Journal of Experimental Psychology:
Human Perception and Performance
2001, Vol. 27, No. 5, 1090–1115

0096-1523/01/$5.00 DOI: 10.1037//0096-1523.27.5.1090

Matching Multicomponent Objects From Different Viewpoints: Mental Rotation as Normalization?

Bert Willems and Johan Wagemans
University of Leuven

Novel multicomponent objects were created, and 3 distractors were created for each object by changing the relations between the parts of the object. In a set of 5 experiments, target objects were presented as a motion sequence of multiple views or as a single view. Participants were asked to determine whether an image of an object, viewed from another viewpoint, was the same as the target object. The axis of rotation was aligned with one of the environmental axes or with the main axis of the object. The effects of viewpoint on performance imply that the matching of objects is viewpoint dependent and requires a process of normalization to undo the differences between the perceptual description and the stored object descriptions. The lack of a systematic effect of the axis of rotation, however, suggests that this normalization is best understood as not involving a 3-D transformation of stored 3-D object models.

Several theories have been formulated to explain how humans recognize objects from different viewpoints. Despite considerable differences between these theories, they all start from the same underlying idea: For recognition or identification of an object, a match has to be established between the specific retinal stimulation of the object when one is looking at it and a more abstract representation of the object stored in memory. Theories differ from one another based on their assumptions about the nature of the perceptual description computed from the retinal stimulation before this perceptual description is matched, about the nature of the stored object description, and about the nature of the matching process (see Table 1 for an overview).

Viewpoint-Dependent Versus Viewpoint-Independent Theories of Object Recognition

The dimension used most frequently to distinguish among theories of object recognition is whether the perceptual descriptions are viewpoint specific or viewpoint invariant. Viewpoint-independent or viewpoint-invariant theories of object recognition, historically the oldest group of theories, claim that both the perceptual descriptions and the stored memory descriptions of objects are viewpoint-independent or object-centered representations of the object's 3-D structure. This means that the description derived from the retinal image will be the same, regardless of the viewpoint from which the object is seen. Each object is also taken to be stored in memory by means of one single description that can be triggered by each view of that object. Because both the perceptual descriptions and the stored memory descriptions of objects are viewpoint independent, there is no need for a normalization process to bring both types of representations into agreement; a direct mapping can thus be made. Because of this direct mapping, these theories predict that the recognition of objects does not depend on viewpoint (e.g., no differences in recognition time or error rate), under the assumption that the same viewpoint-independent perceptual description can be derived equally easily from all views.

This approach is exemplified by Marr and Nishihara's (1978) computational theory of object recognition and by Biederman's (1987) recognition-by-components (RBC) theory. For example, Biederman (1987) proposed that objects are represented as an arrangement of simple viewpoint-invariant parts (geons) and relations, called a *geon structural description* (GSD). The major assumption of RBC theory is that the same GSD will be recoverable from different viewpoints, if the same combination of relatively stable image features or nonaccidental properties (such as parallelism and symmetry; see Lowe, 1987; Wagemans, 1993) is visible. Because this will not always be the case, RBC theory predicts restricted rather than complete viewpoint invariance (for a discussion of this, see Biederman & Gerhardstein, 1995; Tarr, 1995; Tarr & Bülthoff, 1995).

In contrast, viewpoint-dependent or viewpoint-specific theories propose that object recognition is based on the matching of 2-D views of the object with the stored object description. Some of these theories (e.g., Ullman, 1989) assume that the object descriptions stored in memory are 3-D, whereas other theories (e.g., Poggio & Edelman, 1990; Ullman & Basri, 1991) assume that a collection of different 2-D views is stored in memory (i.e., the multiple-views approach). Of course, it cannot be the case that all possible views of all previously seen objects are stored in memory. Hence, some sort of normalization process will always be needed

---

Bert Willems and Johan Wagemans, Department of Psychology, University of Leuven, Leuven, Belgium.

This work was part of a doctoral dissertation of Bert Willems under supervision of Johan Wagemans. This research was supported by Research Grant CAW 96/07 from the Regional Impulse Program for the Humanities and from the Research Program of the Fund for Scientific Research–Flanders (FWO G.0130.98) to Johan Wagemans.

We thank Rob van Lier for useful advice during stimulus construction and statistical analyses. We also thank Géry d'Ydewalle, Pierre Jolicoeur, Larry Parsons, Mary Peterson, Rob van Lier, and an anonymous reviewer for helpful comments on previous drafts.

Correspondence concerning this article should be addressed to Johan Wagemans, Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium. Electronic mail may be sent to johan.wagemans@psy.kuleuven.ac.be.

Table 1
Overview of Current Theories of Object Recognition

Theory	Perceptual description	Stored description	Matching process
	Viewpoint-invariant theories		
Marr and Nishihara (1978)	3-D object model	3-D object model	Direct mapping
Biederman (1987)	GSD	GSD	Direct mapping
	Viewpoint-specific theories		
Ullman (1989)	2-D	3-D	Alignment
Poggio and Edelman (1990)	2-D	Multiple 2-D	View interpolation
Ullman and Basri (1991)	2-D	Multiple 2-D	Linear combination

Note. GSD = geon structural description (see text for a further explanation).

to relate the 2-D perceptual description to one of the stored memory representations. Because this normalization process is most likely time consuming and error prone, these theories predict that object recognition is dependent on viewpoint (e.g., significant differences in recognition time or error rate). More specific predictions about response times (RTs) and error rates can be derived from different viewpoint-dependent theories of object recognition (e.g., Bülthoff & Edelman, 1992; Bülthoff, Edelman, & Tarr, 1995; Edelman & Bülthoff, 1992) and will, in fact, be considered in Experiment 3.

A large number of empirical studies have addressed the question of whether object recognition depends on viewpoint (see Tarr, 1995, for a review). The majority of the findings leave no doubt that viewpoint-specific effects can be considerable. Whether these viewpoint-specific effects are typical for everyday object identification is still a matter of debate (e.g., Biederman & Gerhardstein, 1993, 1995; Biederman & Bar, 1999; Hayward & Tarr, 1997; Tarr & Bülthoff, 1995; Tarr, Bülthoff, Zabinksi, & Blanz, 1997; Tarr, Williams, Hayward, & Gauthier, 1998; Tjan & Legge, 1998). More interesting, however, is the following question: Given that viewpoint-specific effects are obtained, what is causing them? In principle, differences in RTs or error rates could arise from the difficulty in deriving perceptual descriptions from 2-D images. In this way, even theories that postulate object-centered perceptual descriptions could accommodate viewpoint-specific effects; that position has not been maintained, except to attribute sudden increases in RT or error rate to particularly bad views, such as when the viewing direction coincides with the object's elongation axis or when a critical part is occluded (see the criteria to test restricted viewpoint invariance as discussed by Biederman & Gerhardstein, 1993). The source of viewpoint-specific effects has generally been attributed to the normalization process needed to establish a match between the perceptual descriptions derived from the retinal stimulation and the representations in memory derived from previous encounters of that object.

Mental Rotation as Normalization?

Even at the level of the normalization process, a wide range of possibilities remains. The research presented here was aimed at testing some of the predictions that seem to follow from different proposals about possible normalization procedures. More specifically, our experiments concerned the possibility that an incremental transformation process known as *mental rotation* is used to try to establish a match between an internal 3-D model and a current 2-D view. R. N. Shepard and colleagues (see R. N. Shepard & Cooper, 1982, for a review) have established clearly that the time to discriminate standard shapes from mirror-reversed shapes varies linearly with increasing orientation difference between simultaneously or successively presented views, both within the picture plane and in depth. In subsequent research, they have provided evidence that the matching process is an analog in the sense that intermediate positions between the two views of one object seem to be represented internally during this process, as if the physical rotation between the two positions in space is simulated mentally.

Because similar dependencies of RTs on misorientation have been obtained in the literature on object recognition (e.g., Jolicoeur, 1985, 1988; Tarr & Pinker, 1989; see also Leek, 1998), it is tempting to assume that a similar mental rotation process underlies these viewpoint-specific effects as well. However, R. N. Shepard and colleagues (e.g., R. N. Shepard & Cooper, 1982) have always used tasks that involved handedness discrimination, not object identification as such. In fact, to obtain the specific mental rotation effects that have been observed (i.e., inverted U-shaped functions with the reversal at 180°), one has to assume that the object's top is localized first, which seems to require some viewpoint-invariant mechanism prior to mental rotation (for a further discussion, see Corballis, 1988; Gibson & Peterson, 1994; Hinton & Parsons, 1988; Jolicoeur, 1990; Tarr, 1995).

There is another problem with the possibility of mental rotation as a normalization process. It seems to follow logically from the notion of mental rotation that the representation being rotated is 3-D. It is hard to imagine how the 2-D views could be rotated to compensate for the effects of orientation in depth (e.g., try rotating a picture of a car to discover how it looks from a different angle). A fundamental problem with 3-D solid objects is self-occlusion, with some object parts being occluded from some viewpoints and new parts coming into view after rotation in depth (see Van Lier & Wagemans, 1999, for empirical research demonstrating how object regularity may be used to overcome this problem). Nevertheless, there is some ambiguity about this issue within the viewpoint-dependent approach to object recognition. For example, in an article defending the multiple-views approach, Tarr (1995, p. 67) writes: "Thus, the most plausible explanation for viewpoint dependency in recognition is that unfamiliar viewpoints between familiar views are matched to these views by interpolation and that unfamiliar viewpoints beyond familiar views are matched by normalization through the shortest 3-D path."

Perhaps the notion of an aspect graph (e.g., Koenderink, 1984; Van Effelterre, 1994) with explicit representations of qualitatively different, characteristic views (e.g., Blanz, Tarr, & Bülthoff, 1999; Cutzu & Edelman, 1994; Palmer, Rosch, & Chase, 1981) and of the transformation paths inducing the topological transitions between them (for an example, see Tarr, 1995, Fig. 10, p. 77) may be compatible with the pattern of RT results obtained in object recognition experiments. Even then, it seems to stretch the notion of mental rotation beyond what is acceptable if it is intended to be more than a metaphor. In any event, this issue deserves more attention than it has received so far. In a recent monograph on object recognition and visual cognition, Ullman (1996, p. 47) concluded: "It is still unclear, therefore, whether this process [of mental rotation] is an integral part of ordinary object recognition, or a special process that is used for limited purposes only."

An argument for the possibility of an analog process of mentally rotating an object description is usually provided in the context of a linear dependency of performance (RTs, accuracy) on the difference between the stored (canonical) orientation and the tested orientation (R. N. Shepard & J. Metzler, 1971). However, this result can only be considered a necessary condition, not a sufficient condition, because other normalization procedures predict the same dependency of performance on orientation difference without assuming this analog process of mental rotation (Poggio & Edelman, 1990; Tarr, 1995; Ullman & Basri, 1991). Thus, to decide whether this analog process of mental rotation is used, one needs other criteria that can be considered both necessary and sufficient. One of these is the presence of internal representations of intermediate orientations during the process of mental rotation. If it can be shown that the orientations in between the two to-be-aligned orientations are also represented during the response interval (in between the stimulus presentation and the response), an analog process of mentally rotating the object descriptions can safely be assumed (e.g., Cooper, 1976).

Another criterion that is used is based on the observation that the dependency of RTs is mediated by the apparent direction of rotation of the object, which is manipulated by inducing a motion after effect (Corballis & McLaren, 1982) or by presenting a sequence of views instead of a static view (Jolicoeur & Cavanagh, 1992). However, because an absence of this effect by itself cannot be interpreted as the absence of mental rotation (this would be a conclusion based on the null hypothesis), this criterion is only sufficient when it can be shown that the effect is apparent in a comparable experiment with the same design and stimuli but with a different task (where it can consequently be concluded that the objects were mentally rotated to solve the task). Based on this reasoning, Jolicoeur, Corballis, and Lawson (1998) showed that mental rotation is not involved in the recognition of plane-rotated objects, whereas it is involved in a direction-of-facing task using the same set of objects.

In this study, we want to use a different criterion for deciding whether mental rotation is involved in the recognition of in-depth rotated objects, based on the same reasoning as the one used by Jolicoeur et al. (1998). This time, the criterion is based on the observation that the dependency of performance on orientation differences is mediated by the axis of rotation that is needed to undo this difference of orientation (Parsons, 1987). Several studies have found these effects of rotation axis on performance when object descriptions are mentally rotated. For example, Pani (1993; see also Pani & Dupree, 1994) asked observers to predict the outcome of certain rotational motions and found that when the object was normal to the axis of rotation, the rotations were well understood and the outcome could be predicted easily. In contrast, when the axis of rotation was oblique to the object, performance depended on the orientation of the axis relative to the environment: With vertical orientations, performance was much better than with oblique orientations. In subsequent research (Pani, William, & Shippey, 1995), these results were generalized to the perception of rotational motion: When participants were asked to indicate the orientation of the axis and planes of rotation of a variety of objects (square, cube, octahedron, and tetrahedron), they performed well only when the axis and planes of rotation were aligned with the principal directions of the environment or when the object was elongated along the axis of rotation or constituted a generalized cone about it.

Parsons (1995) also found that humans have difficulties with some kinds of mental rotations but not with others. His participants had to imagine a Shepard-like object rotate about an axis and angle to accurately envision its appearance. The objects could be rotated over four different rotation axes. He found that performance was poorest when there was no coincidence among the axis of rotation, an object limb, and a principal environment axis; performance was better with alignment of the axis of rotation with a principal axis of the viewer's visual frame but not with one of the object's limbs; performance was still better with alignment between the axis of rotation and the object's major axis but with no alignment with one of the environmental axes; and performance was best with full coincidence among a principal axis of the viewer's visual frame, the object's major limb, and the axis of rotation.

Again, because an absence of this effect of manipulated path of rotation by itself is not conclusive as a test for the involvement of mental rotation in the context of the recognition of depth-rotated objects, we have to show that a task that uses the same set of stimuli and that is used in an experiment with a comparable design is able to reveal the above-mentioned effects. If this is the case, it is safe to conclude that in the task where the effect of the axis of rotation is absent the involvement of mental rotation is not very likely, whereas in the task where the effect was found this effect is due to the analog process of mentally rotating the object descriptions.

The Present Study

In sum, to test the role of mental rotation in 3-D object perception, we have included two variables in the experiments reported in the following discussion. First, we manipulated our observers' familiarity with a certain test view of a target object. This allowed us to test the effect of viewpoint on performance (difference in RTs or accuracy between seen views and unseen views).

However, these effects of viewpoint can easily be explained by any theory that postulates some sort of normalization procedure. Therefore, we included as a second variable the systematic manipulation of the orientation of the axis in space about which the objects were rotated to create viewpoint differences. From studies addressing the mental imagery and visual perception of rotational motion, we know that the orientation of the motion axis can have dramatic effects. This variable was included and was considered diagnostic for the use of mental rotation in our experiments, because an effect of viewpoint on recognition performance as such is insufficient.

So, evidence for a process of mental rotation to generalize performance to previously unseen views would consist of an effect of viewpoint combined with an effect of the specific path of rotation that relates the unseen views to the object descriptions stored in memory. If mental rotation were used as a process of normalization in the context of viewpoint-specific object perception, one should expect similar effects of the orientation of the axis of rotation. This rationale motivated the manipulation of rotation axis in our experiments.

Some of our experiments were replicated using only enantiomorphs (i.e., mirrored versions) as distractor objects. This additional manipulation was motivated by the claim of some researchers (e.g., Corballis, 1988) that it is only when the handedness of objects is required for discriminating between different objects that a process of mental rotation is induced (i.e., the rotation-for-handedness hypothesis). In that case, an effect of rotation axis could be expected when the only distractors used were the enantiomorphs.

Experiment 1

Novel objects were created that consisted of one major, elongated part and four smaller parts, each attached to a different side of the major part. This type of object was supposed to be theoretically neutral in the debate between viewpoint-dependent and viewpoint-independent theories of object recognition in the sense that it seemed to take an average position on many of the relevant dimensions that distinguish the types of objects typically used in the empirical research surrounding this debate. In Experiment 1 (and also in Experiment 3), objects were presented as sequences of 13 images, 30° apart, which simulated a continuous rotation of the object about a particular axis. Shortly afterward, a test image was presented that showed either the same object (target) or one that was similar but not identical (distractor). We derived the distractor objects from target objects by displacing the smaller parts, hence changing the global spatial configuration while maintaining the same components. Participants were asked to indicate whether the test image came from the object that they had seen in motion just before or from a slightly different one.

