Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 44 (2004) 1717–1730

Vision Research

www.elsevier.com/locate/visres

Rotation direction affects object recognition

Quoc C. Vuong *, Michael J. Tarr

Department of Cognitive and Linguistic Sciences, Brown University, 190 Thayer Street, Providence, Rhode Island, 02912, USA

Received 28 March 2003; received in revised form 8 December 2003

Abstract

What role does dynamic information play in object recognition? To address this question, we probed observers' memory for novel objects rotating in depth. Irrespective of object discriminability, performance was affected by an object's rotation direction. This effect was obtained despite the same shape information and views being shown for different rotation directions. This direction effect was eliminated when either static images or animations that did not depict globally coherent rotation were used. Overall, these results suggest that dynamic information, that is, the spatiotemporal ordering of object views, provides information independent of shape or view information to a recognition system.

Keywords: Object perception; Object recognition; Motion; Rotation; Spatiotemporal

1. Introduction

Visual perception is a dynamic process that incorporates as much of the sensory input as possible for the task at hand (Gibson, 1979). Thus, beyond shape, motion is a natural source of information for recognizing objects. Imagine seeing an object rotate in depth. As it rotates, certain features become visible while others become occluded. This example presents an interesting case for studying object recognition because of the dual nature of the information available for recognition. The rotating object has a certain appearance (in terms of visible features, surfaces, and parts) to the observer at each moment in time. We refer to this instantaneous appearance of an object at a given moment in time as a *view* (e.g., Tarr & Bülthoff, 1998). At the same time, the object's appearance changes in a regular and predictable manner (e.g., Stone, 1998, 1999). We refer to this continuous spatiotemporal sequence of views as *dynamic information*. [1] Within the object-recognition literature a good deal of attention has been paid to how the information in one or more views affects object recognition (Biederman, 1987; Biederman & Gerhardstein, 1993; Bülthoff & Edelman, 1992; Tarr, 1995; Tarr & Pinker, 1989).

Recently, attention has focused on *how* dynamic information afforded by object motion may be used to recognize faces and objects (e.g., Knappmeyer, Thornton, & Bülthoff, 2003; Lander & Bruce, 2000; Liu & Cooper, 2003; Mitsumatsu & Yokosawa, 2003; Pike, Kemp, Towell, & Phillips, 1997; Stone, 1998, 1999; Thornton & Kourtzi, 2002). Independent of different approaches to object recognition (e.g., Biederman, 1987; Tarr & Bülthoff, 1998), we can ask what role(s) dynamic information might play in recognition (for a review of the role of motion in face recognition see, O'Toole, Roark, & Abdi, 2002). The possibilities are surprisingly (although probably not to Gibson!) rich and include:

- Object motion may enhance the recovery of information about shape (e.g., Ullman, 1979).
- Object motion may provide observers with more views (regardless of their temporal order) relative to a stationary object (e.g., Pike et al., 1997).
- Object motion may enhance an observer's ability to find meaningful edges and segment a scene into discrete objects (e.g., Rubin & Albert, 2001).

* Corresponding author. Address: Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tuebingen, Germany. Tel.: +49-7071-601-735; fax: +49-7071-601-616.

E-mail address: quoc.vuong@tuebingen.mpg.de (Q.C. Vuong).

[1] Motion, and motion perception, is a continuous process unfolding over time. However, a given motion may be decomposed into an ordered view sequence, where each view captures the instantaneous appearance of the moving object at a particular point in time.

0042-6989/$ - see front matter
doi:10.1016/j.visres.2004.02.002

- Object motion may specify a physical process and, as such, bias how different views are encoded in visual memory (e.g., Kelly & Freyd, 1987).
- Object motion may provide information about how image features change over time (e.g., Stone, 1998, 1999).
- Object motion may allow observers to anticipate views of objects (e.g., Mitsumatsu & Yokosawa, 2003).

These possibilities are not mutually exclusive. Any or all of these may confer an advantage for dynamic displays over static ones. Thus, although it is often the case that dynamic information improves recognition, there is not always a single definitive explanation for this improvement. At the same time, these possibilities suggest that the visual system derives information from object motion that specifies more than 3D shape or 2D views. Thus, our goal in the present study was twofold. First, we examined whether the visual system recovers only spatial information (3D shape, 2D views, image features, and so on) from object motion or whether the visual system *also* recovers dynamic information—that is, how spatial information unfolds over time. Second, we examined whether the visual system recovers dynamic information *by default* or only under restricted circumstances, for examples, when objects have similar shapes and parts or when observers have time to learn characteristic motion.

Building on the work of Stone (1998, 1999) and Liu and Cooper (2003), we examined whether the direction of rigid depth rotation affected object recognition. The rotation direction allowed us to change dynamic information while maintaining shape and view information. Given our goal, our approach differs from these previous studies in the following ways. First, we used a same/different discrimination task in which observers matched the *identity* of a study and test stimulus on each trial. The study stimulus was an animation of an object rotating in depth clockwise or counterclockwise about the vertical axis. Across trials, the test stimuli were seen views sampled from the animation and unseen views extrapolated from the implied trajectory. By using a static test image *combined* with this sampling procedure, we probed whether the visual system relies only on spatial information (e.g., shape, textures, parts, contours, and so on) available in the animation or whether the visual system *also relies* on dynamic information. Note also that this task relies on short-term memory for the study stimuli, which may be more sensitive at detecting whether dynamic information is used by default for recognition purposes (Thornton & Kourtzi, 2002).

Second, we varied how difficult it is to recognize objects on the basis of their 3D shape and/or 2D views (Hayward & Williams, 2000). Lastly, we randomly determined the rotation direction for each object on a trial-by-trial basis so that there was no association between an object and a rotation direction. Under these conditions, we argue that a *direction effect* on performance in this task would provide evidence (1) that motion specifies dynamic information independently of 3D shape and 2D views, and (2) that the visual system uses this dynamic information for recognition by default.

2. The role of object motion in object recognition

Motion is often cited as an important source of information for inferring 3D shape via structure-from-motion processes (see Todd, 1995, for a review of structure-from-motion). Most theorists have tried to characterize the minimal conditions necessary to successfully recover shape (usually with respect to an Euclidean or Affine geometry; Ullman, 1979), and the psychological validity of these conditions (Domini & Caudek, 1999; Todd & Bressan, 1990). What has largely been neglected, however, is the potential role of motion for object recognition beyond what it may tell observers about shape per se.

One reason for this omission is the assumption that recognition is largely shape based (e.g., Biederman, 1987) and that motion only serves to facilitate the recovery of shape for purposes of recognition (Marr & Nishihara, 1978). However, the effectiveness of shape information derived from motion for recognition is unclear. For example, it has been pointed out that some of the observed recognition advantages seen for rotating objects compared to static objects may be accounted for by the additional views that are necessarily provided by motion (e.g., Pike et al., 1997). At the same time, Lawson, Humphreys, and Watson (1994) found that observers could identify real-world objects and animals more accurately when they were presented as a coherent sequence of views that yielded an apparent rotation in depth than when the sequence of the *same views* were scrambled (but see Harman & Humphrey, 1999). Thus, there is evidence that motion specifies information beyond simply seeing more of an object—whether this is in the form of a 3D structure derived from its motion, a set of 2D views, or something else remains to be determined.

Alternatively, there has been a growing interest in examining whether the visual system encodes motion for the purposes of object recognition. One impetus for this interest is the simple fact that motion information can convey very subtle information regarding people, such as emotions (e.g., Bassili, 1978), gender (e.g., Mather & Murdoch, 1994), and even individuals (e.g., Cutting & Kozlowski, 1977; Hill & Pollick, 2000; Knappmeyer et al., 2003). Similarly, for novel objects, Stone (1998, 1999) found that observers were impaired in their ability to recognize studied amoeba-like objects when they rotated in the opposite direction from their studied direction. Recently, Liu and Cooper (2003) replicated Stone's results with block objects that were easy to discriminate from each other. In principle, the same 3D structure and 2D views should be recovered regardless of rotation direction. Again, the results from these studies suggest that visual system is sensitive to dynamic information at a level that goes beyond how motion specifies 3D shape or surface features.

