

Intrinsic orientation and study viewpoint in recognizing spatial structure of a shape

XIAOOU LI

Institute of Psychology, Chinese Academy of Sciences, Beijing, China

WEIMIN MOU

*Institute of Psychology, Chinese Academy of Sciences, Beijing, China
and University of Alberta, Edmonton, Alberta, Canada*

AND

TIMOTHY P. MCNAMARA

Vanderbilt University, Nashville, Tennessee

Two experiments dissociated the roles of intrinsic orientation of a shape and participants' study viewpoint in shape recognition. In Experiment 1, participants learned shapes with a rectangular background that was oriented differently from their viewpoint, and then recognized target shapes, which were created by splitting study shapes along different intrinsic axes, at different views. Results showed that recognition was quicker when the study shapes were split along the axis parallel to the orientation of the rectangular background than when they were split along the axis parallel to participants' viewpoint. In Experiment 2, participants learned shapes without the rectangular background. The results showed that recognition was quicker when the study shape was split along the axis parallel to participants' viewpoint. In both experiments, recognition was quicker at the study view than at a novel view. An intrinsic model of object representation and recognition was proposed to explain these findings.

People experience an object visually from a single viewpoint at any given time. In most situations, the viewpoint from which an object is first experienced is different from the viewpoint from which it is later recognized. Recognition must therefore rely on a representation and a process that can accommodate the change of the viewpoints from study to recognition. The models proposed to conceptualize the nature of such mental representation and process can be divided into two categories: structural description models and view-based models (Hayward, 2003).

Structural description models claim that an object is represented as a set of parts with a specific structural description of the spatial relations among the parts, and recognition relies on matching the structure descriptions between the test object and the objects in memory (e.g., Biederman, 1987; Marr & Nishihara, 1978). For example, a hand can be represented as five fingers and one palm, with specific spatial relations among the parts. The view-based models claim that snapshots of an object are represented at all learning viewpoints and that recognition relies on the normalization between the recognition view and the closest snapshot in memory (e.g., Tarr & Pinker, 1989, 1990; Ullman, 1989).

The view-based models have found support in findings that object recognition performance is better at a familiar

view than at a novel view, and that performance at a novel view decreases as the angular distance between the novel view and the closest study view increases (e.g., Cooper, 1975; Jolicœur, 1988; Tarr & Pinker, 1989, 1990). In contrast, the structural description models have found support in findings that object recognition can be viewpoint independent when it relies on distinctive features that are also viewpoint independent (e.g., the number of parts; see Biederman & Gerhardstein, 1993). Recent studies have shown that object recognition can rely on both structure and view information (e.g., Foster & Gilson, 2002).

In this project, we proposed and tested an intrinsic model of object representation and recognition derived from the intrinsic model of spatial memory (Mou, Fan, McNamara, & Owen, 2008; see also Mou & McNamara, 2002; Mou, McNamara, Valiquette, & Rump, 2004), which claims that interobject spatial relations in a layout of objects are represented with respect to the intrinsic orientation¹ (intrinsic reference direction) of the layout. The study location of the observer is also represented with respect to the intrinsic orientation of the layout. Recognizing a previously learned configuration requires the observer to (1) identify the intrinsic orientation of the test scene, (2) align the intrinsic orientation of the test scene with the represented intrinsic orientation of the layout, and (3) compare the spatial rela-

W. Mou, wmou@ualberta.ca

tions in the test scene with the represented spatial relations of the layout.

Mou et al. (2008) demonstrated that scene recognition was dependent on both intrinsic orientation and viewpoint. Participants were instructed to learn a layout of objects along an intrinsic axis that was different from their viewing direction, and were then given a scene recognition task in which they had to distinguish triplets of objects from the layout at different views from mirror images of those test scenes. The results showed that recognition latency increased with the distance between the test view and the study view. More importantly, participants were able to recognize intrinsic triplets of objects, which contained two objects parallel to the intrinsic axis that participants were instructed to use when they remembered the layout, faster than they did nonintrinsic triplets of objects, which did not contain two objects parallel to the instructed intrinsic axis. This pattern occurred for all test views. The intrinsic-orientation effect can be explained because participants were quicker in identifying the intrinsic orientation for the intrinsic triplets than for the nonintrinsic triplets. The view effect can be explained because participants had to align the intrinsic orientation in the test triplet with the represented intrinsic orientation of the layout in memory, when the test view was different from the study view, but did not need such alignment when the test view was the same as the study view.