Why did we use these particular objects and this particular paradigm? The choice of this set of objects was motivated by three considerations. First, we wanted to use nonexisting objects. The reason for this is that we wanted to have control over which views were exposed to the observers. If we had used objects that were already encountered by our observers, it would have been impossible to control for this variable. The second consideration was that we did not want to leave open the possibility of our observers solving the task by means of a distinguishing feature that was unique for each to-be-recognized object (see Tarr & Bülthoff, 1995). Although we believe that in some cases object recognition is helped by means of a distinguishing feature, it is almost certain that this is not always the case (see Lawson, 1999; Vanrie, Béatse, Wagemans, Sunaert, & Van Hecke, in press; Vanrie, Willems, & Wagemans, in press). Therefore, objects used in our experiments could not be distinguished from each other by means of one feature that was unique for each object in the set. The third consideration was to make sure that our observers did not need to derive the metric details of the objects to discriminate between the targets and the distractors. The task could be solved based on the structural configuration of the different components, rather than on the exact position of these components relative to each other.

This particular paradigm was introduced to examine whether and how the orientation of the axis of rotation influenced object perception. Two aspects of the paradigm should be noted. First, the task did not involve naming, but it probably induced processes to derive perceptual descriptions from object images that can be taken to occur in natural object identification contexts. In one condition of Experiment 1, target images had been presented previously as snapshots along the motion path; in another condition, target images were 15° in between the previously seen snapshots but on the same motion path. We were interested to know (a) whether seen views would be easier than interpolated views and (b) whether certain rotation axes would yield better performance than others.

Second, the reason we presented a sequence of images simulating a continuous rotation of the object about a particular axis in space was not to induce strong percepts of apparent rotation (we have not formally optimized the parameters to that end). Instead, we only wanted to present observers with a sequence of separate views in such a way that (a) they received a series of separate views, (b) they could easily integrate the views in one single 3-D structural object description, and (c) they could easily see the object rotate in space. All of these conditions should favor either viewpoint independency or viewpoint dependency of mental rotation. If neither of these alternatives occurs, we have a novel finding that allows us to conclude viewpoint dependency and thus normalization, but normalization that is fundamentally different from mental rotation.

Method

Participants. Twelve advanced undergraduate or graduate students at the University of Leuven, who were between 18 and 25 years old, volunteered to participate in Experiment 1. They were all naive about the purpose of the experiment, and they had never seen the stimulus objects before. All of them had normal or corrected-to-normal vision.

Apparatus. Experiment 1 was carried out on a PC with a Pentium 133 Mhz processor. Stimuli were displayed on an SVGA computer screen with 1,024 × 768 spatial resolution and 75 Hz temporal resolution. Responses were given by pressing the Z key (for *same* responses) or the *M* key (for *different* responses) on a QWERTY keyboard.

Stimuli. Objects were constructed with the aid of the software package 3D Studio (Autodesk, Inc., 1993). We created eight novel target objects and made use of Biederman's (1987) geon theory to select the components constituting the objects. For each target object, a different large geon was used as the object's major part. Then, four other small geons were attached to this basis geon in a systematic way (see Figure 1). The components were sampled randomly from the set of 36 geons as described by Biederman (see Appendix for a complete description). Two of the components were attached at opposite sides near the middle of the basis geon (i.e., halfway along the elongation axis), and the other two components were attached at the ends of the basis geon (i.e., one near the top and one near the bottom, at opposite sides). In this way, all objects had a similar global configuration

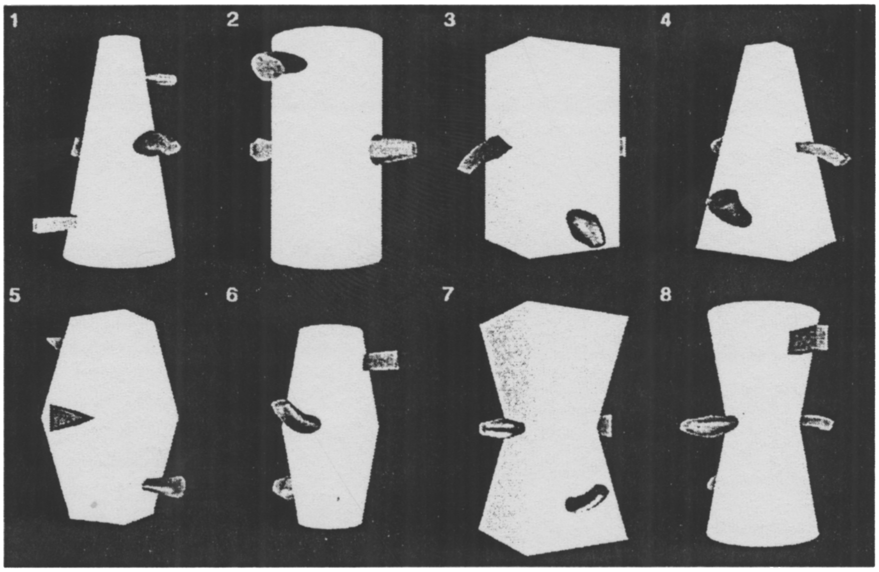

Figure 1. The six target objects (1–6) used in Experiments 1–5, and the two practice objects (7 and 8) used during instructions and practice only. All objects are shown in an arbitrarily chosen view.

with a different basis geon and different side parts. Of the eight target objects, six (1–6 in Figure 1) were used in the experimental trials and two (7 and 8 in Figure 1) were used to explain the task and to allow some practice. The objects were rendered under full-light conditions and in color. The small side parts were always colored copper, and the basis geon was always yellow. The objects were presented against a black background.

For each target object, we constructed three distractor objects with the same basis geon but with a changed configuration of components (see Figure 2 for an illustration with Object 1): In Distractor 1A, the two opposite components that were attached at the middle of the basis geon exchanged their position; in Distractor 1B, the two components that were attached at the top and at the bottom of the basis geon exchanged their position; and, in Distractor 1C, left and right as well as top and bottom components exchanged their position. So, each target object had its own distractor objects. By constructing the distractor objects in this way, we ensured that the task could not be performed on the basis of some distinctive feature. Target and distractor objects also did not differ in some minute metric detail (e.g., size, curvature, aspect ratio). Moreover, because only the first distractor type made target and distractor objects in effect enantiomorphs, the task was not a mere handedness-discrimination task. Instead, our object family and the specific target–distractor combinations seemed to require the derivation of a rather detailed GSD.

Procedure. Experiment 1 was controlled by a program designed with the aid of Superlab Pro (Cedrus Corp., 1997). A trial consisted of a

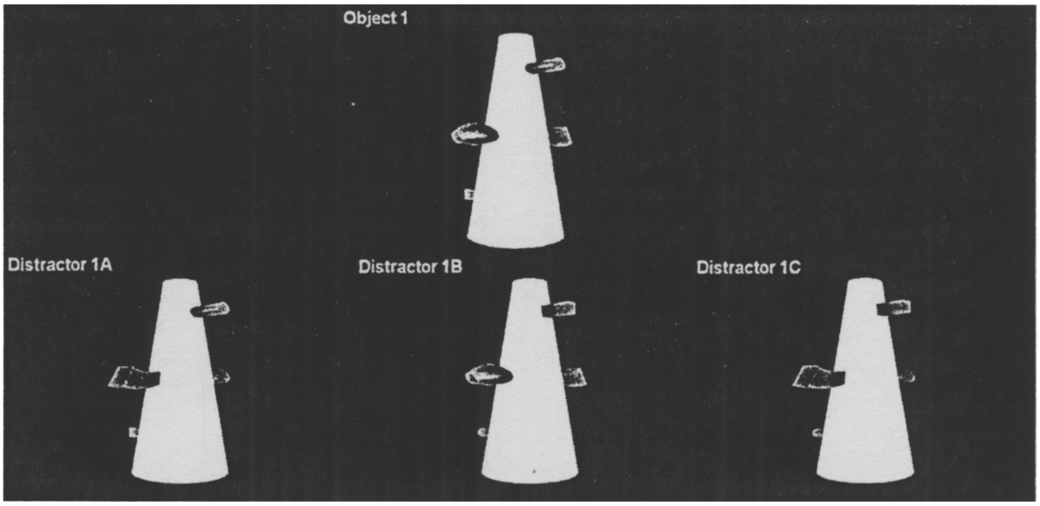

Figure 2. Distractor objects were derived from target objects in three different ways. This is illustrated here for Object 1. In Distractor 1A, components at the left and right sides of the object exchanged their position. This procedure in effect creates enantiomorph objects. In Distractor 1B, components at the top and bottom positions (at the front and back sides of the object) exchanged their position. In Distractor 1C, left and right as well as top and bottom components exchanged their position (see also Appendix).

sequence of 13 images (explained later) of a target object followed, after an interval of 2 s, by a test image either of the same object (i.e., a target view) or of one of its three distractors (i.e., a distractor view). All images were presented in a central area of the computer screen (about 10 × 11 cm) that was viewed from a viewing distance of 50 cm (i.e., about 11.5° × 12.5° of visual angle). Participants had to indicate whether the test image was a target view or not by pressing the *Z* key for *same* (target) and the *M* key for *different* (distractor), respectively. Instructions required participants to respond as fast and accurately as possible.

An apparent motion sequence consisted of 13 subsequent views, 300 ms each, of a rotating target object. Each image in the motion sequence differed from its predecessor and successor by a 30° rotation along a particular axis and in a particular direction (clockwise or counterclockwise, chosen randomly for each object). So, when the 13 views were shown, the object had made a complete rotation of 360° (see Figure 3 for an example). For each object, the first image in the motion sequence was always the same arbitrarily chosen view (i.e., 0° orientation). By showing this motion sequence, we gave the participants' visual system every chance to build a 3-D model of the target object (Edelman & Bülthoff, 1992).

In this motion sequence, we also manipulated the orientation of the rotation axis. The objects could be rotated over six different rotation axes (for a schematic overview, see Table 2):

- O_nE_n: The axis of rotation is not aligned with the main axis of the object and is not aligned with one of the axes of the environment.
- O_nE_x: The axis of rotation is not aligned with the main axis of the object but is aligned with the environment's horizontal axis (*x*-axis).
- O_nE_y: The axis of rotation is not aligned with the main axis of the object but is aligned with the environment's vertical axis (*y*-axis).
- O_aE_n: The axis of rotation is aligned with the main axis of the object but is not aligned with one of the axes of the environment.
- O_aE_x: The axis of rotation is aligned with the main axis of the object and is aligned with the environment's horizontal axis (*x*-axis).
- O_aE_y: The axis of rotation is aligned with the main axis of the object and is aligned with the environment's vertical axis (*y*-axis).

Table 2
Overview of the Different Rotation Axis Conditions

Rotation axis condition	Aligned with the object's main axis?	Aligned with one of the environmental axes?
O_nE_n	Not aligned	Not aligned
O_nE_x	Not aligned	Aligned with the *x*-axis
O_nE_y	Not aligned	Aligned with the *y*-axis
O_aE_n	Aligned	Not aligned
O_aE_x	Aligned	Aligned with the *x*-axis
O_aE_y	Aligned	Aligned with the *y*-axis

Rotations about the optical axis (*z*-axis) were not included because they do not cause a rotation in depth. Only one object was used in each rotation axis condition. Across participants, all target objects were used in all rotation axis conditions. So, each participant had a specific combination of object and rotation. We used this procedure to exert control over the type of object

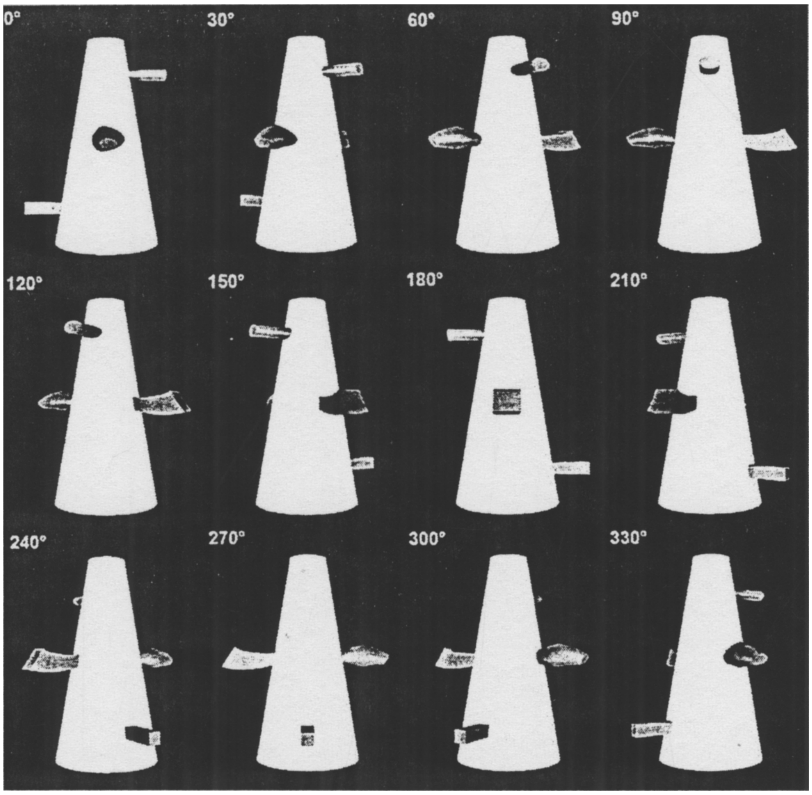

Figure 3. Example of a series of static snapshots, 30° apart and presented for 300 ms each, in Experiments 1 and 3: Object 1 in rotation axis condition O_aE_y (i.e., rotation axis aligned with the object axis and with the vertical axis of the environment; see also Table 2).

representations stored in visual memory by each participant, while at the same time avoiding the possibility that any rotation axis effects would be confounded with object-specific characteristics (which may be considerable; see, e.g., Van Lier & Wagemans, 1999).

After the motion sequence and an interval of 2 s, a test image was shown until the participant pressed one of the designated response keys. In one half of the trials, the test image was a target view (i.e., a match trial) and in the other half it was a distractor view (i.e., a nonmatch trial). A new random order of match and nonmatch trials was used for each participant.

To control the views presented to each participant, we randomly subdivided the participants in two groups. In the first, the test views of the target objects were always the same as those shown in the apparent motion sequence. Without the first and last view in the motion sequence (which were identical), 11 possible test views remained for each rotation (i.e., 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330°). The same views also occurred for the distractors; a different random choice from the three possible distractor objects was made on each trial. In the second group, test views were obtained by rotating the object 15° away from the views shown in the motion sequence along the same motion path (in effect, creating interpolated views). By omitting the view that was 15° away from the last view of the motion sequence (i.e., the 345° view), again 11 test views of the target object were obtained for each rotation (i.e., 15°, 45°, 75°, 105°, 135°, 165°, 195°, 225°, 255°, 285°, 315°). The same views also occurred for the distractors. Figure 4A illustrates the difference between seen and interpolated views by showing the positions of the views within the view sphere.

The total number of test trials for each participant was 264: (11 Target Views + 11 Distractor Views) × 6 Rotation Axes × 2 Repetitions. The two repetitions were presented in two successive blocks of 132 trials, separated by a short break of about 5 min.

The experiment was performed individually in a dimly lit room, with participants seated at about 50 cm from the monitor. First, the task was explained with the two additional objects (constructed on the basis of similar principles as the target objects). It was made clear that distractor objects were always similar to the target objects. To illustrate the task and the target–distractor relations, we gave each participant about 20 trials with the practice objects. Each participant received practice trials similar to those in the experiment (i.e., previously seen views for the first group and interpolated views for the second group). When the participant understood the task, the experimenter started the first series of 132 test trials. The total duration, including instructions, practice, and a break, was about 1 hr for each participant.

Results

Percentage of correct responses (% CRT) and RTs associated with correct responses were used as dependent variables. One participant, performing around chance level (< 55% correct), was replaced by another one (with the same object–rotation pairing, to fill the design as outlined earlier). In addition, RTs more than 3 *SD* below or above the mean (per observer) were also eliminated as outliers before we analyzed the data (i.e., less than 1% of the RTs). The same procedure was used in all subsequent experiments.

The data entered in the analyses of variance (ANOVAs) were averaged across 11 test views, per block, per observer, separately for match and nonmatch trials and separately for the six rotation axis conditions. Whether the test views were seen before or were interpolated between seen views was a between-subjects variable (called *test view type*).