The studies cited above provide strong evidence that object motion can become encoded in the object representation. However, this encoding seems to rely on repeated experience with objects moving in a characteristic manner, either during the course of an experiment (e.g., Stone, 1998, 1999) or from daily experience (e.g., Lander & Bruce, 2000). On the other hand, related studies on object priming (e.g., Kourtzi & Shiffar, 1999) and representational momentum (Freyd, 1987) suggest that motion has immediate effects on observers' perception of and memory for objects. For example, Kourtzi and Shiffrar found that a two-frame apparent motion sequence of an object rotating 120° in depth primed unseen views of that object within this trajectory. Similarly, Munger, Solberg, Horrocks, and Preston (1999) found that visual memories for the final depth orientation of a shaded cube were overestimated in the implied direction of rotation by about 2°. These studies suggest that the visual system uses dynamic information by default across a wide variety of tasks and stimuli.

Given the studies we have reviewed thus far, our preferred explanation for any direction effect is that the visual system automatically recovers dynamic information, in addition to shape and view information, from a moving object, and that this additional information serves as input to a recognition system. However, one issue we cannot directly address is that object motion may instead serve as input into an attentional system (e.g., Cavanagh, Labianca, & Thornton, 2001; Harman & Humphrey, 1999). We will return to this issue in Section 8.

3. Alternative explanations

In addition to the issue raised above, there are "static" effects that are not directly related to the fact that our stimuli are rotating in depth. Here we outline two candidate alternatives and how we addressed each in the present study.

3.1. Shape effects

As mentioned earlier, shape is often assumed to play a dominant role in object recognition. That is, one could argue that motion and other non-shape cues are used only in the "atypical" case when shape is non-informative, as with visually similar objects (Biederman & Ju, 1988). To address this issue, we used two sets of novel objects: a set of "easy" to discriminate objects that would facilitate invariant recognition performance with respect to the rotation direction (e.g., Biederman, 1987), and a set of "hard" to discriminate objects which might be influenced by the rotation direction under this alternative. The "easy" objects were readily decomposed into distinctive volumetric parts, and these parts were arranged in a very consistent manner across the set of objects. Furthermore, these objects had a well-defined axis of elongation and symmetry. In contrast, the "hard" objects were difficult to parse into distinctive parts, the arrangement of parts was arbitrary, and there were no well-defined axes of elongation or symmetry. If motion information was important only for objects that were difficult to discriminate from each other on the basis of their static shape information, then we predict that the direction of rotation should only affect the "hard" objects but not the "easy" ones.

3.2. Viewpoint and serial-position effects

Our dynamic stimuli consisted of an ordered-sequence of views. Thus, one could argue that observers simply remembered some views, particularly those toward the end of the sequence, and matched test images to these views because they were actually seen. Under this alternative, observers are encoding a set of views, rather than using motion per se (e.g., Bülthoff & Edelman, 1992; Tarr, 1995). Therefore, it seems unnecessary to propose any mechanism sensitive to dynamic information for object recognition. For example, static test images would be recognized more or less quickly and/or accurately depending on their rotations in depth from these encoded views (e.g., Tarr, Williams, Hayward, & Gauthier, 1998). This issue was addressed as follows. First, we compared how well observers can recognize two static views of an object that differed by a rotation in depth to how well observers can recognize a static view that differed by a rotation in depth from the last frame of an animation. If observers are simply remembering the final view in an animation, for example, we expect no difference between these two cases. Second, we randomized the frame order to change the available dynamic information relative to an ordered-sequence (Harman & Humphrey, 1999; Lawson et al., 1994). Again, if observers are remembering a set of views, we expect no difference in performance between these two cases.

4. The spatiotemporal link between animations and views

To test whether observers are sensitive to possible dynamic information of a moving object, we manipulated the *spatiotemporal* relationship between an animation and a subsequent static view. On each trial, observers were shown a sequence of views, each depicting the object from a slightly different viewpoint, comprising (and perceived as) a rotating object, followed by a brief blank interval, and then a static test view. Any view along a pre-defined 360° rotation of the object could be selected as the starting frame of the animation and the subsequent frames in the sequence traversed approximately 75° along this pre-defined rotation trajectory. Five static test views were selected with respect to this animation (created on a trial-by-trial basis), as shown in Fig. 1. The first three test views consisted of the first, middle, and last frame of the animation sequence. Note that these views are determined by the rotation direction. The remaining test views were novel views *not* shown during the animation: the pre-test view preceded the first frame of the animation and the post-test view followed the last frame of the animation.

It is important to note that there is *both* a spatial and a temporal relationship between the test view and any of the frames in the animation. The spatial relationship is the angular difference between the test view and the set of frames in the animation. The temporal relationship, on the other hand, is the temporal proximity from the test view to any of the frames presented in the animation. Across the three experiments reported, the interaction between these two relationships suggests that neither strictly spatial information (e.g., angular difference between the final view in an animation and a static test image) nor strictly temporal information (e.g., the recency of the final view in the animation) is sufficient to account for the results.

5. Experiment 1

Our goal in Experiment 1 was to examine whether observers are sensitive to an object's incidental rotation direction using a same/different discrimination task (Thornton & Kourtzi, 2002). That is, each object rotated in depth clockwise or counterclockwise randomly on a trial-by-trial basis so that observers could not associate any particular object with a rotation direction.

5.1. Method

5.1.1. Participants

Forty volunteers (23 females/17 males) were recruited from the Brown University community (undergraduates and graduate students). All participants provided informed consent and were paid for their time.

5.1.2. Stimuli

Two sets of novel 3D objects were used throughout the experiments reported here. The first set consisted 24 "easy" objects constructed from volumes such as bricks, cylinders, wedges, and so on. Fig. 2 shows examples of the objects from different viewpoints. These objects were based on those originally created by Biederman and Gerhardstein (1993). Each object consisted of a unique central part, two lateral parts of the same volume dif-

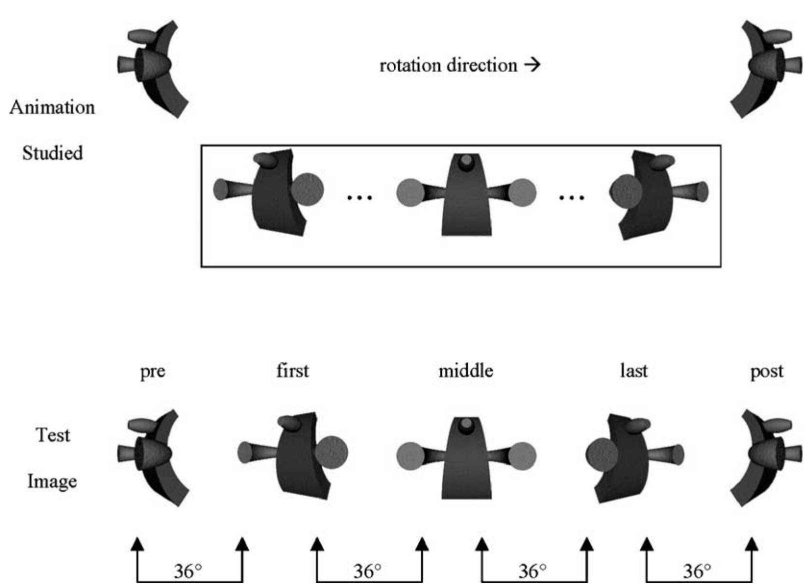

Fig. 1. An illustration of the spatiotemporal link between the animation and the test view used in Experiment 1. A novel object is shown rotating smoothly in depth with a particular direction of rotation (clockwise or counterclockwise). In this example, the object is rotating counterclockwise, as indicated by the direction arrow. Observers were shown the sequence of views highlighted by the box as the animation. Following this animation, observers were either shown the first, middle, or last frame of the animation, as the test view. Observers were also shown novel views that either preceded the first fame (pre) or followed the last frame (post) as test views. Note that consecutive test views were separated by a 36° rotation in depth.