We generalized the intrinsic model from scene recognition to object recognition because two observations suggested that these two visual processes share very similar mechanisms. First, it has been shown that both object and scene recognition rely on the identification of intrinsic orientation. As described previously, identifying intrinsic orientation is important in scene recognition (e.g., Mou et al., 2008). Similarly, it has been reported that identifying intrinsic orientation is important for shape (object) perception and recognition (e.g., Hinton, 1979; Hinton & Parsons, 1988; Palmer, 1989; Rock, 1973). For example, a square tilted 45° can be seen as either a tilted square or an upright diamond, depending on whether an edge or a vertex is identified as the top. Second, it has been shown that both object and scene recognition are viewpoint dependent. As reviewed previously, performance in object recognition is better at the familiar views and decreases as the angular distance between the novel view and the closest familiar view increases (e.g., Cooper, 1975; Jolicœur, 1988; Tarr & Pinker, 1989, 1990). Similarly, it has been reported that performance in scene recognition is better at the familiar views and decreases as the angular disparity between the novel view and the closest study view increases (e.g., Burgess, Spiers, & Paleologou, 2004; Christou & Bühlhoff, 1999; Diwadkar & McNamara, 1997).

In sum, the intrinsic model of object recognition claims that people represent the spatial structure of an object and their own study viewpoint with respect to the intrinsic orientation of the object. When they recognize an object, they first need to identify the intrinsic reference direction of the object, then align the intrinsic reference direction of the test object with the represented intrinsic reference direction established from the learned viewpoint (e.g.,

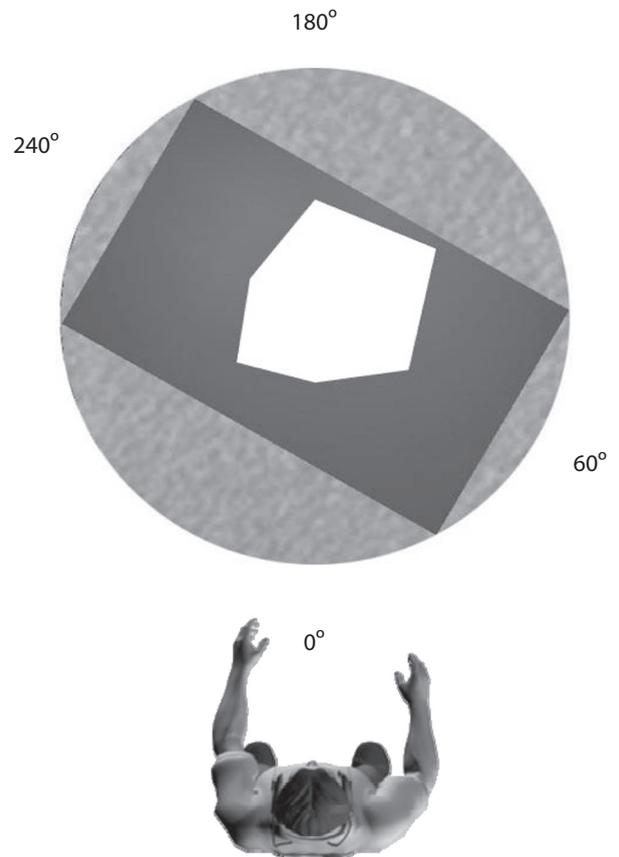


Figure 1. One sample of the study shapes.

Ullman, 1989). According to this model, both intrinsic-orientation-dependent and viewpoint-dependent object recognition are expected. However, to the best of our knowledge, previous investigations of object and shape recognition have not dissociated viewpoint dependence from intrinsic-orientation dependence. Two experiments in this project were conducted to seek such dissociation in shape recognition.

EXPERIMENT 1

In Experiment 1, participants learned shapes of irregular hexagons surrounded by a rectangular background on a round table (Figure 1). The longer axis of the rectangular background was 60° different from the study viewpoint. The study viewpoint was arbitrarily labeled as 0° , and all other directions were labeled counterclockwise from the study viewpoint. Target test shapes were created by splitting study shapes along a 0° – 180° axis or along a 60° – 240° axis and presenting them at the study view (view of 0°) or at a novel view (view of 30°) by rotating shapes 30° clockwise (Figure 2). The purpose of this experiment was to examine both the intrinsic-orientation effect (split along 0° – 180° vs. 60° – 240°) and the view effect (familiar view vs. novel view) on participants' recognition of the target test shapes. We assumed that the intrinsic orientation of a shape was determined by the external rectangular back-