Response times. Overall, match trials (2,835 ms) were faster than nonmatch trials (3,089 ms), $F(1, 22) = 5.47$, $p < .05$. The difference between seen views (2,568 ms) and interpolated views (3,355 ms) was even larger, $F(1, 22) = 8.25$, $p < .01$. There was no Trial Type × Test View Type interaction, $F < 1$. Both main effects are shown in Figure 5A. There was absolutely no main effect of rotation axis, $F < 1$, and rotation was also not involved in interaction effects, all *F*s < 1. One may wonder whether there is a systematic effect of the angular disparity between the final view in the apparent motion sequence and the test view in the match trials. This was not the case, $F < 1$.

Accuracy. There was only one statistically significant effect on % CRT: Match trials (81%) were responded to more correctly than nonmatch trials (60%), $F(1, 22) = 50.43$, $p < .0001$. The means for seen views and interpolated views followed the same trend as the RTs (73% vs. 68%, respectively) but the effect was not statistically reliable, $F(1, 22) = 1.42$, $p = .25$ (see Figure 5B). Again, there was no main effect of rotation axis, $F < 1$, although there was now a trend for an interaction with trial type, $F(5, 110) = 1.92$, $p = .10$, and even for a three-way interaction, $F(5, 110) = 1.47$, $p = .20$. The angular disparity between the final view in the motion sequence and the test view did not have any systematic effect, $F < 1$.

Additional experiment. We did an additional experiment with the same number of participants and the same procedure, except that the only distractor was the enantiomorph of the target object (i.e., only Distractor 1A). In contrast to the experiment with all the distractors included, there was no overall difference in response times between seen views (2,472 ms) and interpolated views (2,538 ms), $F < 1$. However, the trend in the Trial Type × Test View Type interaction, $F(1, 10) = 3.10$, $p = .11$ (see Figure 6A) suggested that the difference was in the expected direction for the match trials, with seen views yielding faster RTs (2,260 ms) than interpolated views (2,428 ms). More central to the motivation behind the experiment was the influence of the orientation of the rotation axis. This time, the main effect of rotation axis approached statistical significance, $F(5, 50) = 2.21$, $p = .067$. However, an a posteriori analysis (Tukey's honestly significant difference [HSD]) revealed that the only pairwise difference contributing to the main effect ($p = .051$) was that between rotation axis conditions O_nE_y and O_aE_x (2,766 vs. 2,346 ms, respectively). Rotation was not involved in interaction effects. With accuracy as the dependent variable, there were no statistically significant effects in this experiment (but trends were similar as for RTs; see Figure 6B).

Discussion

Even when an object is presented from all different sides, in a sequence of 12 different views, 30° apart, there is a remainder of view specificity in the performance of observers who are differentiating target objects from distractors with the same components in a different spatial configuration. RTs were faster for the views that were contained in the apparent motion sequence than for the views that were only 15° in between those views. Nevertheless, the conditions seemed to favor the construction of viewpoint-independent 3-D perceptual representations like Biederman's (1987) GSDs.

First, objects were constructed with idealized geon components. Components should be easy to derive from the images and to categorize according to the RBC system. Second, it should be possible to represent the type of structural relations between the components in an object-centered GSD (top vs. bottom, left vs. right) that was needed for task performance in this experiment. It is useful here to distinguish between different versions of RBC

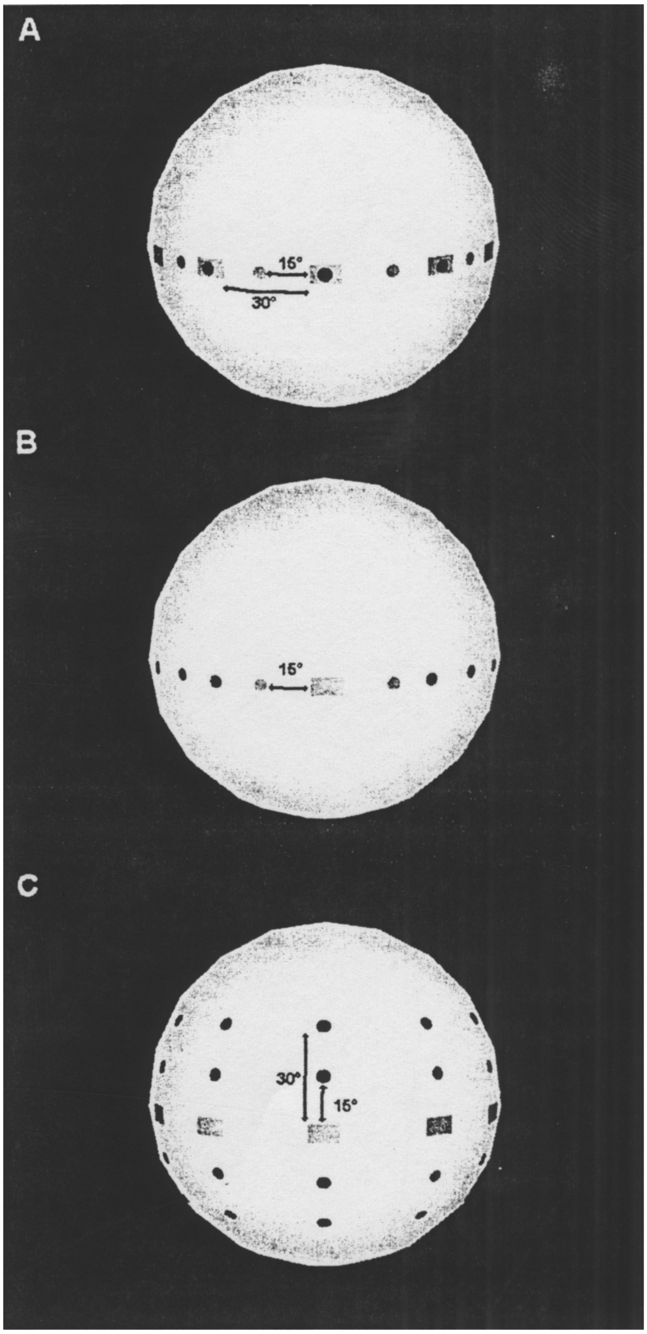

Figure 4. Illustration of the different target and test views (denoted by rectangles and circles, respectively) in all experiments by positioning them in a view sphere. For illustration purposes, only one rotation axis is shown (by changing the orientation of the objects within the view sphere, one can obtain all different rotation axis conditions). A: In Experiment 1, test views were always on the path induced by the apparent motion sequence of target views. In one condition, test views were identical to views presented before (0°, 30°, 60°, . . .); in another condition, they were in between seen views (15°, 45°, . . .). B: In Experiment 2, only one target view was presented and the test view could differ from it by 0° to 90° (both clockwise and counterclockwise, in 15° steps) on one of six paths in space (see Table 2). C: In Experiment 3, the target object was again shown as a series of views but the test views were now rotated away from the target views by 15° or 30° on a path orthogonal to the path defined by the apparent motion sequence.

theory. In principle, RBC theory is supposed to be capable of distinguishing a briefcase from a file drawer and a mug from a bucket (see Biederman, 1987, Figure 3, p. 119). However, a subsequent neural-network implementation of RBC theory (Hummel & Biederman, 1992; Hummel & Stankiewicz, 1996) leaves some uncertainty about the capacity of the model to distinguish between our target and distractor objects. Indeed, more recent work from the same laboratory (Shyi, Goldstone, & Hummel, 1998) has indicated that so-called *bound object representations,* with parts being bound to a specific location on a larger component, are the most difficult to compute. Topological properties, such as connected versus nonconnected, are much more salient for the perceptual system (for some empirical evidence, see Saiki & Hummel, 1998; Van Lier & Wagemans, 1998). Finally, the coherent apparent motion sequence should have worked in favor of an integrative 3-D model and against a view-based representation format. We do not want to stress this finding too much (because it awaits further corroboration in more powerful experimental designs), but it constitutes a puzzling result.

More central to the major goal of this article and the remainder of the research to follow is the question of whether the orientation

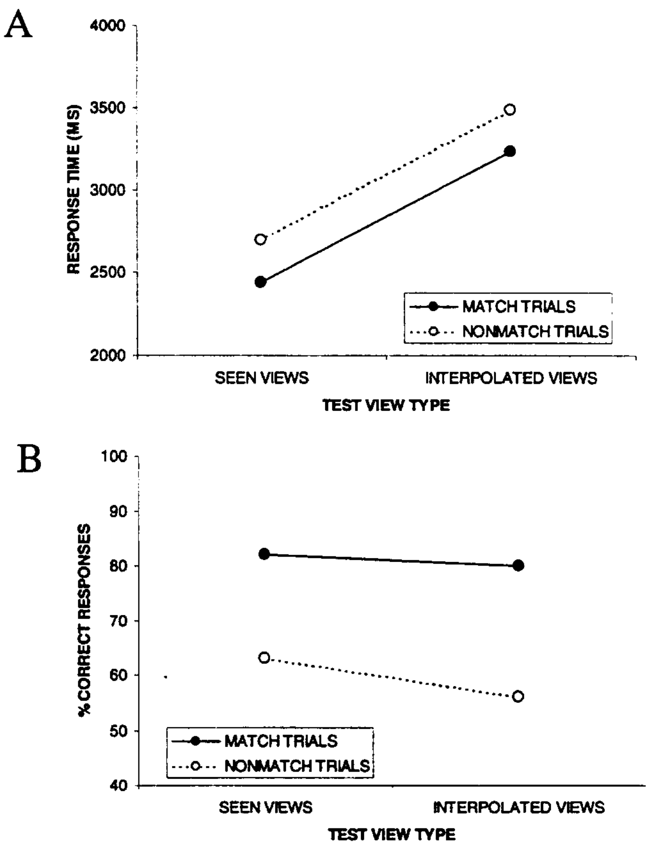

Figure 5. Performance levels in Experiment 1 with a sequence of views and three types of distractors. A: Response times (in milliseconds) for match and nonmatch trials (indicated by solid and dashed lines, respectively), separately for the two experimental groups, depending on whether test views were seen before (seen views) or in between those seen before (interpolated views). B: Percentage of correct responses, separately for match and nonmatch trials and for seen versus interpolated views.

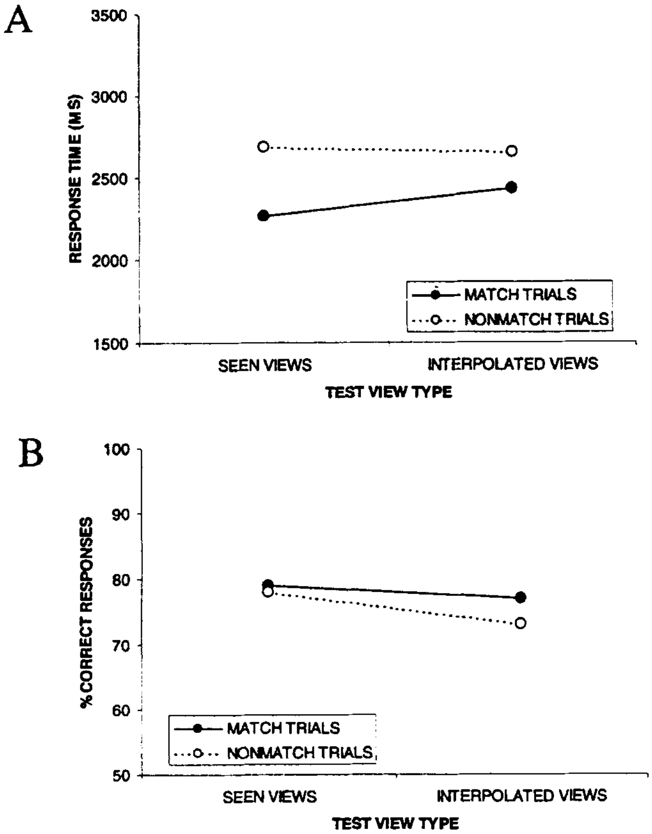

Figure 6. Performance levels in the additional Experiment 1 with a sequence of views and only enantiomorphs as distractors. A: Response times (in milliseconds), separately for match and nonmatch trials and for seen versus interpolated views. B: Percentage of correct responses, separately for match and nonmatch trials and for seen versus interpolated views.

of the axis about which the object is seen to rotate affects performance afterwards (regardless of whether one believes object-centered or viewer-centered perceptual representations are used). The results were clearcut: There was no sign of any systematic difference between the rotation axis conditions for RTs or for % CRT. Even when the only distractors used were enantiomorphs (handedness discrimination), a task that should be able to trigger a mental rotation procedure according to Corballis's (1988) rotation-for-handedness hypothesis, we did not find this expected differential effect of rotation axis. Nevertheless, the orientation of the rotation axis was shown to have strong effects on mental rotation tasks in quite a number of studies (e.g., Just & Carpenter, 1985; Pani, 1993; Pani & Dupree, 1994; Pani et al., 1995; Parsons, 1987, 1995; Shiffrar & Shepard, 1991).

So, this pattern of results raises doubts as to the role of mental rotation (i.e., the procedure that incrementally transforms one view into another as if the object moves along the shortest path in space) in the context of viewpoint-dependent object perception. This notion is further put to test in Experiments 2 in which the test view will have to be matched to only one target view instead of to a whole sequence of views.

Experiment 2

In Experiment 1, by presenting a sequence of views in apparent rotation about one of six differently oriented axes and by presenting test views that were always on the same path in space, we tested whether the orientation of the rotation axis had an influence on the visual system's encoding of the 3-D object. It turned out to have a very limited effect. None of the factors involving rotation axis was statistically significant. Even in the additional experiment, where enantiomorphs were used as the only distractors, only one pairwise difference between rotation axis conditions approached significance, whereas the absence of a Rotation × Match interaction suggested that this effect probably had little to do with the matching as such.

A more stable effect resulted from the use of two different types of test views, those that occurred as snapshots in the apparent rotation sequence (i.e., seen views) and those that were intermediate between the snapshots in the apparent rotation sequence (i.e., interpolated views). Seen views were matched faster and more accurately than interpolated views although the difference between them was only 15° and the experimental paradigm seemed a priori to favor the extraction of a more abstract 3-D representation rather than a mere collection of view-specific 2-D representations. Because the experiments were not designed to have maximal statistical power to test it, this difference did not always get the necessary statistical support; the fact that the trend was systematically in the same direction increases our confidence that it is not a spurious result.

Taken together, these results present a challenge to current theories of viewpoint-dependent object recognition. On the one hand, the experimental data seem to support the notion of view-specific representations, while, on the other hand, the lack of a rotation axis effect seems to argue against the use of an incremental mental rotation procedure as the most prototypical normalization process. To bring our experimental paradigm closer to what is more commonly used in the context of viewpoint-specific versus viewpoint-invariant object recognition, we decided to present only one target view per object (instead of a whole sequence) and to examine the effect of angular disparity between test and target view over a much larger extent (instead of only 0° or 15°). We continued to examine whether the matching performance was affected by the orientation of the axis about which test and target views had to be rotated to be brought into correspondence. Perhaps this variable, known to influence spatial reasoning and imagery tasks (e.g., Pani et al., 1995; Parsons, 1995), would become more prominent in a matching task in which the normalization process would have to mentally undo much larger viewpoint differences.

Method

Participants. Six first-year psychology students at the University of Leuven volunteered to participate for course credit. All of them were naive and had normal or corrected-to-normal vision.

Apparatus and stimuli. The same set of target and distractor objects was used as in Experiment 1.

Procedure. In contrast to Experiment 1, a trial consisted of only one view of the target object for 2 s, followed (after 2 s) by a test view of either the same object or one of its three distractor objects (chosen randomly). Participants had to respond *same* in the first case and *different* in the second (by pressing the *Z* key or the *M* key, respectively). Although we did not show an apparent rotation of the target object about a well-defined axis in space, rotation axis was included as an experimental variable by manipulating the shortest path needed to align the view of the target object with the test view. This shortest path was one of the six rotation axes included in Experiment 1. Each participant had only one of the six objects in each rotation axis condition, but, across participants, all target objects were used in all rotation axis conditions.

After the first view of the target object and a 2-s interval, a test view was shown until one of the two designated response keys was pressed. The test view differed from the target view by a variable amount from 0° (i.e., no difference) to 90° in 15° increments, both clockwise and counterclockwise (i.e., 13 possible test views: −90°, −75°, −60°, −45°, −30°, −15°, 0°, 15°, 30°, 45°, 60°, 75°, 90°); the same 13 views of a distractor object were also included (a different random choice from the three possible distractor objects was made on each trial). Figure 4B illustrates the positions of the test views within the view sphere.

The total number of test trials for each participant was 312: (13 Target Views + 13 Distractor Views) × 6 Rotation Axes × 2 Repetitions. The two repetitions were presented in two successive blocks of 156 trials, separated by a short break of about 5 min.