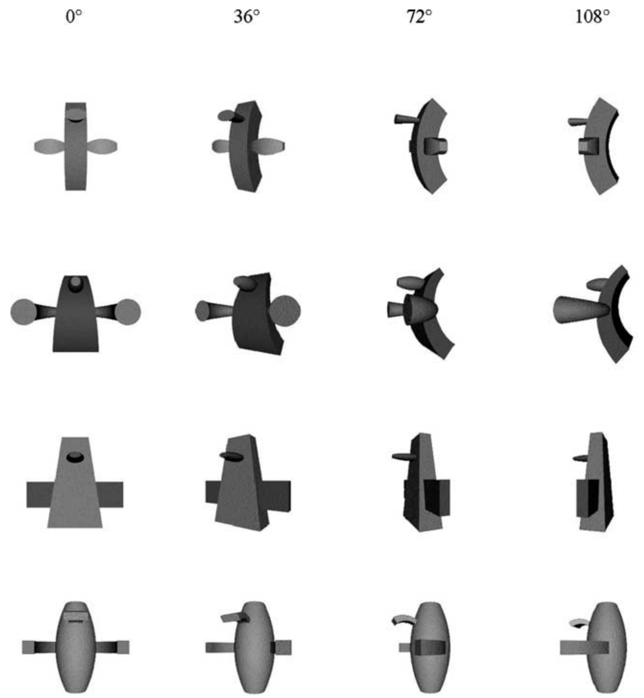

Fig. 2. Examples of the novel objects used in Experiments 1–3. The 0° was arbitrarily defined as the "frontal" view of the objects, and the other views shown are depth rotations from this frontal view.

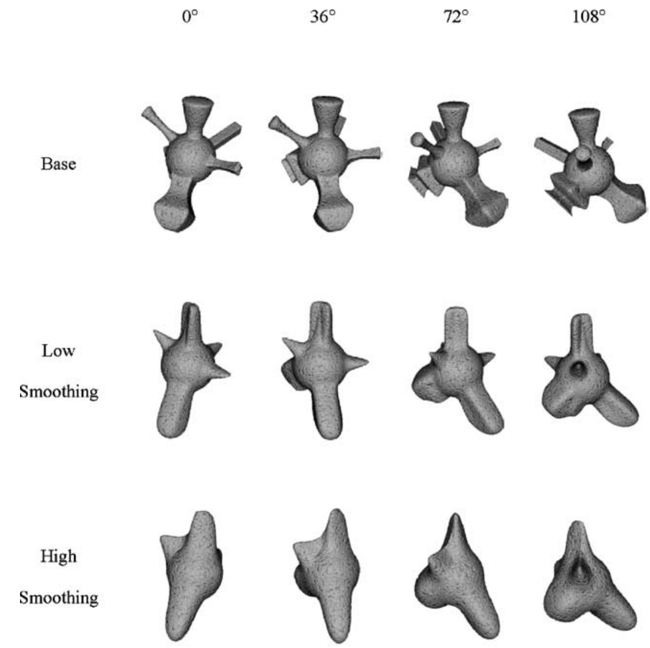

Fig. 3. Examples of the novel objects used in Experiments 1 and 2. The "low-smoothing" and "high-smoothing" objects were derived from the base object by smoothing corners and sharp edges by different amounts. The 0° view is arbitrarily defined, and the other views shown are depth rotations from 0°.

ferent from the central part, and a frontal part that was a different volume from both other parts. The central volume was the largest part, with the other parts approximately 50–70% smaller. The side of an object with the single frontal part was arbitrarily designated as its "front face" or 0° view.

The second set consisted of amoeba-like "hard" objects. In creating this set, nine base objects were first modeled. Each base consisted of a sphere with six parts randomly distributed across the surface of the sphere and placed at arbitrary depths along the surface normal. Unlike the previous set, the parts included both simple volumes, such as cones and boxes, and complex volumes, such as a "vase-like" shape. We created two variants of each base by smoothing out discontinuities, such as corners, to different degrees. Fig. 3 shows one of the base objects and its corresponding smoothed-versions (low, high). For these objects, an arbitrary point 800 units away from the center of the object and elevated 60° above the horizontal meridian defined the 0° view.

All objects were modeled in 3D Studio Max 4.0 (Discreet, Montreal, Quebec). They were illuminated by an ambient light source to ensure that all surface features were uniformly visible across the entire viewing sphere. A circular path with a radius of 160 arbitrary units was centered roughly about the center of mass of each object. A virtual camera moved along this path, and rendered a grayscale image of the object every 3.6° starting from the 0° view (front face) for a total of 100 views (Figs. 2 and 3). The objects were rendered against a black background. In addition, a texture was applied to the objects to give them a bumpy appearance in the final rendered images. When the 100 views were played in one sequential order, the object appeared to smoothly rotate clockwise about a vertical axis. Playing the views in reverse order produced a smooth counterclockwise rotation.

5.1.3. Design

Half the participants were tested with the "easy" object set and the other half were tested with the "hard" object set. For each set of objects, there were two trial types (same, different) and five test views (pre, first, middle, last, post; see Fig. 1). Each object appeared once in each of the 10 conditions, for a total of 240 trials ("easy" objects) or 270 trials ("hard" objects). These trials were completely randomized for each participant.

On each trial, participants were presented with an animation of a rotating object followed by a static test view. The animation consisted of a 21-frame sequence, randomly selected from the 100 possible views. The 21 frames were shown in either ascending order for clockwise rotation or in descending order for counterclockwise rotation. The rotation direction was randomly determined on each trial. Each frame was shown for approximately 25–35 ms; thus, the total duration of the animation was $\sim$735 ms. The animation showed the object rotating 75.6° in depth. Note that the short duration of each frame coupled with the small angular displacement on consecutive frames resulted in a strong impression of smooth motion.

The five test views were determined with respect to the sequential order of the 21-frame sequence selected (see Fig. 1). To illustrate: suppose frames 15–35 (inclusive) were selected with clockwise rotation (ascending frame numbers). The first test view would be frame 15, the middle test view would be frame 25, and the last test view would be frame 35. The pre-test view would be frame 5, and the post-test view would be frame 45. Note that the angular difference between consecutive test views is always 36° (10 frames). Note also that the pre-test view (i.e., frame 5) is 36° away from the first frame of the animation, and that the post-test view (i.e., frame 45) is 36° away from the last frame of the animation. On "different" trials, the test view was determined as described, but the view was randomly selected from one of the animations of the remaining objects (recall that all objects had the same 0° reference point).

5.1.4. Procedure

The experiment was run on an Apple iMac computer. The resolution of the monitor was set to 800 pixels × 600 pixels. PsychToolbox for Matlab 5.0 (Mathworks, Natick, MA) was used to control stimulus presentation and response collection (http://www.psychtoolbox.org/; Brainard, 1997; Pelli, 1997). Participants sat approximately 50 cm from the monitor in a normally lit room. At that viewing distance, each object in any particular view subtended approximately 9° × 9° of visual angle.

The procedure on each trial was as follows. First, a fixation cross was presented at the center of the screen for 500 ms. Following fixation, the animation was presented for approximately 735 ms. This animation was followed by a 500 ms (for "easy" objects) or a 1000 ms (for "hard" objects) blank interval, followed by the test view. The test view remained on the screen until participants responded by pressing either the "same" or "different" button on the keyboard to indicate their decision. No mask was used after the animation in order not to disrupt any perceptual processing induced by the animation. Participants were instructed to respond as quickly and as accurately as possible. All participants practiced with 20 randomly selected trials to become familiar with both the procedure and the response mapping.