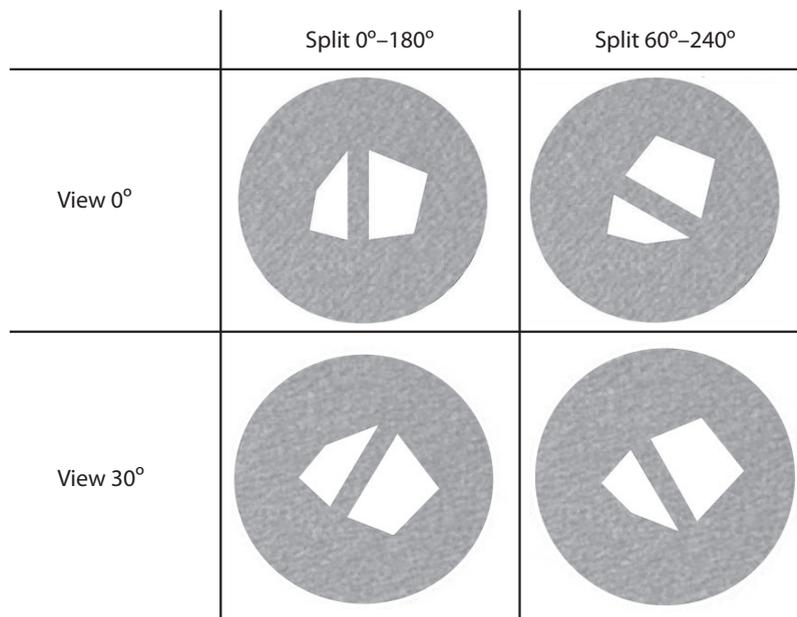


Figure 2. Sample of the target test shapes in the combination of splitting axis (0°–180°, 60°–240°) and test view (0°, 30°).

ground (e.g., Palmer, 1989; Pani & Dupree, 1994). Hence, we expected that shape recognition should be better when the shape was split along the axis of 60°–240° than along the axis of 0°–180°.

Method

Participants. Twelve university students (6 women, 6 men) participated in return for monetary compensation.

Materials and Design. The experiment was conducted in a room (6.0 × 6.0 m) with walls covered in black curtains. The virtual environment with shapes, rectangular backgrounds, and round tables was displayed on the room floor in stereo with an I-glasses PC/SVGA Pro 3D head-mounted display (HMD, I-O Display Systems, Inc., Sacramento, CA). Participants sat in chairs 51 cm high, 180 cm from the table center. Their head movements were tracked with an InterSense 900 motion tracking system (InterSense Inc., Bedford, MA). As illustrated in Figure 1, the study hexagon was presented with a rectangular background (60.6 cm × 100 cm) that fit inside the round table (diameter of 117 cm), and tilted 60° from participants' viewpoint.

Twelve study hexagons were created with the following constraints: The six vertices of the hexagons were located in the directions of 0°, 60°, 180°, 120°, 240°, and 300° from the center of the table; the distances between the vertices and the center of the table were randomized within the range from 15 cm to 35 cm and different from each other, such that the hexagons were not symmetric along any axis. As shown in Figure 2, four target test shapes were created from each study hexagon by splitting the study hexagon into two quadrilaterals 4 cm apart from each other along the axis of 0°–180°, or the axis of 60°–240°, and presenting the quadrilaterals from the view of 0° or from the view of 30° by rotating the quadrilaterals 30° clockwise. Four distractor test shapes were created for each study hexagon by mirror-reflecting the target test shapes about the viewing axis at test. The test shapes were presented only on the round table with the removal of the rectangular background.

Each study hexagon was followed by eight test shapes (four targets and four distractors). Participants were required to judge whether each test shape was the same as the study shape, regardless of the view change. Participants were told to imagine that there was

no gap between the quadrilaterals. Accordingly, there were 12 blocks of trials corresponding to 12 study hexagons; for each block there were eight recognition trials. The 12 study hexagons and the eight test shapes for each study hexagon were presented in random order.

The two independent variables were splitting axis (0°–180° vs. 60°–240°) and test view (0° vs. 30°). The dependent variables were response accuracy and latency. Response latency was measured as the time from the presentation of the test shapes to the response of the participant.

Procedure. Before entering the experiment room, each participant was first instructed to get to know the experiment procedure and requirements, then blindfolded and seated in the chair at the viewing position. All the lights in the room were turned off. The blindfold was removed and the participant, holding a mouse in the right hand, wore the head-mounted display.

Each study hexagon was displayed with the rectangular background on the round table for 60 sec. Participants were required to remember the shape. Then, eight test shapes were presented sequentially. Prior to the presentation of each test shape, only a red cross was presented at the center of the table for 4 sec and participants were required to fixate the cross. Each test shape was displayed on the round table until participants pressed the mouse buttons (left button for targets and right button for distractors). Participants were required to respond as quickly and accurately as possible. There were four extra shapes for practice.