Experiment 2 was performed individually in a dimly lit room, with participants seated at a comfortable distance from the monitor (about 50 cm) and the keyboard (about 25 cm). First, the task was explained with the two additional objects (the same as those used in Experiment 1). It was made clear that distractor objects were always similar to the target objects. To illustrate the task and the target–distractor relations, we gave each participant about 20 trials with these practice objects. When the participants understood the task, the experimenter started the first series of 156 trials. The total duration, including instructions, practice, and a break, was about 1 hr for each participant.

Results

One participant was replaced by another one (with the same object–rotation pairing) because he performed around chance level (< 55% correct). A small number of RT outliers (< 1%) were also removed from the data before the ANOVAs were run. The data entered in the ANOVAs were averages across two views (one from each block) in the case of an angular disparity of 0° or across four views (two from each block) in the case of all other viewpoints (i.e., clockwise and counterclockwise rotated views were collapsed). This was done separately for each participant, separately for match and nonmatch trials, and separately for the six rotation axis conditions. When none of the trials brought together in a cell of this data matrix was responded to correctly, the column average was taken to fill that cell (i.e., the mean across the five remaining participants for that particular experimental condition). This happened in 7% of the cases.

RTs. The overall effect of the angular disparity between target and test view (i.e., angular disparity) was highly significant, *F*(6, 30) = 26.66, *p* < .0001, as was its interaction with trial type, *F*(6, 30) = 10.79, *p* < .0001, due to the fact that only match trials showed the angular disparity effect (see Figure 7A). Linear regression of RTs for match trials against angular disparity yielded an excellent fit (R^2 = .982) and an estimated normalization rate of 39°/s. Rotation axis produced no reliable main effect, *F*(5, 25) = 1.77, *p* = .15. Moreover, none of the interactions with rotation axis were statistically reliable: *F* < 1, for Angular Disparity × Rotation; *F* < 1, for Trial Type × Rotation; and *F*(30, 150) = 1.32, *p* = .14, for the three-way interaction. We have also computed linear regressions for each of the rotation axis conditions separately. As can be seen in the top of Table 3, the normalization rates are rather similar for all of these conditions; if anything, the condition that normally produces poorest performance in spatial reasoning tasks (i.e., O_nE_n) produced the fastest normalization rate here.

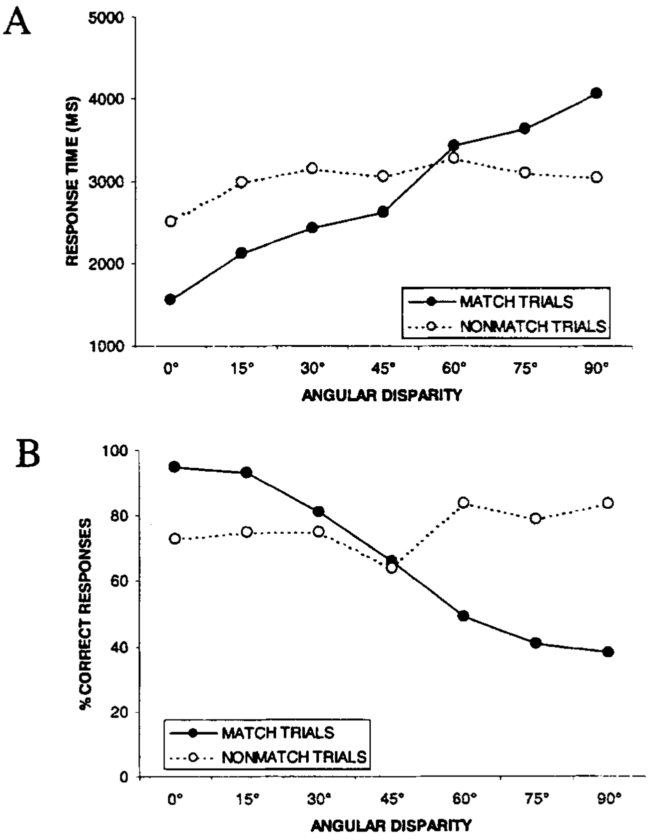

Figure 7. Performance levels in Experiment 2 with only one target view and three types of distractors. A: Response times (in milliseconds) as a function of the angular disparity between the test view and the target view, separately for match and nonmatch trials. B: Percentage of correct responses as a function of the angular disparity between the test view and the target view, separately for match and nonmatch trials.

Accuracy. A similar pattern of results was obtained for % CRT. Angular disparity produced highly significant effects, both as a main effect, *F*(6, 30) = 18.11, *p* < .0001, and in interaction with trial type, *F*(6, 30) = 19.83, *p* < .0001 (see Figure 7B). In contrast, rotation axis produced no significant effects, neither as a main effect, *F* < 1, nor in interaction with other variables: *F* < 1, for Angular Disparity × Rotation; *F*(5, 25) = 1.62, *p* = .19, for Trial Type × Rotation; and *F*(30, 150) = 1.42, *p* = .09, for the three-way interaction.

Additional experiment. As for Experiment 1, we did an additional experiment with the same number of participants and the same procedure but with only enantiomorphs as distractors. The main effect of angular disparity was highly significant for both RTs, *F*(6, 30) = 11.84, *p* < .0001, and % CRT, *F*(6, 30) = 19.44, *p* < .0001, as was the interaction of this variable with trial type: *F*(6, 30) = 3.81, *p* < .01, for RTs (see Figure 8A); *F*(6, 30) = 3.96, *p* < .005, for accuracy (see Figure 8B). These interactions were due to a stronger effect for match trials than for nonmatch

Table 3

Linear Regressions of Response Times Against Angular Disparity for Match Trials in Experiment 2

Rotation axis condition	Intercept	Slope	*r*	Normalization rate	% correct responses
Main Experiment 2 (with three types of distractors)					
O_nE_n	1,658	18.55	.922	53.9°/s	77.38
O_nE_x	1,646	25.56	.901	39.1°/s	62.50
O_nE_y	1,667	29.00	.970	34.5°/s	68.65
O_aE_n	1,930	20.18	.838	49.5°/s	65.08
O_aE_x	1,317	39.45	.958	25.4°/s	62.50
O_aE_y	1,423	30.91	.967	32.3°/s	63.49
Additional Experiment 2 (with only enantiomorph distractors)					
O_nE_n	1,428	15.06	.852	66.4°/s	83.13
O_nE_x	1,701	22.47	.985	44.5°/s	75.99
O_nE_y	1,436	23.74	.958	42.1°/s	78.97
O_aE_n	1,120	27.19	.934	36.8°/s	74.21
O_aE_x	1,702	19.83	.627	50.4°/s	73.61
O_aE_y	1,349	17.33	.820	57.7°/s	71.03

Note. O_nE_n = axis of rotation is not aligned with the main axis of the object and is not aligned with one of the axes of the environment; O_nE_x = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's horizontal axis (*x*-axis); O_nE_y = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's vertical axis (*y*-axis); O_aE_n = axis of rotation is aligned with the main axis of the object but is not aligned with one of the axes of the environment; O_aE_x = axis of rotation is aligned with the main axis of the object and is aligned with the environment's horizontal axis (*x*-axis); O_aE_y = axis of rotation is aligned with the main axis of the object and is aligned with the environment's vertical axis (*y*-axis).

trials. None of the effects involving rotation axis were statistically significant. Linear regression of RTs for match trials against angular disparity yielded a good fit ($R^2 = .923$) and an estimated normalization rate of 48°/s. The separate linear regressions for each of the rotation axis conditions (see the bottom of Table 3) suggest that the normalization rates are rather similar for all of these conditions; if anything, as in the experiment with all the distractor types, the O_nE_n condition produced the fastest normalization rate.

Discussion

Two results of Experiment 2 stand out. First, with an increasing angular disparity between test and target view, matching became increasingly more difficult as indicated by a strong decrease in the accuracy (% CRT) as well as a strong increase in RTs for correct match trials. The linear regressions revealed rather slow normalization rates around 40°/s, more in line with classic mental rotation studies (e.g., Parsons, 1987; R. N. Shepard & Cooper, 1982) than with object recognition experiments (e.g., Tarr, 1995). This makes it even more strange that, second, the orientation of the axis along which the test view was rotated away from the target view had no effect at all. Because rotation axis is known to affect many different tasks involving spatial reasoning or imagery in 3-D space, we conclude that the normalization process used to match the different views is very unlikely to make use of a mental rotation in 3-D space.

Experiment 3

In Experiments 1 and 2, we found several effects of viewpoint specificity in our participants' performance. In Experiment 1, participants responded faster to previously seen views than to views that were only 15° away from them, and in Experiment 2, performance was linearly dependent on the angular disparity between the seen view of the target object and the tested view. By manipulating the kind of 3-D rotation, we showed that this pattern of results need not be interpreted as evidence for the use of a mental transformation process that mimics the physical rotation in 3-D space.

Theoretical alternatives exist that are able to explain the viewpoint dependency of 3-D object recognition without having to invoke a process of mental rotation. All of these theories assume that 3-D objects are stored in memory as a collection of 2-D views (i.e., the multiple-views approach; for reviews, see Bülthoff et al., 1995; Tarr, 1995). Incoming perceptual descriptions are matched against a 2-D view, which is created either by linearly combining the stored views (Ullman, 1996; Ullman & Basri, 1991) or by interpolating between the stored views (Poggio & Edelman, 1990). Although these two theories can be regarded as variants of the same underlying idea (i.e., that information sufficient for recognition can be found in the 2-D locations of features), in some circumstances they predict different results.

The first goal of Experiment 3 was to test the different predictions made by the two groups of multiple-views approaches. According to a theory of recognition based on the linear combination of stored views (Ullman & Basri, 1991), performance will be good when test views are on a path between previously seen views because the seen views then span the view space of the test views. Performance will be harmed when views orthogonal to that path are extrapolated (because no view is available to span the orthogonal dimension of the view space). Performance will probably be harmed more when the coordinates of a feature vary in only one

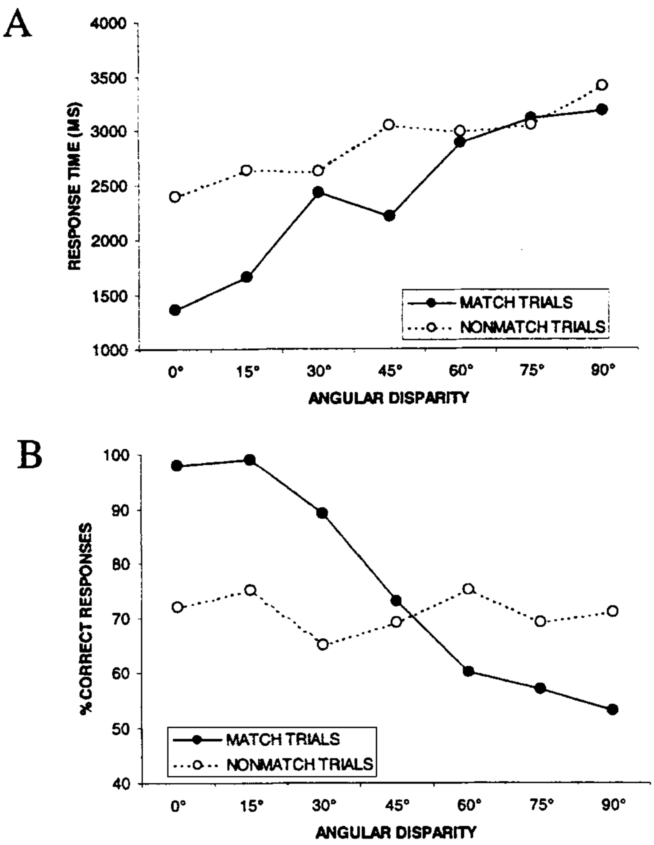

Figure 8. Performance levels in the additional Experiment 2 with only one target view and only enantiomorphs as distractors. A: Response times (in milliseconds) as a function of the angular disparity between the test view and the target view, separately for match and nonmatch trials. B: Percentage of correct responses as a function of the angular disparity between the test view and the target view, separately for match and nonmatch trials.

direction (x or y) of the 2-D image (i.e., in our rotation axis conditions O_nE_x, O_nE_y, O_aE_x, and O_aE_y) than when the coordinates vary more generally (i.e., in our rotation axis conditions O_nE_n and O_aE_n). Ullman (1998) has also explained how object identification is possible based on views that have never been seen before by using a linear-combinations scheme. In contrast, according to the view-interpolation model (Poggio & Edelman, 1990), equal performance is predicted as long as the distance between the previously seen views and the test views remains the same. When this distance increases, accuracy should decrease. Edelman (1999) has also explained how object identification is possible based on views that have not been seen before by using a view-interpolation scheme. RTs are usually less important in comparisons of human performance against computational models because they are much more implementation dependent (see Bülthoff & Edelman, 1992). Experiment 3 was designed to test these two predictions.

A second goal of Experiment 3 was to provide stronger evidence for the finding that even small differences between views presented in a target sequence and test views immediately afterward resulted in measurable behavioral differences (Experiment 1). Not all of the differences were statistically reliable in that experiment, but the number of participants was rather small and the effect of viewpoint was tested between two different groups of participants. Therefore, in Experiment 3, the number of participants was increased while every variable was included in the design as a within-subjects variable.

To investigate the two predictions made by the different multiple-views approaches and at the same time to provide stronger evidence for the conclusions drawn in Experiments 1 and 2, we designed Experiment 3 so that the views at which the to-be-recognized objects were shown (i.e., test view type) were of one of the following four types: (a) views that were already shown in the motion sequence (no generalization); (b) views rotated 15° away from the views already shown in the motion sequence, on the same path of rotation as shown in the motion sequence (15° interpolation); (c) views rotated 15° away from the views already shown in the motion sequence but now on a path orthogonal to the path of rotation (15° extrapolation); and (d) views 30° rotated away from the seen views, on a path orthogonal to the seen path of rotation in the motion sequence (30° extrapolation). At the same time, we again manipulated the axis of rotation that was used to create the views that were shown in the motion sequence.

Method

Participants. Participants were 12 students at the University of Leuven who were paid a small amount for participating in Experiment 3. All of them were naive about the purposes of the experiment, and all of them had normal or corrected-to-normal vision.

Apparatus and stimuli. The same apparatus and set of target objects were used as in Experiments 1 and 2.

Procedure. As in Experiment 1, a trial consisted of a sequence of a target object (13 views, 30° apart, 300 ms each) followed (after 2 s) by a test view of either the same object or of the mirror-reversal of the target object (handedness discrimination). Participants had to indicate whether the test view could be the target object (*Z* key) or not (*M* key).

The rotations were the same as those used in Experiments 1 and 2, and every two participants had a unique combination of object and rotation axis condition (as in Experiments 1 and 2). Across participants, all target objects were used in all rotation axis conditions.

In Experiment 3, we manipulated the amount of generalization needed to bring the test view in alignment with one of the views already shown in the motion sequence. The test view was either a view that was already shown in the motion sequence (no generalization; see Figure 4A), a view 15° rotated away from the views shown in the motion sequence but rotated about the same axis as used to generate the motion sequence (15° interpolation; see Figure 4B), a view 15° rotated away from the views already seen in the motion sequence but rotated about an axis orthogonal to the axis used to generate the motion sequence (15° extrapolation; see Figure 4C), or a view 30° rotated away from the seen views again about an orthogonal axis (30° extrapolation; see Figure 4C). The same views were also shown for the mirror-reversed versions of the target object (distractors).

The total number of trials for each participant was 528: 11 Viewing Positions × 2 Object Types × 6 Rotation Axes × 4 Test View Types. Because of the large number of trials, we blocked the trials per object–axis combination so that the number of trials per block was more manageable. The duration of such a block was about 20 min; in one session, two blocks were done by each participant with a 5 min rest in between. Participants thus performed three sessions, often on consecutive days. The rest of the procedure was as in Experiments 1 and 2.

Results

We analyzed % CRT and RTs associated with correct responses as dependent variables. One participant was replaced by another (with the particular object–rotation pairing to counterbalance the design), because she performed around chance level (< 55% correct). A small number of RT outliers (< 1%) were also removed from the data before the ANOVAs were run.

RTs. As in Experiments 1 and 2, match trials (2,048 ms) were faster than nonmatch trials (2,141 ms), although this difference was now only marginally significant, $F(1, 11) = 4.72, p = .052$. The effect of test view type was highly significant, $F(3, 33) = 11.22, p < .0001$ (1,996 ms, 2,059 ms, 2,089 ms, and 2,234 ms for no generalization, 15° interpolation, 15° extrapolation, and 30° extrapolation, respectively; see Figure 9A for the interaction between object type and test view type).