5.2. Results

Given the overall difficulty of the "hard" objects relative to the "easy" ones, we analyzed mean correct response times (RTs) and mean sensitivity (d' scores) separately for each type of objects. To compute the d' scores, hits were defined as responding "same" on "same" trials, and false alarms were defined as responding "same" on "different" trials. For the "easy" objects, we removed RTs that were greater than 2500 ms or less than 400 ms. This RT range eliminated less than 2.5% of correct trials. For the "hard" objects, RTs from correct trials greater than 4000 ms and RTs less than 400 ms were removed. Again, this range removed less than 2.5% of correct trials. For this and subsequent experiments reported, we only analyzed response times from "same" trials.

In Experiment 1, both RTs and d' scores were submitted to a repeated-measures analysis of variance (ANOVA) with test view (pre, first, middle, last, post) as the only within-subjects factor. An $\alpha = 0.05$ was adopted as the significance level for all analyses.

5.2.1. Response times

The RT data from "same" trials for Experiment 1 are shown in the top panels of Figs. 4 and 5. For both "easy" and "hard" objects, there was a significant effect of test view ("easy": $F(4, 76) = 11.91$, $p < 0.01$ and "hard": $F(4, 76) = 15.56$, $p < 0.01$). We also analyzed the familiar test views (first, middle, and last views)

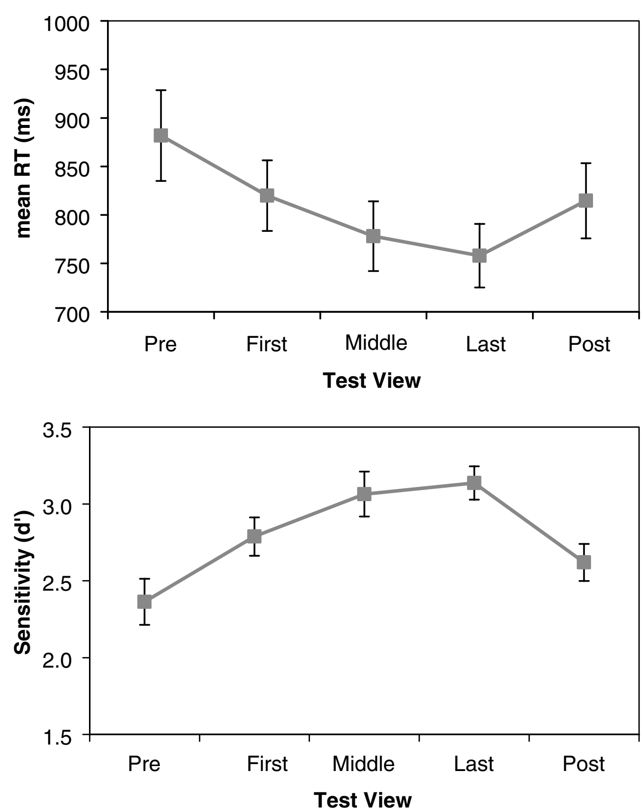

Fig. 4. Results obtained from Experiment 1 for the "easy" objects. Mean response times (top panel) for same trials and mean d' scores (bottom panel) were averaged across participants as a function of test view. Errors bar show standard error of participant means.

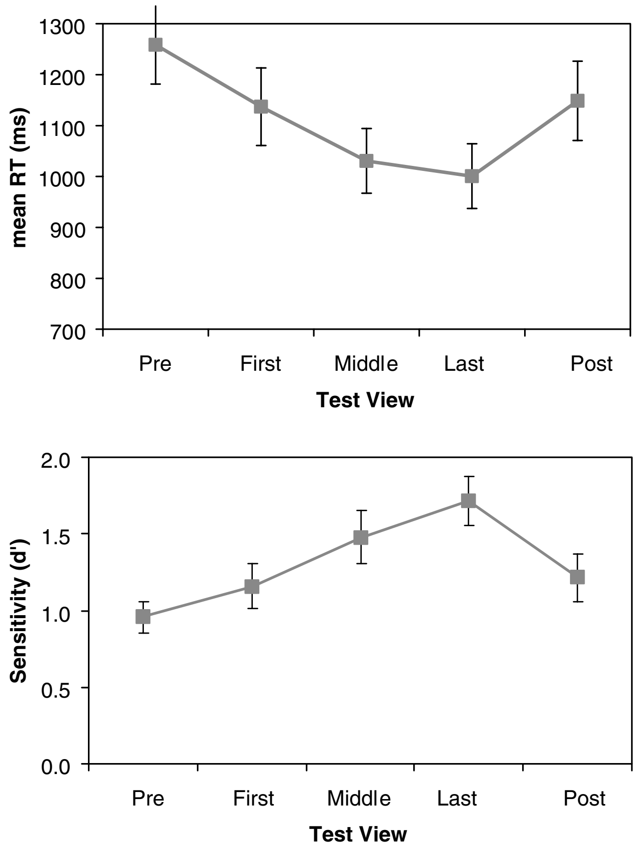

Fig. 5. Results obtained from Experiment 1 for the "hard" objects. Mean response times (top panel) for same trials and mean d' scores (bottom panel) were averaged across participants as a function of test view. Errors bar show standard error of participant means.

separately to examine whether or not the novel test views (pre-test and post-test views) were driving the results. Without the novel test views, the main effect remained significant for both object sets ("easy": $F(2,38) = 9.37$, $p < 0.01$, and "hard": $F(2,38) = 10.13$, $p < 0.01$). Lastly, a post-hoc t-test revealed that participants responded more quickly to post-test views than to pre-test views for both object sets ("easy": $t(19) = 2.86$, $p < 0.01$ and "hard": $t(19) = 2.40$, $p < 0.05$).

5.2.2. Sensitivity

The d' data for Experiment 1 are shown in the bottom panels of Figs. 4 and 5. As with response times, there was a highly significant main effect of test view for both types of objects ("easy": $F(4,76) = 7.69$, $p < 0.01$, and "hard": $F(4,76) = 8.89$, $p < 0.01$). When only the familiar test views were analyzed, there was only a marginally significant main effect for "easy" objects ($F(2,38) = 2.57$, $p = 0.09$) but a significant effect for "hard" objects ($F(2,38) = 6.88$, $p < 0.01$). A post-hoc test did not find a significant difference between post-test views and pre-test views for "easy" objects ($t(19) = 1.45$, $p = 0.16$), and found only a marginally significant difference for "hard" objects ($t(19) = 1.89$, $p = 0.07$). Although there were differences in the patterns of performance in response times and sensitivity, there was no indication of a speed-accuracy tradeoff in any portion of the experiment.

5.3. Discussion

There were two results to highlight in Experiment 1. First, we found an effect of test view on performance for familiar views. Participants responded more quickly and more accurately for familiar test views that were sampled towards the end of the animation. Second, we found differences in recognition performance across the two novel test views. For these test stimuli, we found that participants were faster with the post-test view than with the pre-test view. Differences in performance were obtained for both familiar (shown during the animation) and novel test views, suggesting that the test view effect is not due to a seen/not-seen distinction. Rather, we interpret the effects of test view on performance as a consequence of the rotation direction depicted in the animation on a by-trial basis, since all test views were defined with respect to this direction.

By comparing this *direction effect* across the "easy" and "hard" objects, we show that motion information is not only used in restricted circumstances, as when objects are difficult to discriminate from each other (e.g., Biederman & Ju, 1988). Rather, object dynamics, independent of object shape and views, affect recognition in the same way that any aspect of an object's appearance, such as viewpoint, may affect recognition (e.g., Tarr, Williams, et al., 1998). Indeed, Stone (1998, 1999) and Liu and Cooper (2003) separately found that long-term memory of "hard" and "easy" novel objects was affected by rotation direction. Here we show a similar direction effect for the short-term memory of "easy" and "hard" objects.

