child*. New York: International University Press.

Lawsson (1999). Achieving visual object containing across flame rotation and depth rotation. *Acta Psychologica* 102, 221–245.

Luyat, M., Ohlmann, T., & Barraud, P. A. (1997). Subjective vertical and postural activity. *Acta Psychologica*, *95*, 181–193.

Massironi, M., & Luccio, R. (1989). Organizational versus geometric factors in mental rotation and folding tasks. *Perception*, *18*, 321–332.

McMullen, P. A., & Jolicoeur, P. (1990). The spatial frame of reference in object naming and discrimination of left-right reflections. *Memory & Cognition*, *18*, 99–115.

Newcombe, N., & Huttenlocher, J. (1992). Children's early ability to solve perspective-taking problems. *Developmental Psychology*, *28*, 635–643.

Paillard, L. (1991). Motor and representational framing of space. In J. Paillard (Ed.), *Brain and space* (pp. 163–182). Oxford: Oxford University Press.

Pani, J. R. (1993). Limits on the comprehension of rotational motion: Mental imagery of rotations with oblique components. *Perception*, *22*, 785–808.

Pani, J. R. (1997). Descriptions of orientation in physical reasoning. *Current Directions in Psychological Science*, *6*, 121–126.

Pani, J. R., & Dupree, D. (1994). Spatial reference systems in the comprehension of rotational motion. *Perception*, *23*, 929–946.

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

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

Parsons, L. M. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 709–730.

Parsons, L. M. (1995). Inability to reason about an object's orientation using an axis and angle of rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1259–1277.

Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., Jerabek, P. A., & Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature*, *375*, 54–58.

Piaget, J., & Inhelder, B. (1967). *The child's conception of space*, (pp. 209–246). New York: Norton (Original work published 1948).

Presson, C. C. (1982). Strategies in spatial reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 243–251.

Presson, C. C., & Montello, D. R. (1994). Updating after rotation and translational body movements: Coordinate structure of perspective space. *Perception*, *23*, 1447–1455.

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1157–1165.

Rieser, J. J., Guth, D. A., & Hill, E. W. (1986). Sensitivity to perspective structure while walking without vision. *Perception*, *15*, 173–188.

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

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, *19*, 280–293.

Rock, I., & Heimer, W. (1975). The effect of retinal and phenomenal orientation on the perception of form. *American Journal of Psychology*, *70*, 493–511.

Rock, I., Wheeler, D., & Tudor, L. (1989). Can we imagine how objects look from other viewpoints?. *Cognitive Psychology*, *21*, 185–210.

Sekiyama, H. (1982). Kinesthetic aspects of mental representation in the identification of left and right hands. *Perception & Psychophysics*, *32*, 89–95.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, *91*, 417–447.

Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review*, *1*, 2–28.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

Shepard, R. N., & Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition*, *18*, 161–193.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Shiffrar, M. M., & Shepard, R. N. (1991). Comparison of cube rotations around axes inclined relative to the environment or to the cube. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 44–54.

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, *9*, 315–320.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, *2*, 55–82.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1494–1505.

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

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint-dependent. *Nature Neuroscience*, *1*, 275–277.

Willems, B., & Wagemans, J. (submitted). *Matching multi-component objects from different viewpoints: Normalization but not mental rotation.*

Wraga, M., Creem, S. H., & Proffitt, D. R. (submitted). *Updating scenes after imagined object- and viewer-rotations.*