The third variable of interest was the orientation of the rotation axis. This time, the main effect of rotation axis approached statistical significance, $F(5, 55) = 2.16, p = .071$. However, an a posteriori analysis (Tukey's HSD) revealed that the only pairwise difference contributing to the main effect ($p = .074$) was that between rotation axis conditions O_nE_x and O_aE_y (2,380 vs. 1,792 ms, respectively).

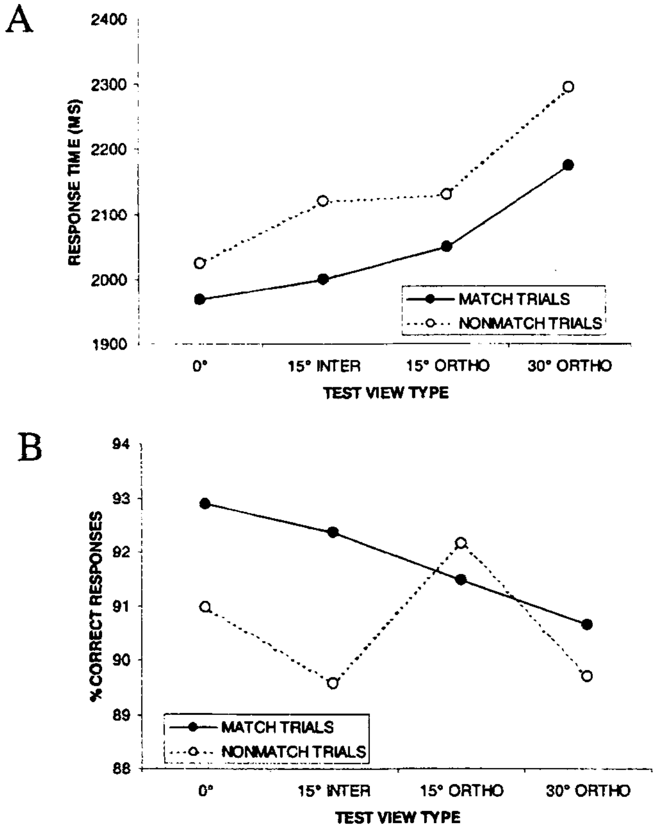

Figure 9. Performance levels in Experiment 3 with a sequence of views and only enantiomorphs as distractors. A: Response times (in milliseconds), separately for match and nonmatch trials and for four different test view types: 0° (views already seen in the motion sequence), 15° interpolation (INTER; i.e., along the induced motion path), 15° extrapolation (i.e., orthogonal [ORTHO] to the induced motion path), and 30° extrapolation (i.e., also orthogonal to the induced motion path). B: Percentage of correct responses, separately for match and nonmatch trials and for the same four generalization conditions.

The rotation axis was also involved in an interaction effect with test view type: $F(15, 165) = 2.09, p < .05$. However, according to a posteriori analyses (Tukey's HSD), the effect of rotation axis was quite different from the one found in the literature on spatial reasoning in all of the four test view types. For none of the four test view types did manipulation of the rotation axis result in systematically better performance when it was aligned with the object's main axis or with one of the environmental axes.

The interaction between rotation axis and object type was marginally significant, $F(1, 11) = 4.72, p = .052$. An a posteriori analysis (Tukey's HSD) revealed that the difference between match and nonmatch trials was only significantly different for rotation axis conditions where the axis of rotation was not aligned with the object's main axis (i.e., O_nE_n, O_nE_x, and O_nE_y). All other interaction effects did not reach the required level of statistical significance, largest $F = 1.08$).

To test the predictions of the linear combination and view interpolation models, we tested some a priori contrasts within the first level of object type (i.e., only for the match trials) and across the different levels of rotation axis. These planned comparisons revealed that the trials with no generalization were not significantly different from trials with a 15° interpolation ($p = .56$) and that the 15° interpolation was not significantly easier than the 15° extrapolation ($p = .22$) but that the 15° extrapolation was significantly easier than the 30° extrapolation ($p < .001$). When we tested the same two contrasts, but this time only for those rotation axis conditions where the features needed to recognize the objects vary only in one image direction, the same conclusions could be drawn, although the difference was now somewhat less pronounced ($p = .13$, $p = .11$, and $p < .01$, respectively).

Accuracy. There were no statistically significant effects on % CRT in Experiment 3. The largest F value was obtained for the main effect of object type, $F(1, 11) = 1.56, p = .23$ (see Figure 9B for the interaction between object type and test view type). It is important to note the lack of a speed–accuracy trade-off that would reduce the validity of the RT effects.

Discussion

The purpose of Experiment 3 was twofold. First, the experiment was designed to provide stronger evidence for the effect of viewpoint in the absence of a systematic effect of the alignment of the axis of rotation with one of the environmental axes or with the object's main axis. For a discussion of this replicated result, see the Discussion in Experiment 1. Because the trials were now blocked per object–axis combination (i.e., each trial of one block showed the same apparent rotation of the target object), it was even more strange that under these circumstances effects of viewpoint were found (after some trials, it seems that participants would have had every chance to store a 3-D object description of the target object).

The second purpose of Experiment 3 was to test the predictions made by the different normalization approaches. The analysis on the RTs did not reveal that extrapolations became significantly more difficult than interpolations, not even when participants differentiated between those rotation axis conditions inducing features to vary in only one image direction and those inducing these features to vary in more image directions. Note also the difference in performance between the 15° and 30° extrapolation conditions. This kind of performance could be expected if recognition was based on the approximation of a multidimensional surface by means of stored views (Poggio & Girosi, 1990), as in the view interpolation model (Poggio & Edelman, 1990). This does not mean, however, that the linear combination approach to object recognition should be rejected as psychologically implausible. For example, our effects were obtained on RT as dependent variable instead of accuracy. Further experiments are needed to test some additional predictions made by these two kinds of theories (see Bülthoff et al., 1995), preferrably by comparing human performance against implemented versions of the models (e.g., Edelman, 1999; Edelman & Duvdevani-Bar, 1997; Liu, Knill, & Kersten, 1995; Tjan & Legge, 1998).

Experiment 4

In Experiments 1–3, we found very little evidence for systematic effects of the 3-D transformations on our participants' performance in a variety of tasks. When we manipulated the shortest path needed to align the tested view of the target object with a view already shown in the motion sequence of the target object or a single view of the same target object, the performance of the participants was not systematically influenced by the axis of rotation being aligned with either the object's main axis or one of the environmental axes. Because the views that needed to be aligned were always separated in time, the participants had to compare the incoming test view with a view already stored in memory. However, it could be argued that a different process occurs when the two views that need to be aligned are shown simultaneously. Indeed, there is evidence in the literature that suggests important differences between these two kinds of paradigms (sequential vs. simultaneous matching). For example, the putative rates of mental rotation are generally much lower when the two views of the target object are presented simultaneously than when the target view has to be aligned with a view of this object stored in memory, although other factors could be invoked to explain this difference (for a review, see S. Shepard & D. Metzler, 1988). This means that the slopes of the RTs as a function of the angular difference between the to-be-aligned views are generally larger for simultaneous matching tasks than for sequential matching tasks. Perhaps the effect of rotation axis becomes stronger when normalization is slower.

To test for this possibility, we replicated Experiment 2 (with only Distractor 1A, resulting in a handedness discrimination task) with the to-be-aligned views presented simultaneously.

Method

Participants. Participants were 6 members of the Laboratory of Experimental Psychology at the University of Leuven. All of them were naive about the purposes of the experiment, and all of them had normal or corrected-to-normal vision. Two of them had participated in one of the preceding experiments. Because the views in Experiment 4 were always visible on the screen, it was not necessary for all participants to be completely unfamiliar with the particular views of the objects used.

Apparatus and stimuli. The only difference from Experiments 1–3 was that the stimuli were somewhat smaller to allow them to be presented side by side on the screen (about 9.5° × 11.4° of visual angle).

Procedure. As in Experiment 2, a trial consisted of two views, but this time they were not presented sequentially but side by side on the screen: One was a view of the target object, and the other was a test view of either the same object or of the mirror-reversal of the target object (handedness discrimination). All trials were shown twice, once with the target shown on the left side of the screen and once on the right. For participants, there was no meaningful distinction between target and test views: Their task was only to indicate whether the two views could be depicting the same object (*Z* key) or not (*M* key).

Rotation axis was again included as an experimental variable by manipulating the shortest path needed to align the view of the target object with the test view. This shortest path was again one of the six rotation axes used in Experiments 1–3. This time, however, it was not necessary to present the objects in only one rotation axis condition, because it was no longer important to have control over which views of the target object were already shown to each participant (as was the case in all of the preceding experiments). Therefore, each target object was used in the six different rotation axis conditions.

The angular disparity between the two depicted objects was manipulated in steps of 15°, from −60° to 60°, resulting in 10 possible test views. (The 0° condition was included twice, because the data for the two different directions of rotation could then be collapsed when calculating average RTs and percentages of correct answers.)

The total number of trials for each participant was 1,440: 10 Viewing Positions × 2 Object Types × 2 Modes of Presentation (target object left or right) × 6 Rotation Axes × 6 Objects. Because of the large number of trials, we blocked the trials per object so that the number of trials per block was more manageable. The duration of such a block was about 15 min; in total, there were three sessions of about 35 min, each containing two blocks with the same object, interrupted by a short break. The rest of the procedure was as in Experiments 1–3.

Results

RTs. The ANOVA on the RTs revealed three main effects. First, match trials (1,647 ms) were faster than nonmatch trials (1,793 ms), $F(1, 5) = 9.61, p < .05$. There was also a main effect of the angular disparity between the two objects depicted in the trial, $F(4, 20) = 25.89, p < .00001$, and a main effect of rotation axis, $F(5, 25) = 3.92, p < .01$. A posteriori analysis (Tukey's HSD) revealed only a statistically significant difference between rotation axis conditions O_aE_x and O_aE_y ($p = .005$).

There was also a statistically significant interaction effect between angular disparity and rotation axis, $F(20, 100) = 3.69, p < .0001$. The slow performance for rotation axis condition O_aE_x relative to the other rotation axis conditions appeared only when the angular disparity between the two views of the object was large (as can be seen in Figure 10A for the match trials only). None of the other interaction effects was reliable, largest $F = 1.88$.

Separate linear regressions for each of the rotation axis conditions (see Table 4) suggested again that the normalization rates were rather similar for all of these conditions, with the O_aE_n condition producing the fastest normalization rate.

Accuracy. ANOVA on the percentage of correct answers revealed a main effect of the angular disparity, $F(4, 20) = 21.78, p < .00001$, and a main effect of rotation axis, $F(5, 25) = 3.75, p < .05$. There was no difference between match and nonmatch trials, $F < 1$.

However, these main effects have to be interpreted with caution because all two-way interactions were statistically significant: Object Type × Angular Disparity, $F(4, 20) = 12.44, p < .0001$, Object Type × Rotation Axis, $F(5, 25) = 3.3, p < .05$, Angular

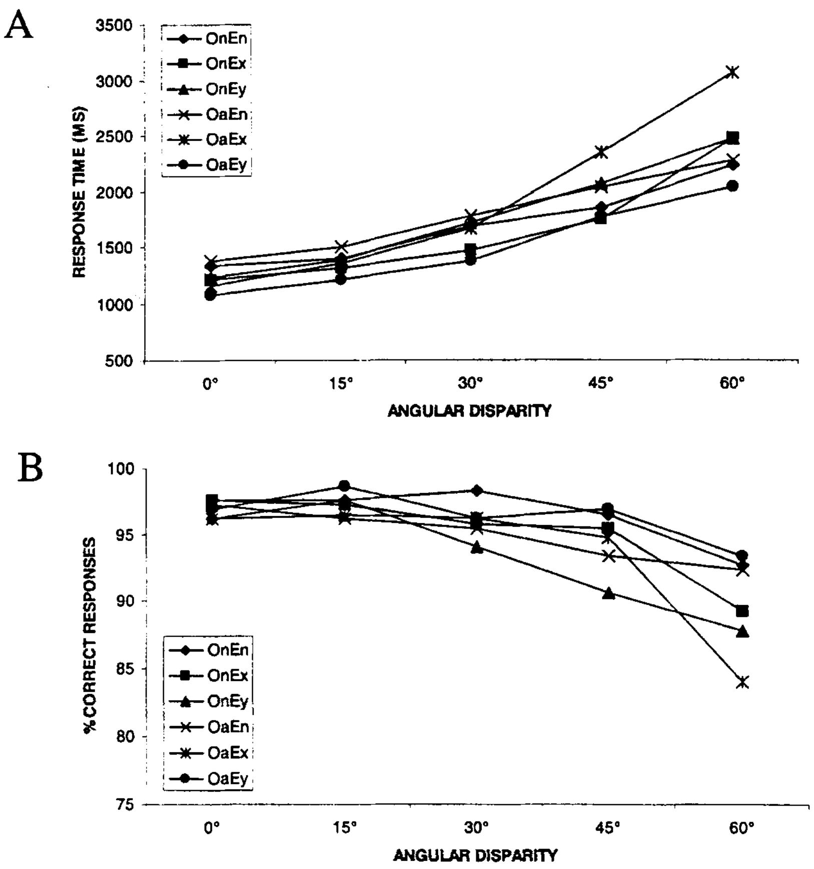

Figure 10. Performance levels in Experiment 4. A: Response times (in milliseconds) for the correct match trials for the six different rotation axis conditions, as a function of the angular disparity between the two simultaneously shown views. B: Percentage of correct responses for the match trials for the six different rotation axis conditions, as a function of the angular disparity between the two simultaneously shown views. O_nE_n = axis of rotation is not aligned with the main axis of the object and is not aligned with one of the axes of the environment; O_nE_x = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's horizontal axis (x-axis); O_nE_y = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's vertical axis (y-axis); O_aE_n = axis of rotation is aligned with the main axis of the object but is not aligned with one of the axes of the environment; O_aE_x = axis of rotation is aligned with the main axis of the object and is aligned with the environment's horizontal axis (x-axis); O_aE_y = axis of rotation is aligned with the main axis of the object and is aligned with the environment's vertical axis (y-axis).

Disparity × Rotation Axis, $F(20, 100) = 2.29, p < .01$ (see Figure 10B for this interaction effect). The three-way interaction did not reach the standard level of statistical significance, $F(20, 100) = 1.44, p = .12$.

Discussion

In Experiment 4, we once again found no evidence for the statement that the viewpoint-dependent performance, when two views of the same object are compared, can be seen as a time-consuming process of aligning 3-D object models of the object in question. This can be inferred from the finding that mental rotation of an object in 3-D space to see what it looks like seems more difficult when the axis of rotation is not aligned with one of the environmental axes or the object's main axis (Parsons, 1995). This pattern was not at all present in the data of Experiment 4. So, we can conclude that in these different experimental set-ups, we found no evidence that our participants mentally rotated a 3-D object description over the specified path of rotation to see whether the two views shown in each trial are of the same object or not. This is an empirical argument in addition to the theoretical problem of inferring the specific rotation that could be used to align the two object descriptions.

Another result that should be noted is the relatively fast normalization rates (around 60°/s) as compared with the rates obtained in the sequential matching task (Experiment 2, around 40°/s). In the literature (for a review, see S. Shepard & D. Metzler, 1988), it was found that the regression of RTs on the angular disparity resulted in larger slopes (i.e., slower normalization rates) for simultaneous matching tasks (our Experiment 4) than for sequential matching tasks (our Experiment 2). Whereas the normalization rates obtained in Experiment 4 are in agreement with those obtained in previous studies with simultaneous matching

Table 4
Linear Regressions of Response Times Against Angular Disparity for Match Trials in Experiment 4

Rotation axis condition	Intercept	Slope	*r*	Normalization rate	% correct responses
O_nE_n	1,535	14.69	.942	68.1°/s	98.47
O_nE_x	1,173	17.36	.904	57.6°/s	91.81
O_nE_y	1,267	19.06	.993	52.5°/s	93.33
O_aE_n	1,520	12.53	.986	79.8°/s	94.03
O_aE_x	987	33.16	.971	30.2°/s	95.28
O_aE_y	1,033	17.42	.966	57.4°/s	93.75

Note. O_nE_n = axis of rotation is not aligned with the main axis of the object and is not aligned with one of the axes of the environment; O_nE_x = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's horizontal axis (*x*-axis); O_nE_y = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's vertical axis (*y*-axis); O_aE_n = axis of rotation is aligned with the main axis of the object but is not aligned with one of the axes of the environment; O_aE_x = axis of rotation is aligned with the main axis of the object and is aligned with the environment's horizontal axis (*x*-axis); O_aE_y = axis of rotation is aligned with the main axis of the object and is aligned with the environment's vertical axis (*y*-axis).