6. Experiment 2

In Experiment 2 our goal was to examine whether viewpoint effects might account for the results obtained in Experiment 1. For a wide range of objects, several studies have found a viewpoint effect—a linear increase in response times and/or error rates with a linear increase in viewpoint differences between two views of an object (e.g., Bülthoff & Edelman, 1992; Humphrey & Khan, 1992; Tarr, Williams, et al., 1998). Can the direction effect reported in Experiment 1 be explained by this strictly spatial relationship between the test views and the frames of the animation? To that end, we tested participants in a same/different discrimination task in which both studied and test stimuli were static images. If the recognition patterns obtained in Experiment 1 were due to the angular difference between views, then a similar pattern should be seen in Experiment 2 for both sets of objects.

6.1. Method

6.1.1. Participants

Twenty naïve participants from the Brown University community (13 females/7 males) were recruited for this experiment. All participants provided informed consent and were paid for their time. These participants were tested with the "easy" objects. We also tested 10 participants (4 females/6 males) with the "hard" objects. Seven of these 10 observers had participated in Experiment 1; however, several weeks had passed since they had originally participated in Experiment 1.

6.1.2. Stimuli

The same two sets of objects were used in Experiment 2. However, here, individual views were presented as static study and test images.

6.1.3. Design and procedure

There were two trial types (same, different) and four angular differences between the study image and the test image (0°, 36°, 72°, and 108°). These differences were chosen because they correspond to the angular differences between the last frame of the animation used in Experiment 1 and the last test view, the middle and post-test views, the first test view, and the pre-test view, respectively. Each of the 24 "easy" objects or 27 "hard" objects appeared once in the eight possible conditions for a total of 192 or 216 trials. These trials were completely randomized for each participant.

On each trial, participants were presented with a static study image followed by a static test image. The study image was randomly selected from the 100 possible views of the entire 360° rotation and shown for approximately 735 ms (i.e., the duration of the animation in Experiment 1). The test image was a view of the object that was rotated by 0°, 36°, 72°, or 108°. The direction of rotation, clockwise or counterclockwise, was randomly determined on each trial. On "different" trials, the test image was randomly selected from the views for one of the remaining objects. The same experimental setup and procedure as in Experiment 1 was used in the present experiment.

6.2. Results

Response time and sensitivity data for "easy" and "hard" objects were submitted to separate repeated-measures ANOVA with angular difference (0°, 36°, 72°, 108°) as a within-subjects factor.

6.2.1. Response times

The RT data from "same" trials for Experiment 2 are shown in the top panels of Figs. 6 and 7. For both types of objects, the ANOVA revealed a main effect of angular difference ("easy": $F(3,57) = 17.43$, $p < 0.01$, and "hard": $F(3,27) = 9.08$, $p < 0.01$). However for the "easy" objects, this main effect appears to be driven mostly by the 0° condition, in which the study and target images were physically identical. An ANOVA excluding this condition revealed no significant main effect of angular difference ($F(2,38) = 0.15$, $p > 0.05$). By comparison, for the "hard" objects, an analysis without the 0° condition revealed a significant effect of angular difference ($F(2,18) = 4.37$, $p < 0.05$).

6.2.2. Sensitivity

The d' data for Experiment 2 are shown in the bottom panels of Figs. 6 and 7. The ANOVA revealed a significant effect of angular difference for both types of objects ("easy": $F(3,57) = 36.67$, $p < 0.01$, and "hard": $F(3,57) = 18.86$, $p < 0.01$). We also analyzed the d' data excluding the 0° angular difference. Unlike response times, this analysis revealed a significant effect of

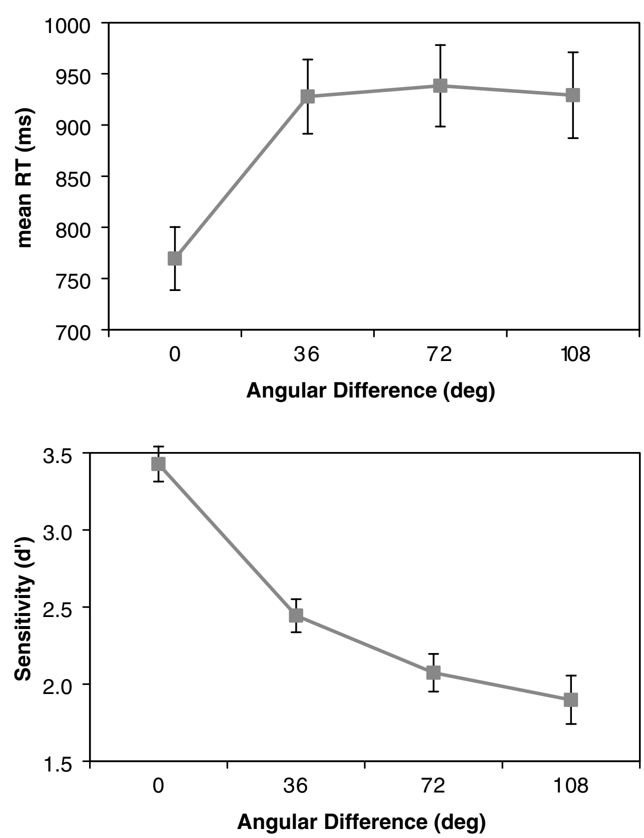

Fig. 6. Results obtained from Experiment 2 for the "easy" objects. Mean response times (top panel) for same trials and mean d' scores (bottom panel) were averaged across participants as a function of angular difference between study and test stimuli. Errors bar show standard error of participant means.

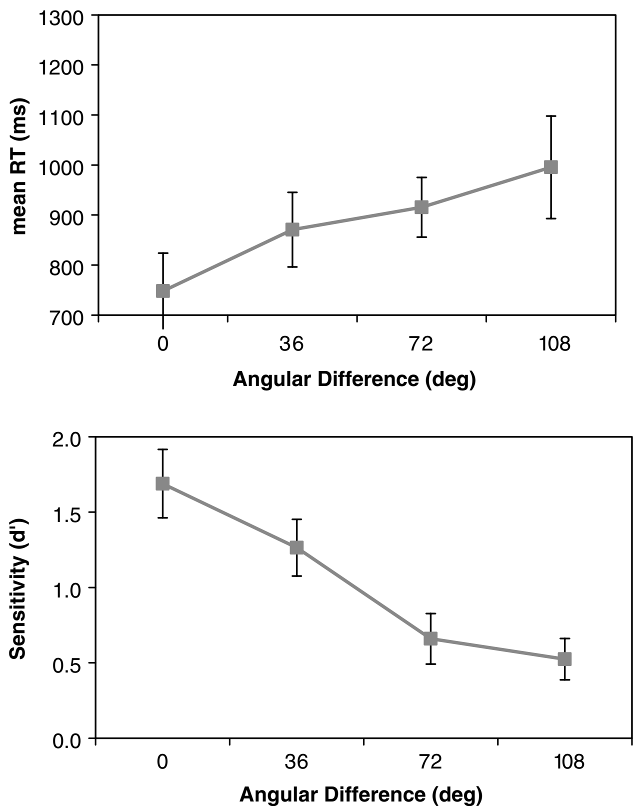

Fig. 7. Results obtained from Experiment 2 for the "hard" objects. Mean response times (top panel) for same trials and mean d' scores (bottom panel) were averaged across participants as a function of angular difference between study and test stimuli. Errors bar show standard error of participant means.

angular difference for both types of objects ("easy": $F(2,38) = 5.61, p < 0.01$, and "hard": $F(2,18) = 13.02$, $p < 0.01$). There were no indications of speed-accuracy tradeoffs.