Results and Discussion

Mean response latency for correct responses to targets is plotted in Figure 3 as a function of splitting axis and test view. As shown in Figure 3, participants recognized shapes quicker when the shapes were split along the axis of 60°–240° than when the shapes were split along the axis of 0°–180°. Participants recognized shapes quicker at the study view than at the view of 30°. Mean response latencies per participant per condition were analyzed in repeated measures ANOVAs with terms for splitting axes and test view. The main effect of splitting axis was significant [$F(1,11) = 4.92, p < .05, MS_e = 4.43$]. The main

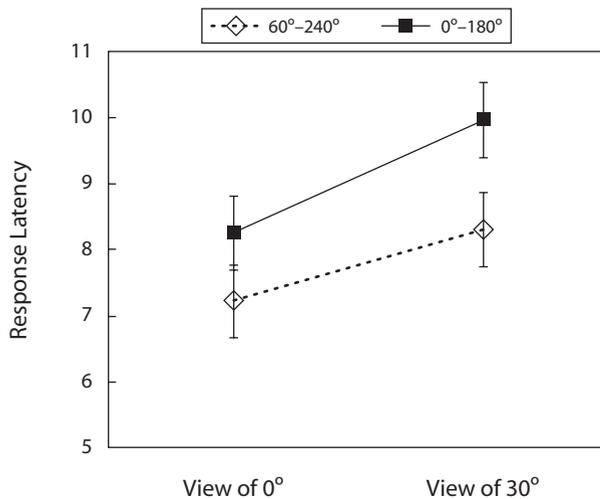


Figure 3. Mean response latency in correctly recognizing target shapes as a function of splitting axis and test view in Experiment 1. Error bars are confidence intervals corresponding to ± 1 standard error, as estimated from the ANOVA.

effect of test view was significant [$F(1,11) = 9.46, p < .05, MS_e = 2.44$]. The interaction between splitting axis and test view was not significant [$F(1,11) = 0.29, p > .05, MS_e = 4.12$].

Mean accuracy for targets is listed in Table 1 as a function of splitting axis and test view. Mean response accuracies per participant per condition were analyzed in repeated measures ANOVAs with terms for splitting axis and test view. The test view was the only significant effect [$F(1,11) = 12.0, p < .01, MS_e = .028$]. Participants recognized shapes more accurately at the study view than at the view of 30°. Mean accuracy of responses to distractors is listed in Table 2 as a function of splitting axis and test view. The ANOVA revealed no significant effects, indicating that effects in accuracy for targets were not caused by response biases.²

This experiment successfully dissociated the intrinsic-orientation effect and the view effect. Participants determined the intrinsic orientation of the shape using the longer axis of the rectangular background.

EXPERIMENT 2

The purpose of Experiment 2 was to investigate whether the intrinsic-orientation effect could be dissociated from the view effect when no rectangular background was presented with the study shapes. We hypothesized that under such conditions the intrinsic orientation of the study shape would be determined by participants' study viewpoint (e.g., Rock, 1973; Shelton & McNamara, 1997). Hence, we predicted that shape recognition should be better when the shape was split along the axis of 0°-180° than along the axis of 60°-240°.

Method

Participants. Twelve university students (6 women, 6 men) participated in return for monetary compensation.

Materials, Design, and Procedure. The materials, design, and procedure were identical to those of Experiment 1, except that no rectangular background was presented during learning.

Results and Discussion

As shown in Figure 4, participants were quicker in recognizing target test shapes when the study hexagons were split along the axis of 0°-180° than along the axis of 0°-240° [$F(1,11) = 5.01, p < .05, MS_e = 5.03$]. Participants were quicker in recognition at the view of 0° than at the view of 30° [$F(1,11) = 15.02, p < .01, MS_e = 1.81$]. The interaction between splitting axis and test view was not significant [$F(1,11) = 0.18, p > .05, MS_e = 3.71$]. As shown in Table 1, participants were more accurate in recognizing the target shapes at the view of 0° than at the view of 30° [$F(1,11) = 8.06, p < .05, MS_e = 0.04$]. No other effects were significant. As shown in Table 2, no effects were significant in recognizing the distractor shapes. Hence, without the influence of the rectangular background, we still

Table 1
Mean Accuracy and Standard Deviation (in Percentage) in Recognizing Target Shapes As a Function of Splitting Axis and Test View in the Experiments

Experiment	0°-180°				60°-240°			
	View 0°		View 30°		View 0°		View 30°	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	0.64	0.21	0.50	0.13	0.72	0.13	0.53	0.25
2	0.75	0.17	0.55	0.23	0.72	0.21	0.59	0.21