(between 40°/s and 60°/s), this is clearly not the case with the rates obtained in Experiment 2, which are much slower than those in previous studies with sequential matching (between 300°/s and 600°/s). The latter difference suggests that object complexity plays an important role.

However, this in turn opens the possibility that the nature of the objects itself resulted in the absence of any systematic rotational effects in our experiments. Maybe the objects were too complex for our participants to mentally transform the 3-D perceptual descriptions of the viewed objects toward a previously stored view or a simultaneously shown view of that same object. Experiment 5 was designed to show that in a task where one can be more certain that participants use a mental process of aligning 3-D object descriptions, rotation axis effects like those in the mental imagery and spatial reasoning literature can be found. This would establish clearly that it is not the structure of the objects that is responsible for the lack of any systematic rotational effect.

Experiment 5

The main conclusion of Experiments 1–4 was that the viewpoint-dependent performance, classically attributed to some kind of mental rotation of 3-D object representations, is better explained by theories that do not assume this process of alignment over the shortest path in 3-D space. To further strengthen this conclusion, we need to show that, in a task that more clearly requires mental rotation in 3-D space but uses the same 3-D objects, strong effects of axis of rotation can be obtained. Therefore, we used a task where observers had to judge whether an object presented from a given viewpoint could be rotated to another given view of that object by means of the rotation axis that was explicitly presented in both the images. Because one of the two objects had to be rotated toward the other one (over the specified axis of rotation), we could be relatively certain that our observers would solve the task by means of a process of mental rotation. The finding of an effect of axis orientation in this task would reveal that it is not the difficulty of the objects used in our tasks or a lack of sensitivity of our experiments that caused the negative results in Experiments 1–4.

Method

Participants. Participants were 6 members of the Laboratory of Experimental Psychology at the University of Leuven. All of them were naive about the purposes of the experiment, and all of them had normal or corrected-to-normal vision. Two of them had participated in one of the preceding experiments. In Experiment 5, it was not necessary for the participants to be totally naive to the particular views of the objects used.

Apparatus and stimuli. The only difference from Experiments 1–4 was that the objects were now pierced by a rotation axis. This rotation axis was colored red in order to be highly discriminable from the objects themselves (see procedure for how the axes of rotation pierced every object). The Y junctions and arrow junctions at the endpoints of the rotation axis were clearly visible to provide a clear percept of the orientation of the axis in 3-D space (see Willems & Wagemans, 2000, for an experiment demonstrating the importance of this).

Procedure. A trial consisted of two views of the same object presented side by side on the screen. The objects were pierced by the same axis of rotation (i.e., the position of the axis in 3-D space was the same for the two views), but the second view of the object (always shown at the right side of the screen) was a rotated version of the first view (always shown at the left side of the screen). In half of the trials, the axis shown in red could be used as the axis of rotation to align the two objects. In the other half of the trials, the axis shown in red could not be used as the axis of rotation to align the two objects. The axis of rotation needed for alignment was orthogonal to the axis in red and was not aligned with one of the axes of the environment. Participants were asked whether the two object views could be transformed to each other using the axis of rotation that was shown in red (*Z* key) or not (*M* key).

As in Experiments 1–4, the axis of rotation shown in red was positioned in six different ways with respect to the object's main axis and the axes of the environment (i.e., the same six rotation axis conditions as in Experiments 1–4). Trials were not blocked by axis of rotation.

The angular disparity between the first view of the object (right) and the second view (left) was one of four values: −60°, −30°, 30°, or 60° (see Figure 11).

The total number of trials for each participant was 288: 4 Angular Disparities × 2 Trial Types (match/nonmatch) × 6 Rotation Axes × 6 Objects. The duration of Experiment 5 was about 15 min. The rest of the procedure was as in the preceding experiments.

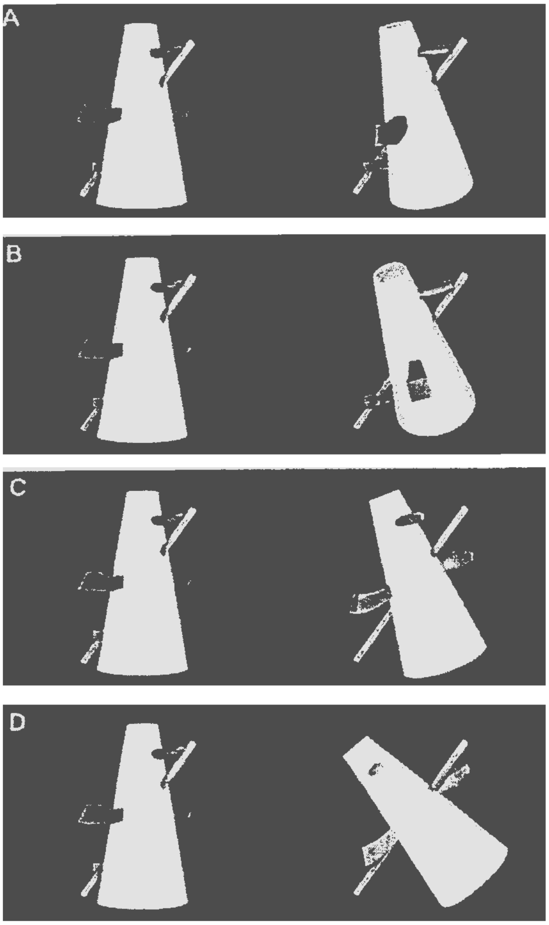

Figure 11. Four trial types used in Experiment 5, illustrated by means of the first object in rotation axis condition O_nE_n (i.e., rotation axis not aligned with main axis or with one of the axes of the environment). A: Match trial with a 30° rotation between the two objects (over the axis of rotation, shown in red in the experiment and reproduced in light gray here). B: Match trial with a 60° rotation between the two objects (over the axis of rotation). C: Nonmatch trial with a 30° rotation between the two objects (over a path of rotation orthogonal to the axis of rotation). D: Nonmatch trial with a 60° rotation between the two objects (over a path of rotation orthogonal to the axis of rotation).

Results

RTs. An ANOVA on mean RTs for each trial type, rotation axis, and angular disparity yielded the following results. In contrast to Experiments 1–4, match trials (2,470 ms) were not reliably faster than nonmatch trials (2,613 ms), $F(1, 5) = 1.81$, $p = .23$. The two other main effects were statistically significant: angular disparity, $F(1, 5) = 13.37$, $p < .05$, and rotation axis, $F(5, 25) = 32.57$, $p < .00001$ (see Figure 12A for the effects of angular disparity and rotation axis for match trials only). A posteriori analysis (Tukey's HSD) revealed that the different rotation axis conditions could be separated in two clusters (O_nE_n, O_nE_x, O_nE_y and O_aE_n, O_aE_x, O_aE_y), which suggests that it was more important for the rotation axis to be aligned with the object's main axis than with one of the environmental axes. Nevertheless, within each cluster, the conditions where the rotation axis was aligned with one of the environmental axes tended to produce faster performance.

There was a significant Trial Type × Angular Disparity effect, $F(1, 5) = 20.87$, $p < .01$. A posteriori analysis (Tukey's HSD) revealed that there was a statistically significant effect of angular disparity for the match trials only ($p < .01$), not for the nonmatch trials ($p = .94$). All the other interaction effects were not reliable (highest $F = 2.25$; lowest $p = .08$).

Accuracy. ANOVA on the proportion of correct answers showed the same main effect of rotation axis, $F(5, 25) = 42.35$, $p < .00001$. However, this time the angular disparity did not yield a reliable main effect, $F(1, 5) = 3.48$, $p = .12$, whereas trial type did, $F(1, 5) = 9.6$, $p < .05$. Furthermore, trial type interacted significantly with rotation axis, $F(5, 25) = 4.7$, $p < .001$, as well as angular disparity, $F(1, 5) = 15.54$, $p < .05$. Angular Disparity × Rotation Axis was unreliable, $F(5, 25) < 1$, but was involved in a reliable three-way interaction with trial type, $F(5, 25) = 5.8$, $p < .01$ (see Figure 12B for the effects of angular disparity and rotation axis for match trials only). A posteriori analysis (Tukey's HSD) revealed that the interaction effect of Angular Disparity × Trial Type was strongest when the axis of rotation was not aligned with the object's main axis (O_nE_n, O_nE_x, and O_nE_y) and was virtually absent when it was aligned with the object's main axis (O_aE_n, O_aE_x, and O_aE_y).

Discussion

These results clearly show that whether the axis of rotation was aligned with one of the environmental axes or with the object's main axis had a profound effect on speed and accuracy of performance. This is in sharp contrast with Experiments 1–4 where the effects of axis of rotation, if present, were small and not systematically related with the axis of rotation being aligned with one of the environmental axes or the object's main axis. The most difficult condition (i.e., largest RTs and lowest accuracy) turned out to be when the axis of rotation was not aligned with the environmental axes or with the object's main axis. Performance improved (i.e., RTs decreased and accuracy increased) when the axis of rotation was aligned with one of the environmental axes or the object's main axis, and the task was easiest when the axis of rotation was aligned with both the object's main axis and one of the environmental axes.

Because this pattern of results is so different from Experiments 1–4, we should ask what induced this remarkable difference. Clearly, an explanation in terms of object properties can be ruled out because the same set of objects was used in both paradigms. An explanation of these contrasting results must thus be found in what participants had to do with the objects in the experiments (i.e., task properties). In Experiment 5, we can be relatively certain that participants needed to mentally rotate a 3-D object description to solve the task. In Experiment 5, profound and systematic effects of the 3-D position of the axis of rotation in space were found. Because in the preceding experiments no such effects were obtained, and we have no other evidence as to the nature of the intermediate process, it seems reasonable to conclude that in all the other tasks participants were most likely not mentally rotating 3-D object descriptions to undo the difference in viewpoint between two views of the same object.

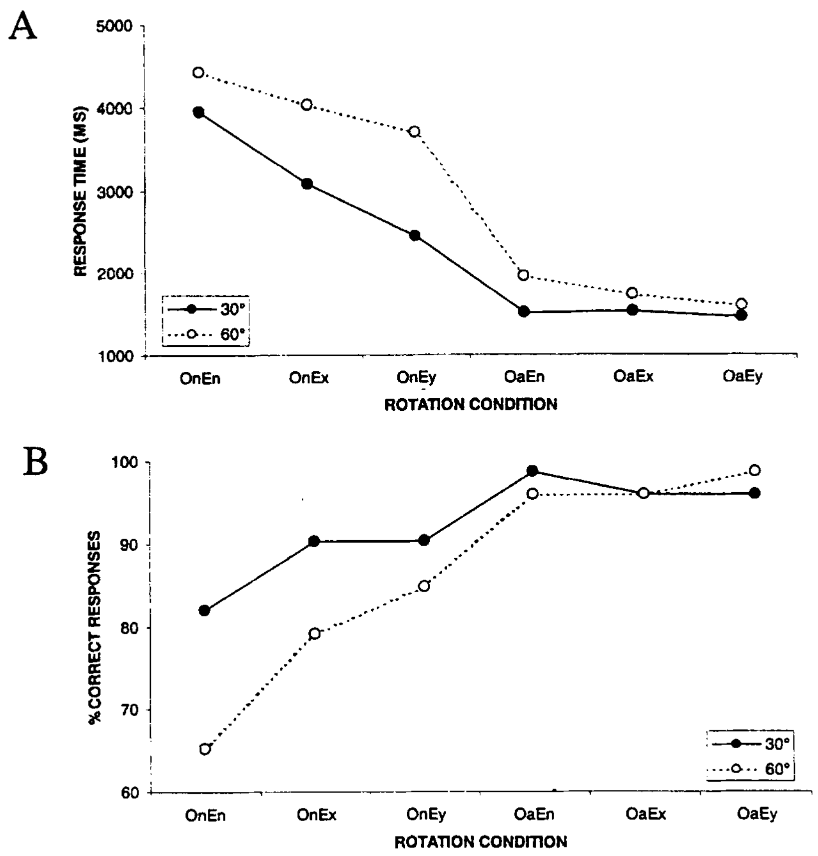

Figure 12. Performance levels in Experiment 5. A: Response times (in milliseconds) for the correct match trials for the six different rotation axis conditions and with an angular disparity between the two objects of 30° or 60°. B: Percentage of correct responses for the correct match trials for the six different rotation axis conditions and with an angular disparity between the two objects of 30° or 60°. O_nE_n = axis of rotation is not aligned with the main axis of the object and is not aligned with one of the axes of the environment; O_nE_x = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's horizontal axis (*x*-axis); O_nE_y = axis of rotation is not aligned with the main axis of the object but is aligned with the environment's vertical axis (*y*-axis); O_aE_n = axis of rotation is aligned with the main axis of the object but is not aligned with one of the axes of the environment; O_aE_x = axis of rotation is aligned with the main axis of the object and is aligned with the environment's horizontal axis (*x*-axis); O_aE_y = axis of rotation is aligned with the main axis of the object and is aligned with the environment's vertical axis (*y*-axis).

What then is the nature of this other process that our participants used to find a match between a shown view of an object and another view of this object stored in memory or between two different views of the same object shown simultaneously? The results of our experiments suggest that it is more appropriate to relate the viewpoint-dependent performance in these other tasks to a process that is based on 2-D transformations (Bülthoff et al., 1995) rather than on a 3-D alignment over the shortest path in 3-D space.

General Discussion

Summary of the Results

In Experiment 1, many different views of each target object were presented in a sequence that suggested apparent rotation of the object about a fixed axis oriented in 3-D space. We found evidence that suggested encoding of viewpoint-specific representations: Test views that were also presented in the target sequence were easier than test views that lay on the same 3-D rotation path but were in between the target views, although their angular separation in depth was only 15°. In Experiments 2 and 4, one view of a target object had to be compared with a large number of test views, with a variable angular separation in depth, either shown simultaneously (Experiment 4) or separated in time (Experiment 2). We found strong evidence for a quasilinear normalization procedure: RTs for match trials increased monotonically with angular separation between test and target views, while the percentage of correct responses decreased.

At the same time, we manipulated the orientation of the axis in space about which target views appeared to rotate in Experiments 1 and 3 and about which target and test views had to be rotated in Experiments 2 and 4 to make them congruent. This manipulation resulted in surprisingly few and small effects, a result which is in agreement with a recent study on identification of everyday 3-D objects in which axis orientation also failed to yield differential depth rotation effects (Newell & Findlay, 1997). In Table 5, the average RTs (top) and average % CRT (bottom) for each of the six rotation axis conditions are ranked from left to right according to task difficulty (i.e., at the left are the largest RTs and the smallest % CRT). As a comparison, the top row indicates the order that could be expected from studies on mental imagery and spatial reasoning (e.g., Pani et al., 1995; Parsons, 1995). Overall, in the first four experiments, there was no indication that axes aligned with the object's main axis or with the environment's cardinal axes were systematically easier than other axes. The ordering in these experiments was not at all as expected, and the differences between the different rotation axis conditions (on which this ordering was based) were most of the time not statistically significant (as indicated by the grouping variables after each rotation axis condition).

Table 5
Overview of the Effect of Rotation Axis for Each Experiment

Experiment and trial type	1 O_nE_n	2 O_nE_x	3 O_nE_y	4 O_aE_n	5 O_aE_x	6 O_aE_y
	Response times					
Experiment 1 (main)						
M + NM	3a	4a	5a	6a	2a	1a
M only	3a	6a	5a	1a	4a	2a
Experiment 1 (additional)						
M + NM	3a	1a	6a	2a	4a	5a
M only	3a	1a	6a	2a	4a	5a
Experiment 2 (main)						
M + NM	3a	5a	4a	2a	6a	1a
M only	3a	4a	5a	2a	1a	6a
Experiment 2 (additional)						
M + NM	2a	3a	4a	1a	6a	5a
M only	2a	1a	6a	3a	4a	5a
Experiment 3						
M + NM	2a	5a	3a	1a	4a	6a
M only	2a	5ab	3abc	1bcd	4cd	6d
Experiment 4						
M + NM	5a	4ab	3ab	1ab	2ab	6b
M only	5a	3ab	4ab	1bc	2bc	6c
Experiment 5						
M + NM	1a	2a	3a	4b	5b	6b
M only	1a	2ab	3b	4c	5c	6c
	Percentage of correct responses					
Experiment 1 (main)						
M + NM	2a	5a	1a	3a	4a	6a
M only	1a	6a	5a	2a	3a	4a
Experiment 1 (additional)						
M + NM	3a	2a	1a	4a	5a	6a
M only	3a	4a	2a	1a	6a	5a
Experiment 2 (main)						
M + NM	2a	4a	6a	1a	3a	5a
M only	1a	2a	4a	3a	6a	5a
Experiment 2 (additional)						
M + NM	6a	5a	1a	4a	3a	2a
M only	1a	6a	5a	3a	4a	2a
Experiment 3						
M + NM	5a	2a	1a	3a	6a	4a
M only	2a	3a	5a	1a	6a	4a
Experiment 4						
M + NM	3a	5a	4ab	2ab	1ab	6b
M only	3a	5a	4ab	2ab	6ab	1b
Experiment 5						
M + NM	1a	2a	3b	5c	6c	4c
M only	1a	2ab	3ab	5b	6b	4b

Note. On top are the six rotation axis conditions in decreasing order of difficulty (from left to right, from most to least difficult), according to the mental imagery and spatial reasoning literature. The next seven pairs of rows are the empirically obtained rank orders in Experiments 1–5 (with different letter symbols indicating significant pairwise difference according to Tukey's HSD comparisons). Response times are ordered from large to small values; percent correct responses are ordered from small to large values. Within each pair, the first row is based on averages across both trial types (match and nonmatch trials [M + NM]), whereas the second row is based on the averages for the match trials only (M only).