6.3. Discussion

In Experiment 2 there were differences in the pattern of results between "easy" and "hard" objects. For the "easy" objects, we did not find a viewpoint effect on response times except for physically identical matches in the two images (Biederman & Gerhardstein, 1993). By comparison, for the "hard" objects, we found a robust viewpoint effect on response times even excluding the 0° angular difference. This difference across the two sets of objects suggests that the viewpoint effect is modulated by how difficult it is to recognize the objects on the basis of their 3D shape or projected 2D views. Indeed, Hayward and Williams (2000) showed robust viewpoint effects when the set of objects was difficult to discriminate from each other but not when the set of objects were easy to discriminate from each other. Rather than constructing "easy" or "hard" models as we did, they varied the context by including objects from the same "family" in the difficult context and different "families" in the easy context.

In contrast, recognition of both "easy" and "hard" objects was affected by the rotation direction in Experiment 1. For the "easy" objects, it is important to point out that sensitivity to the direction of rotation did not come with a concurrent cost to overall recognition performance: the mean response times and sensitivity in Experiments 1 and 2 were 813 and 880 ms, and 2.79 and 2.46. Thus, the results across the two experiments support the hypothesis that observers were sensitive to dynamic information and not to particular views (or distinctive features) seen per se.

In Experiment 2, we found that changing the study stimulus from an animation to a static image had a drastic effect on how quickly the "easy" objects were recognized. However, there are obvious differences in the number and duration of each image across these two different study stimuli. The study animation used in Experiment 1 presented a series of images from slightly different views, each shown very briefly (~30 ms). In contrast, the static study image used in Experiment 2 was the same view of an object presented for a much longer duration (~730 ms). To address this problem, in the last experiment we scrambled the view order of the animation for "easy" objects.

7. Experiment 3

In Experiment 1, we presented views of the study object in sequential order. This presentation was perceived as a smooth, continuous clockwise or counterclockwise rotation in depth. In Experiment 3, we presented views of the study object in random order (Harman & Humphrey, 1999; Lawson et al., 1994). This presentation is perceived as discontinuous rotations back and forth in depth. If observers are sensitive to dynamic information for recognition purposes, then we predict a different pattern of results in response times and sensitivity with respect to those obtained in Experiment 1. In particular, because the scrambled view sequence is no longer perceived to rotate smoothly clockwise or counterclockwise, we expect that the effects of the familiar test frames will be reduced and that there will be no differences in performance between the pretest and post-test views.

7.1. Method

7.1.1. Participants

Forty naíve participants (21 females/19 males) were recruited from Brown University. Half the participants were randomly assigned to a no global-rotation condition (NO group), and the remaining participants were assigned to a weak global-rotation condition (WEAK group). All participants provided informed consent and were paid for their time.

7.1.2. Stimuli

Only the "easy" objects used in Experiment 1 were used in Experiment 3.

7.1.3. Design and procedure

The same design and procedure used in Experiment 1 was used in the present experiment, with the exception that the frame order of the animation was scrambled as follows. First, we grouped the 21-frame animation sequence into seven 3-frame subsequences, and then scrambling the order of these subsequences. The consecutive 3-frame subsequences ensured that there were corresponding features across views to provide local motion information sufficient for structure-from-motion processes (Ullman, 1979).

We used two slightly different scrambling procedures for the two groups of participants. Fig. 8 illustrates an example of both procedures. In the NO group, we scrambled the 3-frame subsequences with the constraint that there were not more than two consecutive sets in succession. With this procedure, there should be no global-rotation direction. In the WEAK group, we maintained the position of the first, fourth, and seventh 3-frame subsequence, and scrambled the remaining sets (with the same constraint as the first procedure). Thus, in the WEAK group, there was a "weak" global coherent rotation clockwise or counterclockwise, established by the fixed 3-frame subsequences. Note that although these scrambling procedures preserve object views (with respect to the unscrambled sequence), they necessarily do not produce smooth rotations back and forth.

Finally, we note the effect of the scrambling procedures on the relationship between the animation and the static test view. The familiar test views were simply determined relative to the new (scrambled) 21-frame sequence. The novel test views were problematic, however, because the global direction of rotation was disrupted. Our solution was to select the pre-test and post-test views relative to the *unscrambled* 21-frame sequence. Note that for the NO group, this meant that the angular differences between the pre-test and post-test views relative to the last view of the scrambled sequence were no longer valid (i.e., they were unlikely to be 108° and 36°, respectively). For the WEAK group, the angular difference was not an issue because the first, fourth, and last 3-frame subsequences of the unscrambled sequence were fixed (Fig. 8).

7.2. Results

The RT and sensitivity data were submitted to a mixed-design ANOVA with scrambling procedure (NO group vs. WEAK group) as a between-subjects factor and test view (first, middle, last) as a within-subjects

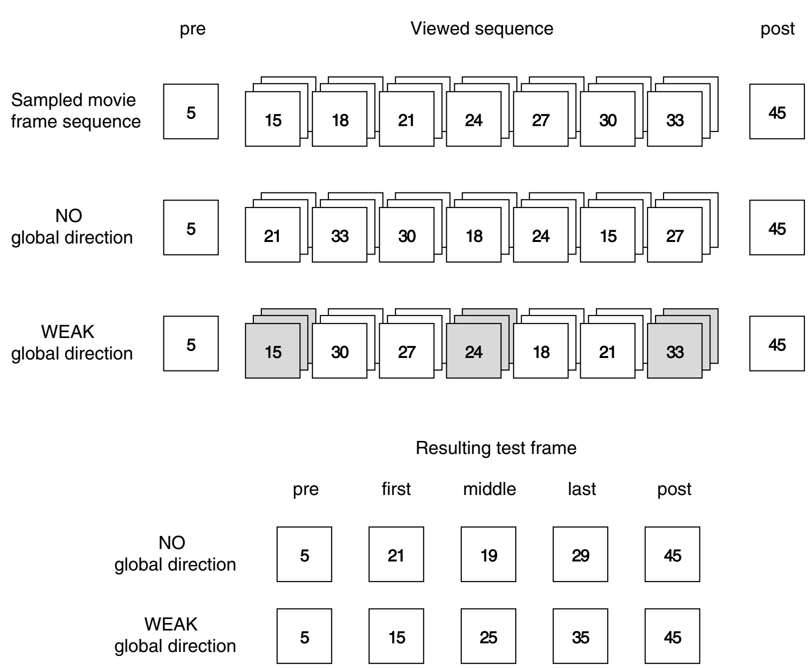

Fig. 8. An example of the NO and WEAK scrambling procedures used in Experiment 3. The 21-frame sequence sampled from the full animation was first grouped into sets of three consecutive frames, and then the order of the 3-frame subsequences was randomized. The main difference between the two procedures is that with the WEAK procedure the order of three of the seven subsequences was fixed (as illustrated by the gray boxes). The bottom half of the figure shows the resulting test views that would be used for each procedure.

factor. We eliminated the pre-test and post-test views from the omnibus ANOVA because they were defined relative to a global direction of rotation, which was eliminated by both rotating the object back and forth in depth. However, we still conducted post-hoc analyses to test for differences for these two arbitrarily labeled conditions.

7.2.1. Response times

The RT data from "same" trials for Experiment 1 are shown in the top panel of Fig. 9. The omnibus ANOVA showed only a significant interaction between scrambling procedure and test view [2] ($F(2, 76) = 4.31$, $p < 0.05$). For the WEAK group, there is a significant linear trend in the response times as a function of the test view ($F(1, 19) = 10.29$, $p < 0.01$). In contrast, for the NO group, there are no a priori trends in response times (linear: $F(1, 19) = 0.06$, $p > 0.05$ and quadratic: $F(1, 19) = 1.06$, $p > 0.05$). Post-hoc tests revealed that the pre-test and post-test views were not significantly different from each other for both scrambling procedures (for both NO and WEAK group: $t(19) < 1$).

7.2.2. Sensitivity

The d' data from Experiment 3 are shown on the bottom panel of Fig. 9. The interaction between scrambling procedure and test view was marginally significant for sensitivity ($F(2, 76) = 16.50$, $p = 0.08$). No main effects were significant (all Fs < 1). For the WEAK group, there were no significant a priori trends in sensitivity (linear: $F(1, 19) = 2.93$, $p = 0.10$ and quadratic: $F(1, 19) = 0.46$, $p > 0.05$). Similarly, for the NO group, there were no significant trends (linear: $F(1, 19) = 0.35$, $p > 0.05$ and quadratic: $F(1, 19) = 1.79$, $p > 0.05$). Like the response time data, post-hoc tests revealed no significant differences between the pre-test views and the post-test views for either scrambling procedure (for both NO and WEAK group: $t(19) < 1$). The results do not indicate any speed-accuracy tradeoffs.

7.3. Discussion

In Experiment 3, we changed the dynamic information of the study stimuli with respect to the dynamic information of the study stimuli in Experiment 1. In this case, we found corresponding differences in the pattern of results with both familiar and novel test views. For the NO direction group, we found that observers re-

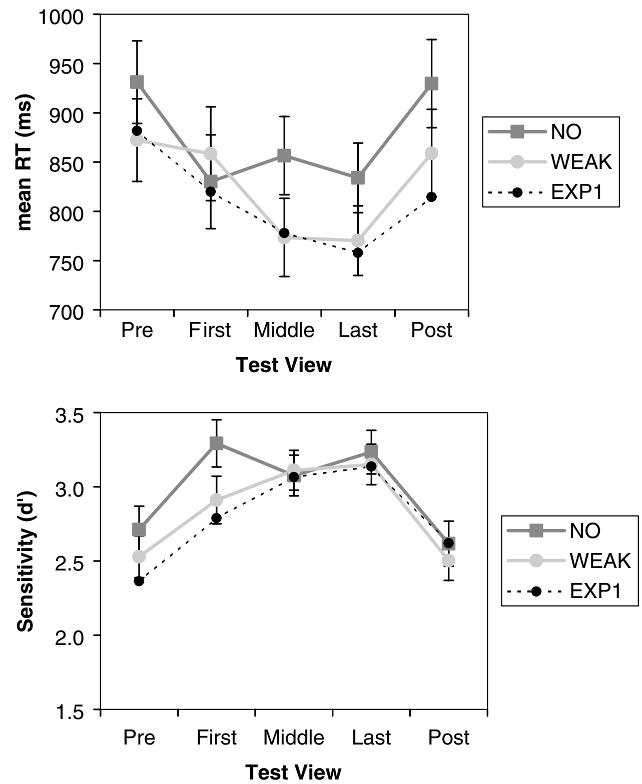

Fig. 9. Results obtained from Experiment 3. The mean response times (top panel) and mean d' scores (bottom panel) were averaged across participants as a function of test view for the NO and WEAK direction group. Errors bar show standard error of participant means. The results obtained from Experiment 1 are also plotted for easy comparison across the three experiments.

sponded equally quickly for all familiar views tested. We also found that observers responded equally fast for the pre-test and post-test views. On the other hand, for the WEAK direction group, there was a small effect of familiar test views. However, in contrast to Experiment 1, we did not find a significant difference in response times for pre-test and post-test views. Based on these findings in conjunction with the lack of a viewpoint effect in Experiment 2 for the "easy" objects, we conclude that the direction effect found in Experiment 1 was not due to observers encoding only static views of objects. Lastly, we note that participants generally responded equally quickly and accurately for the three different study stimuli used in Experiments 1–3 (810 ms and $d' = 2.79$ in Experiment 1; 891 ms and $d' = 2.46$ for Experiment 2; 876 ms and $d' = 2.99$ for the NO group in Experiment 3; 832 ms and $d' = 2.84$ for the WEAK group in Experiment 3).

8. General discussion

One assumption cutting across different theories of object recognition is that recognition is largely driven by

[2] For response times, including the pre-test and post-test views resulted in a significant main effect of test view ($F(4, 152) = 14.71$, $p < 0.01$) and a significant interaction between scrambling procedure and test view ($F(4, 152) = 2.34$, $p < 0.05$). For sensitivity, there was only a significant main effect of test view ($F(4, 152) = 16.50$, $p < 0.01$), and no interaction between scrambling procedure and test view ($F(4, 152) = 1.10$, $p > 0.05$). However, these results are likely to be driven by the novel test views.

object shape and/or views (e.g., Biederman, 1987; Hayward, 1998; Marr & Nishihara, 1978; Tarr, 1995). Here we hypothesize that recognition is also driven by dynamic information, independent of what such information tells observers about shape (Freyd, 1987; Stone, 1998, 1999). Across three experiments, we varied the dynamic information but maintained the same 3D shape and 2D views by changing the rotation direction. Consistent with our hypothesis, changing the dynamic information had different effects on recognition, which manifested as a direction effect on response times and sensitivity in a same/different discrimination task with "easy" and "hard" novel objects. Thus, our results indicate that motion was not used simply to derive shape information or enhance other aspects of shape processing (e.g., segmentation), that what observers encoded went beyond any particular views shown during a study animation, and that the effect was not simply a serial-position effect tied to recent views in the animation. Rather motion afforded dynamic information that was directly used for recognition. Our results also suggest that the visual system uses dynamic information for recognition by default: that is, sensitivity to dynamic information develops quickly (within ∼730 ms) irrespective of object geometry and without necessarily associating any particular objects with a particular direction.

Interestingly, the claim that the visual system processes dynamic information by default is reminiscent of claims for pre-attentive or automatic processing of primitive sensory features such as color or orientation (e.g., Treisman & Gelade, 1980). Indeed, an intriguing line of inquiry to be addressed by future experiments is whether there are primitive features that are *dynamic*. A second direction for future research is how automatic processing of dynamic information may ultimately lead to the development of more long-term visual representation of object motion, such as the rotation direction, as demonstrated by Stone (1998, 1999) and Liu and Cooper (2003) (see also Vuong, 2004). For example, it would be interesting to see whether rotation-reversal could affect performance in our same/different discrimination task *after* observers had some prior experience with the objects rotating in a particular direction.

Our extension of the same/different discrimination task allowed us to examine the spatiotemporal relationship between a dynamic stimulus and a static test view (Bülthoff & Edelman, 1992; Mitsumatsu & Yokosawa, 2003; Thornton & Kourtzi, 2002). In particular, by probing different views along the trajectory of a rotating object, we found that strictly spatial (i.e., view) or strictly temporal information (i.e., recency) does not sufficiently account for our results. With this paradigm, we also found that the rotation direction affected how observers generalize to novel views of the objects (Freyd, 1987; Kourtzi & Shiffar, 1999). Using a priming paradigm, Kourtzi and Shiffrar found generalization to novel views for large rotations (i.e., 120°) but not for small rotations (i.e., 60°). Here we found generalization for small rotations in the implied direction of rotation (i.e., 36°).

Given that we sampled views along a rotation trajectory, there are parallels that can be drawn between the direction effect on recognition and representational momentum on judgments of final positions of dynamic displays (Freyd, 1987). Munger et al. (1999), for example, found that observers watching a shaded cube rotate in depth in a three-frame apparent motion sequence overestimated the true orientation of a fourth test frame. Here we found better recognition for post-test views that were consistent with the rotation direction (Experiment 1). Therefore, representational momentum can provide one possible account of our results since these investigators speculate that this overestimation is caused by the visual system internalizing physical inertia thereby distorting the memory of the object (Freyd, 1987). However, there has been no direct test whether this memory distortion have any repercussions for object recognition. For example, the distractors in representational momentum studies are different views of the same object. By comparison, our distractors were other objects. Thus, our findings establish a possible connection between sensitivity to this "representational inertia" and higher visual functions such as object recognition. Future studies that explore this connection may provide insights into how the visual system utilizes dynamic information for vision in general.