This contrast is evidence against the notion that mental operations similar to those in the spatial reasoning and mental imagery tasks are used in the context of our matching tasks. It is not only an argument against mental rotation along the shortest axis in space; it also appears to be in conflict with alternative 3-D mental transformations such as the spin-precession procedure proposed by Parsons (1987). However, to further validate this result, Experiment 5 required participants to follow this procedure of mentally rotating an object representation over a precalculated 3-D transformation path in order to see whether the objects look the same. The bottom rows in the top and bottom of Table 5 clearly show that under these circumstances we indeed found that performance was systematically related to the alignment of the axis of rotation (3-D transformation) with one of the environmental axes or with the object's main axis. Because the same objects were used as in Experiments 1–4, we could be certain that the task difference was responsible for these contrasting results.

Implications for Theories of Object Recognition

Several implications seem to follow from this pattern of results. First, viewpoint-independent theories of object recognition, as they currently exist in the literature (e.g., Biederman, 1987), appear unable to explain these viewpoint-specific effects. For example, all objects were composed of a relatively small number of idealized geon components that should be easily derivable from all views. The way our target and distractor objects were related may not be typical of everyday object identification, but it has two important advantages: The spatial relations that are critical to the discrimination task (left of, right of, at the top, and at the bottom) should be essential in all structural descriptions (but see Discussion of Experiment 1), and they should allow participants to distinguish targets from distractors even when one component is turned out of sight. Nevertheless, even small angular deviations between test and target views induced strong effects on both accuracy and RTs.

Second, but more central to the present study, we were unable to provide any evidence for the use of an incremental mental rotation procedure as the normalization process used to identify depth-rotated 3-D objects in circumstances that lead to view-specific performance. The surprisingly small and unsystematic effects of rotation axis lead us to conclude that this process of normalization is probably not a procedure by which one object description is mentally rotated into the other, along the shortest path in 3-D space, in order to see whether they look the same. R. N. Shepard and colleagues (e.g., R. N. Shepard & Cooper, 1982) have always been careful to avoid the suggestion that mental rotation was used in object recognition, but many other researchers have been intrigued by the strikingly similar linear trends of angular deviation on 2-D object recognition (e.g., Jolicoeur, 1985, 1988; Tarr & Pinker, 1989). Even though fundamental problems arise when viewpoint-specific effects on 3-D object recognition are attributed to mental rotation along the shortest path in space, some theorists have been tempted to use the notion at least metaphorically (e.g., Tarr, 1995; Ullman, 1989). There is now mounting evidence against the view that mental rotation is an appropriate description of the normalization process used for 3-D object identification, also in a metaphorical sense. Related evidence with a completely different experimental paradigm has recently been obtained that rules out mental rotation even for the case of identification of plane-rotated objects (Jolicoeur et al., 1998).

Third, and more positively, we have obtained evidence that is compatible only with a multiple-views approach to object recognition. Fortunately, within the multiple-views approach, several computational models of viewpoint-specific object recognition have been proposed that include well-elaborated accounts of normalization processes that do not imply 3-D rotations. Three-dimensional objects can be recognized from different viewpoints if the viewpoint-specific perceptual descriptions can be matched against a 2-D view derived from the stored views, either by the linear combination of the stored views (Ullman, 1996; Ullman & Basri, 1991) or by the interpolation between the stored views (Edelman, 1999; Poggio & Edelman, 1990). Several results from Experiment 3 were more congruent with the interpolation model than with the linear combination model, but it should be stressed that the two models are computationally very similar. An essential property of both of these theories is that the 2-D perceptual descriptions provide sufficient information for 3-D object recognition: New views can be computed without regard to 3-D transformations in space. If an effect of specific 3-D paths were to be found, it would be opposite: New views would be somewhat easier to handle in these models if the stored views constituted a more heterogeneous sample. In fact, this factor could explain why the condition with no axis alignment (i.e., O_nE_n) tended to yield the fastest normalization rates. It is worth exploring this effect further in future research.

Relations to Mental Imagery Results

We used, in addition to the viewpoint-dependent results, another criterion for deciding whether or not mental rotation was involved in the process of recognizing objects from previously unseen views. If mental rotation were used, an effect of axis of rotation should have been apparent in the performance of our observers. Because the absence of this effect by itself was not enough (several reasons seem to be valid for explaining this absence), we used the same objects in a task where we could be sure that mental imagery was involved. If the effect were found with this kind of task, it would be safe to conclude that something else was going on in the tasks where the effect was not found or where the effect was not in the expected direction. Because in our set of experiments such a differentiation was found, it seems worthwhile to examine the tasks carefully to try to explain this difference in performance.

Table 6 presents simple computational process models of the tasks that are commonly used in the context of object recognition on the one hand and mental imagery on the other hand. Note that the order in which each component is presented does not necessarily coincide with the order in which they are computed because several of these components can be computed in parallel. We make a distinction between a perceptual description (PD), which is derived from the retinal stimulation, and a nonperceptual description (NPD), which is not derived from the retinal stimulation but is stored in memory or derived from another intermediate representation. In the sequential matching task, a perceptual description (PD_i) of the object has to be computed along with a transformation (tr) that can be used to align the stored object description (NPD) with the perceptual description to find a match between the perceptual description and the transformed object description. Whether the transformation is applied to the PD_i or the NPD is, at the moment, of no importance. In the simultaneous matching task, two perceptual descriptions have to be calculated (PD_i and PD_j), again with a transformation (tr) that is able to align the two perceptual descriptions to find a match between the two. The knowledge of which PD is transformed is of no importance for the underlying argument.

For the tasks used in the mental imagery literature (e.g., Parsons, 1987), a perceptual description and a transformation are computed (how the transformation is computed can differ considerably from the way this happens in the tasks used in object recognition but this is of no concern to our argument). Based on these two, an object description is computed that is stored in memory (NPD) to compare it with the new perceptual description (PD_j) that is computed next. In our Experiment 5 (which we could call a simultaneous imagery task), two perceptual descriptions have to be calculated, along with the transformation that is implied by the shown rotation axis. One of the two perceptual descriptions is then transformed using the computed transformation to find a match between the two (note the computational resemblance with the simultaneous matching task, despite the difference in performance).

Now, if we assume that the same kind of perceptual and stored object descriptions are used in the different tasks (which is likely because the same set of objects was used), the only factor that can

Table 6
Computational Process Models of the Different Tasks

Task	Required components	Decision to be made
Sequential matching task	PD_i, NPD, tr	$\mathrm{tr(NPD)} \leftarrow ? \rightarrow \mathrm{PD}_i$ or $\mathrm{tr}(\mathrm{PD}_i) \leftarrow ? \rightarrow \mathrm{NPD}$
Simultaneous matching task	PD_i, PD_j, tr	$\mathrm{tr}(\mathrm{PD}_j) \leftarrow ? \rightarrow \mathrm{PD}_i$ or $\mathrm{tr}(\mathrm{PD}_i) \leftarrow ? \rightarrow \mathrm{PD}_j$
Sequential imagery task	PD_i, tr, $\mathrm{NPD} = \mathrm{tr}(\mathrm{PD}_i)$	$\mathrm{NPD} \leftarrow ? \rightarrow \mathrm{PD}_j$
Simultaneous imagery task	PD_i, PD_j, tr	$\mathrm{tr}(\mathrm{PD}_j) \leftarrow ? \rightarrow \mathrm{PD}_i$ or $\mathrm{tr}(\mathrm{PD}_i) \leftarrow ? \rightarrow \mathrm{PD}_j$

Note. PD_i = first perceptual description; NPD = nonperceptual description; tr = transformation; PD_j = second perceptual description.

explain the differences in performance is the assumed transformation. A clear distinction has to be made between the transformation as a mathematical entity and the transformation as a psychological process. Although the transformation in the mathematical sense is the same for the two kinds of tasks, the psychological process (i.e., how this transformation is implemented in the brain) is supposed to be different. In the first group of tasks, the psychological transformation or normalization process can be performed automatically and fast, and it is independent of the specific transformation in question. In contrast, for the other group of tasks, it is a difficult and effortful process that is mediated by other factors such as the alignment of the rotation axis with common frames of reference (object axes or environmental axes).

What can be the reason for applying different transformations (as implemented in the brain) in the two groups of tasks? Of course, the weight carried by the transformational component in the underlying process models differs between the two groups: In the first group of tasks, the specific transformation is of little importance (it is the resulting representation that matters most), whereas in the second group the transformation has to be performed exactly in order to solve the task. So, it seems to us that the importance of the transformation is a valid reason for not performing the same transformation for the two groups of tasks. Based on these considerations, we think it is safe to conclude that the mechanisms underlying the recognition of objects from unseen viewpoints do not rely on the mental rotation of stored object descriptions (i.e., comparing the incoming perceptual description with the description obtained by mentally rotating the stored object description).

This finding also corroborates another argument that we have made recently in the context of a pattern-matching task, with dot patterns as stimuli instead of multicomponent solid objects and with affine transformations instead of 3-D rotations. In one study (Wagemans, Van Gool, Lamote, & Foster, 2000), we have shown that minimal information is sufficient to determine affine shape equivalence and that qualitatively or quasi-invariant properties (such as convexity or collinearity) play a more important role in task performance than mental transformations. However, we also obtained systematic effects of the affine transformation parameters on task performance that would lead many people to infer psychological normalization processes that mentally undo the geometric transformation during the matching task. In another study (Wagemans, Van Gool, & Lamote, 1996), we have provided evidence that effects of transformation parameters can reflect measurement errors in the visual system itself (even when it is computing mathematical invariants) and thus need not be taken as evidence for mental transformations (see also Wagemans, Lamote, & Van Gool, 1997, for a similar point with respect to projective shape equivalence of polygons).

In subsequent research (Van Campenhout, Wagemans, Kyllingsbaek, Bundesen, & Larsen, 1998), we were able to induce much more classic "mental transformation" results (with linear regressions of RTs against 3-D angle producing much better fits) by introducing only a small number of changes in the experimental paradigm. Recent research in our laboratory has dissociated mental transformation procedures and invariants-based procedures with neuroimaging techniques (Vanrie, Béatse, et al., in press; Vanrie, Willems, et al., in press). Also relevant in this context is a neuropsychological study by Farah and Hammond (1988). They described a patient whose mental rotation ability was heavily impaired, whereas his ability to recognize inverted pictures of objects was spared. This suggests that the process of mental rotation is not needed to recognize objects from noncanonical viewpoints.

In the same spirit, Niall (1997) has argued that mental rotation is neither necessary nor sufficient to explain changes in RTs for the simultaneous comparison of planar shapes depicted in depth. When two planar shapes were depicted as separated by a small and fixed angular slant difference in depth and then depicted as tilted in depth together (in so-called *tandem rotation*), RTs varied nearly linearly with the magnitude of the tilt in depth, even with constant angular differences. This result suggests that comparison of shapes in different 3-D orientations can vary as a function of slant and tilt without implying the use of mental transformations in 3-D space to undo the transformations. Niall has even argued that the classic mental rotation functions for Shepard's block figures can be predicted from a specific combination of the 2-D compression effects on the separate surface patches (when the block figures are viewed from different angles). Although we do not want to push this idea too far, it does make the point that strong linear effects of a parameter change on a 3-D path (as we obtained in Experiment 2) are perfectly compatible with normalization procedures that work on 2-D views only (much like in the computational models by Poggio & Edelman, 1990, and by Ullman & Basri, 1991). Related arguments about the role of 2-D similarity of views of 3-D objects in determining performance differences have been made before in the context of picture matching of familiar objects (Lawson & Humphreys, 1996; Lawson, Humphreys, & Watson, 1994; for a review, see Lawson, 1999). It will continue to be a challenge for future research to provide a metric for these similarity effects in object representation and object recognition, although some progress has already been made (e.g., Edelman, 1995a, 1995b, 1997, 1999; Edelman & Duvdevani-Bar, 1997).

Relations to Other Recent Work

It is useful to mention two lines of research that are also relevant to the present work: physiological evidence for the role of multiple views in object recognition by monkeys, and the role of an induced motion path in the context of viewpoint effects on object recognition.

Physiological evidence for multiple views. In a series of experiments, combining psychophysical with electrophysiological techniques, Logothetis and his colleagues (e.g., Logothetis & Pauls, 1995; Logothetis, Pauls, Bülthoff, & Poggio, 1994; Logothetis, Pauls, & Poggio, 1995) provided evidence for viewer-centered object representations in the primate (see Logothetis & Sheinberg, 1996, for a review). In an object recognition task, previously used to investigate subordinate-level recognition by humans (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992), rhesus monkeys showed the same pattern of results as humans. When the stimuli were rotated away from the learned viewpoint, recognition became increasingly difficult. However, when the monkeys had learned as few as three views of the object, recognition became invariant over all viewpoints of that object. Moreover, single-cell recordings in inferotemporal cortex (IT) with the same group of trained monkeys revealed that some cells were selectively responsive to particular views of that object, even when the object could be recognized from all viewpoints (see also Perrett et al., 1991). Many results obtained by these researchers are in accordance with theoretical studies that describe view-based models of 3-D object recognition, which do not rely on the mental transformation of 3-D object models to obtain viewpoint invariance (Edelman, 1999; Vetter, Hurlbert, & Poggio, 1995; Perrett, Oram, & Ashbridge, 1998).

Role of object rotation. Kourtzi and Shiffrar (1997) showed that the matching of two identical views of 2-D objects (adapted from Tarr & Pinker, 1989) could become faster when preceded by two successive views of the target object separated by a variable angle of planar rotation. The degree of primed matching depended strongly on the rotation angle between the two prime views, the orientation of the target views relative to the prime views, and the spatiotemporal relations between the prime views. When the target views had been presented previously as prime views (*seen views* in our terminology), significant priming was obtained in all conditions. With novel target views, an interesting interaction occurred. When the spatiotemporal conditions were such that they did not allow apparent motion to be seen between the successive prime views, significant priming was obtained only when the orientation difference between prime and target views was small. When the spatiotemporal conditions favored apparent motion between the successive prime views, priming also occurred for larger orientation differences but only when the target views were intermediate between the two views in the apparent motion sequence; when the target views were outside the rotation path, priming even disappeared for the small angular deviations (except when they could be taken as extrapolations of the perceived rotation, on the same path and in the same direction).

Thus, significant priming across views is possible according to these results, and the generalization fields around known object views (Bülthoff & Edelman, 1992) can even be expanded for views on an apparent rotation path. These findings are in line with earlier empirical studies on picture matching of familiar objects (e.g., Lawson & Humphreys, 1996; Lawson et al., 1994) as well as with theoretical studies (e.g., Edelman & Weinshall, 1991; Wallis & Bülthoff, 1999).

These findings are also in line with recent research on representational momentum (Munger, Solberg, Horrocks, & Preston, 1999), which shows that the apparent rotation of an object seems to induce a dynamic representation of the object (i.e., a representation that also involves the motion of the object). In a subsequent study (Munger, Solberg, & Horrocks, 1999), it was even shown that these dynamic representations are the same representations underlying the matching of two views of an object from different viewpoints.

Although our presentation conditions also favored the perception of apparent rotation between the successive views in the target sequence of views (based on our own observations and those of the participants), our results revealed rather strict viewpoint specificity: In Experiment 1, test views that were only 15° in between previously seen views and on the induced motion path tended to be responded to more slowly than the seen views themselves; in Experiment 3, view differences that were not on the induced motion path but were orthogonal to it tended to require even more time. One could speculate that there is a fundamental difference between 2-D patterns and rotations in the plane, as used by Kourtzi and Shiffrar (1997), and 3-D objects and rotations in depth, as we used here, but more recent results by Kourtzi and Shiffrar (1999) indicate that similar primed matching can be obtained with 3-D objects rotating in depth, even when the prime and test views have different parts visible. Future research will have to determine which of the possible differences between the two research paradigms is the strongest source of this discrepancy (e.g., object recognition vs. primed matching, the complexity of the objects, the type of target–distractor differences, etc.; see also Srinivas & Schwoebel, 1998).

Toward a Synthesis

When the present results are taken at face value, somewhat more detached from the debate between viewpoint-independent and -dependent theories of object recognition, they have implications that may lead to an interesting synthesis when followed up by future research. What are these basic results and their implications?

First, when objects are presented as a series of views, seen views are identified faster as images of a target object than views that are only 15° in between and on the same apparent rotation path. This may be due to a short-term memory advantage similar to other priming effects found in a large number of studies. It is not compatible with a viewpoint-independent theory of object recognition that situates priming effects at the level of GSDs instead of 2-D images or views (e.g., Biederman & Cooper, 1991); it is also at odds with viewpoint-dependent theories that should allow at least these small generalizations to avoid the storage of all possible views.

Second, the absence of rotation axis effects, even in variants of our experiments involving mere handedness discrimination, suggests that our participants were probably not using a mental rotation along the shortest path in space to compare the different object views that they had available. In fact, this is not surprising at all if one knows how poor people are, in general, at finding the orientation of a rotation axis based on perceived object motion (Pani et al., 1995) or at establishing a unique axis and angle that would rotate an object from one view to another (Parsons, 1995). More probably, observers resort to some sort of sloppier, more qualitative comparison based on a small number of salient object characteristics.

The specification of these characteristics is a major task for theories of object recognition. It is at this level that Biederman's (1987) RBC theory has made an important contribution. For example (see Figure 1 and Figure A1, later), Object 1 can be distinguished from Distractor 1B because the small component with a straight edge (Component 1) is near the bottom part of the large component in the target and near the top part in the distractor, whereas the opposite is true for the small component with a curved edge (Component 5). This is relatively easy to see and remember because the large component itself has a clear top–bottom asymmetry based on its expanding cross-section (Component 10). For Object 3, which has no such asymmetrical large component (Component 1), relevant distinguishing characteristics are based on the direction of curvature of some of the small components (e.g., Component 35 pointing away from the object center in the target and pointing to the center in Distractor 3B). Even though these characteristics remain invariant with viewpoint changes, they will be easier to recover from some views than from others, depending on their relative orientation to the viewer. In other words, although these qualitative properties are mathematically invariant, one should not forget that the visual system must acquire information about them based on a limited view. This may cause measurement error (see Wagemans et al., 1996, for the case of affine invariants based on point coordinates). In sum, it may be harder to derive the essential aspects of the perceptual description from some views than from others (see Van Lier & Wagemans, 1999, for the case of object symmetry based on skewed symmetry).

This account may also explain why systematic effects are obtained from the angular difference between two object views to be matched, a third basic result in our experiments. Due to 3-D object rotation, some perceptual features in the 2-D image change such that the similarity of the views will decrease and the required time and effort for a comparison will increase. It is quite likely that the comparison process works in a noncontinuous similarity space (based on qualitatively or quasi-invariant features such as those that determine the different aspects in an aspect graph; see Koenderink, 1984; Van Effelterre, 1994) and thus would lead to rather nonmonotonic effects on a trial-by-trial basis. However, across trials (and certainly across observers), more monotonic effects may occur because of an averaging process (see also Leek, 1998). When phrased in these terms, it becomes clear that the nonaccidental properties that underlie Biederman's (1987) RBC theory may also be the salient perceptual features that are stored in a multiple-views representation such as an aspect graph. We believe it would be more fruitful for future research to address this issue than to stick to a categorical distinction between viewpoint-independent and -dependent theories of object recognition. Of equal importance will be the question of where viewpoint-specific performance differences arise, at the level of deriving perceptual descriptions or at the level of matching perceptual with stored descriptions. Clever new experimental paradigms will have to be designed to disentangle these two possibilities.

References

Autodesk, Inc. (1993). *Autodesk 3D Studio* (Release 3). Sausolito, CA: Author.

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

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39,* 2885-2899.

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23,* 393-419.

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

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception, 28,* 575-599.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, 89,* 60-64.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex, 5,* 247-260.

Cedrus Corp. (1997). *Superlab Pro for Windows* (Version 1.04). Phoenix, AZ: Author.

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

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

Corballis, M. C., & McLaren, R. (1982). Interactions between perceived and imagined rotation. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 215-224.

Cutzu, F., & Edelman, S. (1994). Canonical views in object representation and recognition. *Vision Research, 34,* 3037-3056.

Edelman, S. (1995a). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics, 72,* 207-220.

Edelman, S. (1995b). Representation of similarity in three-dimensional object discrimination. *Neural Computation, 7,* 408-423.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences, 1,* 296-304.

Edelman, S. (1999). *Representation and recognition in vision.* Cambridge, MA: MIT Press/Bradford Books.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3-D objects. *Vision Research, 32,* 2385-2400.

Edelman, S., & Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London, B352,* 1191-1202.

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics, 64,* 209-219.

Farah, M. G., & Hammond, K. M. (1988). Mental rotation and orientation-invariant object recognition: Dissociable processes. *Cognition, 29,* 29-46.

Gibson, B. S., & Peterson, M. A. (1994). Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 299-316.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1511-1521.

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

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

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.), *Attention and performance: Vol. XVI. Information integration in perception and communication* (pp. 93-121). Cambridge, MA: MIT Press.

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

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology, 42,* 461-478.

Jolicoeur, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind & Language, 5,* 387-410.

Jolicoeur, P., & Cavanagh, P. (1992). Mental rotation, physical rotation, and surface media. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 371-384.

Jolicoeur, P., Corballis, M. C., & Lawson, R. (1998). The influence of perceived rotary motion on the recognition of rotated objects. *Psychonomic Bulletin & Review, 5,* 140-146.

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.

Koenderink, J. J. (1984). The internal representation of solid shape and visual exploration. In L. Spillmann & B. R. Wooten (Eds.), *Sensory experience, adaptation, and perception* (pp. 123-143). Hillsdale, NJ: Erlbaum.

Kourtzi, Z., & Shiffrar, M. (1997). One-shot view invariance in a moving world. *Psychological Science, 8,* 461–466.

Kourtzi, Z., & Shiffrar, M. (1999). The visual representation of three-dimensional, rotating objects. *Acta Psychologica, 102,* 265–292.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica, 102,* 221–245.

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 395–416.

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint-specific recognition procedures. *Perception, 23,* 595–614.

Leek, E. C. (1998). The analysis of orientation-dependent time costs in visual recognition. *Perception, 27,* 803–816.

Leeuwenberg, E., & van der Helm, P. (1991). Unity and variety in visual form. *Perception, 20,* 595–622.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research, 35,* 549–568.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex, 3,* 270–288.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology, 4,* 401–414.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5,* 552–563.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621.

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

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, B200,* 269–294.

Munger, M. P., Solberg, J. L., & Horrocks, K. K. (1999). The relationship between mental rotation and representational momentum. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25,* 1557–1568.

Munger, M. P., Solberg, J. L., Horrocks, K. K., & Preston, A. S. (1999). Representational momentum for rotations in depth: Effects of shading and axis. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25,* 157–171.

Newell, F. N., & Findlay, J. M. (1997). The effect of depth rotation on object identification. *Perception, 26,* 1231–1257.

Niall, K. K. (1997). "Mental rotation," pictured rotation, and tandem rotation in depth. *Acta Psychologica, 95,* 31–83.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention & performance* (Vol. 9, pp. 135–151). Hillsdale, NJ: Erlbaum.

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

Pani, J. R., William, C. T., & Shippey, G. T. (1995). Determinants of the perception of rotational motion: Orientation of the motion to the object and to the environment. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1441–1456.

Parsons, L. M. (1987). Visual discrimination of abstract mirror-reflected 3-D objects at many orientations. *Perception & Psychophysics, 42,* 49–59.

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.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without transformations. *Cognition, 67,* 111–145.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86,* 159–173.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263–266.

Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science, 247,* 978–982.

Saiki, J., & Hummel, J. E. (1998). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 227–251.

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

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

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

Shiffrar, 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.

Shyi, G. C.-W., Goldstone, R. L., & Hummel, J. E. (1998, November). *Computing representations for bound and unbound 3-D objects matching.* Paper presented at the Fifth Annual Workshop on Object Perception and Memory (OPAM), Philadelphia, PA.

Srinivas, K., & Schwoebel, J. (1998). Generalization to novel views from view combination. *Memory & Cognition, 26,* 768–779.

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., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8,* 282–289.

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., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1,* 259–331.

Tjan, B. S., & Legge, G. E. (1998). The viewpoint complexity of an object-recognition task. *Vision Research, 38,* 2335–2350.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193–254.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition.* Cambridge, MA: MIT Press/Bradford Books.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition, 67,* 21–44.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-13,* 992–1006.

Van Campenhout, S., Wagemans, J., Kyllingsbaek, S., Bundesen, C., & Larsen, A. (1998, August). *Rotating simple two-dimensional figures through space along the shortest path.* Poster session presented at the 21st European Conference on Visual Perception, Oxford, England. Abstract published in *Perception, 27* (Suppl.), 125a.

Van Effelterre, T. (1994). Aspect graphs for visual recognition of three-dimensional objects. *Perception, 23,* 563–582.

Van Lier, R., & Wagemans, J. (1998). Effects of physical connectivity on the representational unity of multi-part configurations. *Cognition, 69,* B1–B9.

Van Lier, R., & Wagemans, J. (1999). From images to objects: Global and local completions of self-occluded parts. *Journal of Experimental Psychology: Human Perception & Performance, 25,* 1721–1741.

Vanrie, J., Béatse, E., Wagemans, J., Sunaert, S., & Van Hecke, P. (in press). Mental rotation versus invariant features in object perception from different viewpoints: An fMRI study. *Neuropsychologia.*

Vanrie, J., Willems, B., & Wagemans, J. (in press). Multiple routes to object matching from different viewpoints: Mental rotation versus invariant features. *Perception.*

Vetter, T., Hurlbert, A., & Poggio, T. (1995). View-based models of 3-D object recognition: Invariance to imaging transformations. *Cerebral Cortex, 5,* 261–269.

Wagemans, J. (1993). Skewed symmetry: A nonaccidental property used to perceive visual forms. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 364–380.

Wagemans, J., Lamote, C., & Van Gool, L. (1997). Shape equivalence under perspective and projective transformations. *Psychonomic Bulletin & Review, 4,* 248–253.

Wagemans, J., Van Gool, L., & Lamote, C. (1996). The visual system's measurement of invariants need not itself be invariant. *Psychological Science, 7,* 232–236.

Wagemans, J., Van Gool, L., Lamote, C., & Foster, D. H. (2000). Minimal information to determine affine shape equivalence. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 443–468.

Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences, 3,* 22–31.

Willems, B., & Wagemans, J. (2000). The viewpoint-dependency of veridicality: Psychophysics and modelling. *Vision Research, 40,* 3017–3027.

Appendix

Details of Stimulus Objects: All Components and Their Spatial Relations

As outlined in the text, we created eight novel stimulus objects consisting of five components each. We used Biederman's (1987) geon theory to select a representative set of components. For each of the eight stimulus objects, a different large geon was used as the object's major part and four other small geons were attached onto this basis geon, each at one of four possible locations (at different sides of the major part). Two of the small geons were attached somewhere halfway along the elongation axis, at opposite sides of the major part (let us call these *left* and *right*). The other two components were attached near the ends of the basis geon (i.e., one near the top and one near the bottom) and at opposite sides (which could be indicated as *front* and *back*).

In Figure A1, we show schematic drawings of each of Biederman's (1987) 36 possible geons, according to the description system introduced by Leeuwenberg and van der Helm (1991). The numbers in this scheme are used to indicate which components were used as object parts in our set of stimulus objects.

In Table A1, all versions of all stimulus objects are represented by a row of numbers, each indicating a component in a particular role, either as major part or as one of four minor parts: left (L), right (R), top (T), or bottom (B). Four rows are present for each stimulus object: The first one is for the target version and the three following ones are for the three distractor versions. Table A1 clearly indicates how distractors are derived from targets: Distractor A by switching L and R components; Distractor B by switching T and B components; and Distractor C by both switching L and R components and T and B components (see also Figure 2). Distractors B and C were not mere mirror versions of the target objects, because there were always some spatial relations to uniquely define the top and bottom of each object (either because it had a clear bottom side that was larger than the top side, as in Objects 1 and 4, or because at least one of the minor parts had a curved axis, which was the case for all objects).

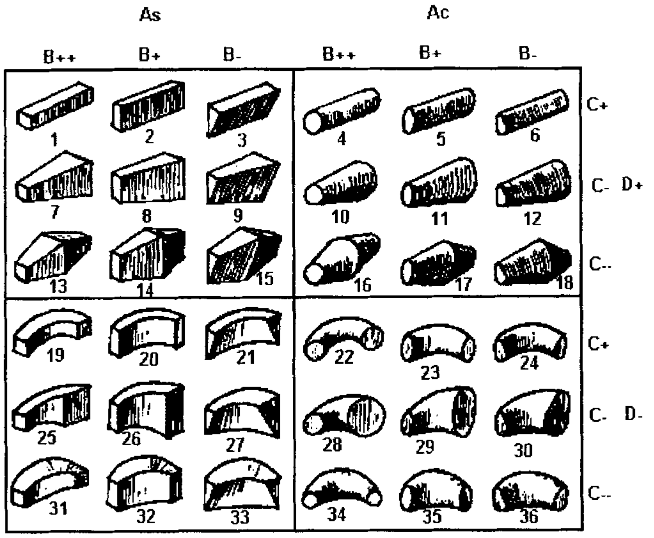

Figure A1. Labeling system used to describe all geons in Biederman's (1987) recognition-by-components (RBC) theory. Columns A indicate whether the cross-section has straight (As) or curved (Ac) edges and columns B describe the symmetry type of the cross-section, which can be either rotational and reflectional (+ +), only reflectional (+), or asymmetrical (−). Rows D indicate whether the axis is straight (+) or curved (−), and rows C describe the size variation when the cross-section is swept about the axis, which can be constant (+), expanding (−), or expanding and contracting (− −). The numbers next to each basis geon are used in Table A1. From "Unity and Variety in Visual Form," by E. Leeuwenberg and P. van der Helm, 1991, *Perception, 20,* pp. 595–622, Figure 15, p. 612. Copyright 1991 by Pion Ltd., London. Adapted with permission.

Table A1
All Stimulus Objects With Their Components and Their Spatial Relations

Object and distractor	Major component	Minor components			
		Left	Right	Top	Bottom
Object 1	10	26	36	5	1
Distractor 1A	10	36	26	5	1
Distractor 1B	10	26	36	1	5
Distractor 1C	10	36	26	1	5
Object 2	4	10	13	30	31
Distractor 2A	4	13	10	30	31
Distractor 2B	4	10	13	31	30
Distractor 2C	4	13	10	31	30
Object 3	1	27	7	6	35
Distractor 3A	1	7	27	6	35
Distractor 3B	1	27	7	35	6
Distractor 3C	1	7	27	35	6
Object 4	7	16	20	15	28
Distractor 4A	7	20	16	15	28
Distractor 4B	7	16	20	28	15
Distractor 4C	7	20	16	28	15
Object 5	13	9	23	33	11
Distractor 5A	13	23	9	33	11
Distractor 5B	13	9	23	11	33
Distractor 5C	13	23	9	11	33
Object 6	16	25	29	3	18
Distractor 6A	16	29	25	3	18
Distractor 6B	16	25	29	18	3
Distractor 6C	16	29	25	18	3
Object 7	7(2×)	17	7	26	22
Distractor 7A	7(2×)	7	17	26	22
Distractor 7B	7(2×)	17	7	22	26
Distractor 7C	7(2×)	7	17	22	26
Object 8	10(2×)	19	35	2	12
Distractor 8A	10(2×)	35	19	2	12
Distractor 8B	10(2×)	19	35	12	2
Distractor 8C	10(2×)	35	19	12	2

Note. Objects 1–6 were used as experimental objects; Objects 7 and 8 were used for practice only.

Received April 3, 1998
Revision received August 9, 2000
Accepted December 11, 2000 ■