At the same time, as raised in Section 1, it is important to bear in mind that attention can influence how dynamic objects are recognized (e.g., Cavanagh et al., 2001; Harman & Humphrey, 1999). For example, the visual system may track distinctive features of an object as it rotates, thereby allowing it to "anticipate" views (or at least the tracked feature) of that object (e.g., Mitsumatsu & Yokosawa, 2003). If the static test view violates this anticipated view, then the attentional system may need to reorient its focus, which may produce delays or more errors. Similarly, in Experiment 3, the scrambling procedure we employed may reduce the possibility of feature tracking, thereby producing no performance differences between pre-test and post-test frames. Alternatively, our scrambling procedure may produce "spatiotemporal discontinuities" when views "jumped" from one view to another non-consecutively. It is possible that attention is drawn to views that are temporally close to these discontinuities thereby ameliorating the effects of familiar and novel test views found in Experiment 1.

However, two aspects of our study suggest that observers were sensitive to the rotation direction *in addition to* any contributions of attention. First, the "easy" and "hard" objects used varied in the amount of "distinctive" features available to be tracked as the focus of attention but both were affected by the randomly determined rotation direction, whereas only the "hard" objects were affected by viewpoint differences when both the study and test stimuli were static images (Experiment 2). Second, the attentional system would also need to reorient to the first test frame since this test frame also violates the anticipated view. However, we found equivalent performance between the first test frame and the post-test frame (Experiment 1). Taken in conjunction with the results of prior object-recognition studies reviewed here (e.g., Bülthoff & Edelman, 1992; Liu & Cooper, 2003; Stone, 1998, 1999), our results are consistent with the hypothesis that dynamic information is encoded in the short-term representation of objects but further studies are needed to directly address attentional effects in encoding dynamic information afforded by moving objects.

8.1. Implications for theories of object recognition

Overall, our present results are consistent with results from previous behavioral studies that have used different tasks and stimuli (Bülthoff & Edelman, 1992; Liu & Cooper, 2003; Stone, 1998, 1999; Thornton & Kourtzi, 2002). Thus, we can formulate three generalizations regarding the role of motion in object recognition.

First, motion information seems to plays a role in recognition across a range of stimulus classes that differ with respect to similarity and familiarity (e.g., Lander & Bruce, 2000; Lawson et al., 1994; Thornton & Kourtzi, 2002). In our study, we found a direction effect for both "easy" and "hard" objects that differed in geometry. Second, different types of object motion seem to play a role in recognition. For example, Thornton and Kourtzi found motion-specific effects using non-rigid facial expressions (Cavanagh et al., 2001; Knappmeyer et al., 2003). Similarly, Stone (1998, 1999) found a direction effect for a complex tumbling motion. Here we found differences in recognition for objects rotating continuously in depth about a single axis (see also Liu & Cooper, 2003).

Third, motion information seems to play a role across different recognition tasks. Stone (1998, 1999), for example, used an old/new recognition memory paradigm in which observers discriminated target objects from distractors. He found that reversing the studied motion direction impaired recognition performance (see also Liu & Cooper, 2003). Recently, Vuong (2004) replicated Stone's results using a task in which observers identified a subset of the "easy" and "hard" objects at the individual level. For "easy" objects, observers only seemed to encode motion direction when they learned them in a "dynamic fog" that degraded both shape and motion information. By comparison, for "hard" objects, observers encoded motion direction irrespective of learning condition. Thus, object geometry may affect whether or not motion direction is encoded in long-term memory for objects, possibly because observers are learning both shape and motion. In contrast, the present results indicate that dynamic information has immediate effects on the short-term memory of objects, irrespective of object geometry.

Lastly, motion also seems to play a role in tasks that require categorical discriminations. For example, Knappmeyer et al. (2003) found that associating facial motions with individual faces biased the perception of those faces, particularly when facial form was ambiguous. To reiterate, the important point suggested by these generalizations is that motion information is used *by default*, and not restricted to particular tasks or stimuli.

8.2. Conclusion

There is a growing body of evidence that observers encode "visually rich" object representations for recognition. That is, more than shape and/or view information is included, by default, in our visual knowledge about objects. For example, there is evidence for the representation of the effects of lighting (e.g., Tarr, Kersten, & Bülthoff, 1998), of color (e.g., Naor-Raz, Tarr, & Kersten, 2003; Price & Humphreys, 1989; Tanaka, Weiskopf, & Williams, 2001), and of motion (e.g., Cavanagh et al., 2001; Knappmeyer et al., 2003; Mather & Murdoch, 1994; Stone, 1998, 1999). Thus, object recognition will depend on the many different measures of object appearance encoded during original viewing. In line with these studies, we provide evidence that these measures are not static; rather, they are affected by the dynamics of the objects (Freyd, 1987). In sum, dynamic information does not simply refine 3D shape information nor does it simply provide more views of an object. Instead, we argue that object dynamics contribute to the richness of its representation.

Acknowledgements

QCV was supported by a post-graduate scholarship from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Both authors would like to thank Anjula Joshi and Giulia Righi for help with data collection. We would also like to thank two anonymous reviewers for comments and suggestions on an earlier draft of this manuscript.

References

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 373–379.

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

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

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

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

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition, 80*, 47–60.

Cutting, J., & Kozlowski, L. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomics Society, 9*, 353–356.

Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 426–444.

Freyd, J. (1987). Dynamic mental representations. *Psychological Review, 94*, 427–438.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin Company.

Harman, K. L., & Humphrey, G. K. (1999). Encoding regular and random sequences of view of novel three-dimensional objects. *Perception, 28*, 601–615.

Hayward, W. G. (1998). Effects of outline shape in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 427–440.

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science, 11*, 7–12.

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light display. *Psychological Science, 11*, 223–227.

Humphrey, G. K., & Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology, 46*, 170–190.

Kelly, M. H., & Freyd, J. (1987). Explorations of representational momentum. *Cognitive Psychology, 19*, 369–401.

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research, 43*, 1921–1936.

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

Landes, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology, 12*, 259–272.

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.

Liu, T., & Cooper, L. A. (2003). Explicit and implicit memory for rotating objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 554–562.

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

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London, 258*, 273–279.

Mitsumatsu, H., & Yokosawa, K. (2003). Efficient extrapolation of the view with a dynamic and predictive stimulus. *Perception, 32*, 969–983.

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.

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception, 32*, 667–680.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences, 6*, 261–266.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition, 4*, 409–437.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology, 41A*, 797–828.

Rubin, N., & Albert, N. (2001). Real-world scenes can be easily recognized from their edge-detected renditions: Just add motion! In *First annual meeting of the Vision Sciences Society, Sarasota, FL, May 4–8*.

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research, 38*, 947–951.

Stone, J. V. (1999). Object recognition: View-specificity and motion-specificity. *Vision Research, 39*, 4032–4044.

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Science, 5*, 211–215.

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

Tarr, M. J., & Bülthoff, H. H. (1998). Object recognition in man, monkey, and machine. *Cognition, 67*, 1–20.

Tarr, M. J., Kersten, D., & Bülthoff, H. H. (1998). Why the visual system might encode the effects of illumination. *Vision Research, 38*, 2259–2275.

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

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

Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic faces. *Perception, 31*, 113–132.

Todd, J. T. (1995). The visual perception of three-dimensional structure from motion. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 201–226). New York: Academic Press.

Todd, J. T., & Bressan, P. (1990). The perception of 3dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48*, 419–430.

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

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Vuong, Q. C. (2004). *Spatio-temporal signatures for object recognition: Motion and shape interact in discriminating, remembering, and learning novel rotating objects*. Unpublished doctoral dissertation, Brown University, Providence, RI, USA.