Table 2
Mean Accuracy and Standard Deviation (in Percentage) in Recognizing Distractor Shapes As a Function of Splitting Axis and Test View in the Experiments

Experiment	0°-180°				60°-240°			
	View 0°		View 30°		View 0°		View 30°	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	0.57	0.21	0.59	0.21	0.48	0.19	0.56	0.23
2	0.53	0.19	0.57	0.17	0.47	0.22	0.51	0.24

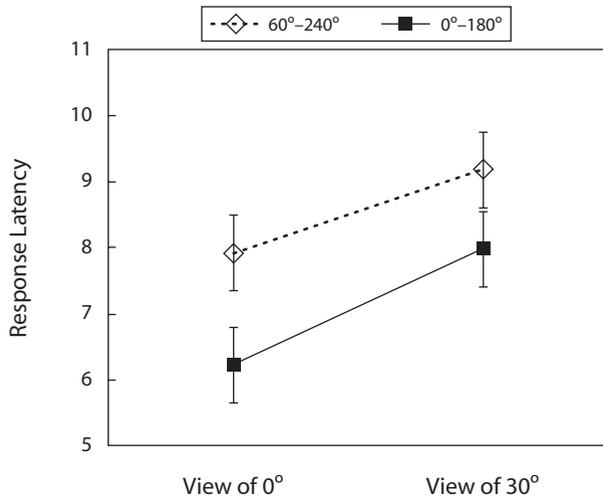


Figure 4. Mean response latency in correctly recognizing target shapes as a function of splitting axis and test view in Experiment 2. Error bars are confidence intervals corresponding to ± 1 standard error, as estimated from the ANOVA.

dissociated the intrinsic-orientation effect and the view effect, and the intrinsic orientation of the hexagons was determined by the study viewpoint of the participants.

GENERAL DISCUSSION

This project successfully dissociated the intrinsic-orientation effect from the view effect in shape recognition. In Experiment 1, the hexagons were presented with one intrinsic axis parallel to the study viewpoint and another intrinsic axis parallel to the orientation of an external rectangular background. The results showed that participants were quicker in recognizing the target shapes when the study hexagons were split along the intrinsic axis parallel to the orientation of the rectangular background than when the study hexagon was split along the intrinsic axis parallel to the study viewpoint, and were quicker at the same view than at the novel view. The intrinsic-orientation effect and the view effect were independent. In Experiment 2, the external rectangular background was removed during learning. The results showed that participants were quicker to recognize the target shapes when the study hexagons were split along the intrinsic axis parallel to the study viewpoint than when the study hexagons were split along the intrinsic axis not parallel to the study viewpoint, and were quicker at the same view than at the novel view. The intrinsic-orientation effect and the view effect were independent.

Although view-based models of object recognition (e.g., Tarr & Pinker, 1989, 1990; Ullman, 1989) can readily explain the view effect, they have difficulty explaining the intrinsic-orientation effect. The view-based models predict that the shapes presented from the same test view (both the study view and the novel view) should be equivalent, regardless of the intrinsic axis along which the shape was split.

These findings, however, can be explained by the intrinsic model proposed above. According to the model, participants needed to identify the intrinsic orientation of

the test shape in shape recognition. We assume that the intrinsic orientation of a test shape was more salient when the study shape was split along the intrinsic orientation than when the shape was split along another intrinsic axis. Hence, identifying the intrinsic orientation was easier in the former condition than in the latter condition, leading to the intrinsic-orientation effect. Because participants represented their study viewpoint with respect to the intrinsic orientation of the shape, when they were tested at the study position and the test view was different from the study view, the test shape was oriented differently from the mental representation of the study shape. Participants needed to align these two intrinsic orientations leading to the view effect. We acknowledge that the intrinsic model is similar to the structural description models (e.g., Biederman, 1987; Marr & Nishihara, 1978), in that it claims that object representation is object-centered but different from structure description models that can also explain the view effect.

The dissociation of the intrinsic-orientation effect and the view effect in object recognition in this project echoes the dissociation of the intrinsic-orientation effect and the view effect in scene recognition reported by Mou et al. (2008; see also Schmidt & Lee, 2006). We acknowledge that in both studies, participants' recognition decisions were primarily based on spatial structure rather than on unique features (e.g., number of parts), because the distractors were the mirror images of the shape or the scene; hence, at least recognition of spatial structures of shapes and that of scenes seem to share similar mechanisms. Such similarity implies that a unified model of visual representation and recognition may be possible to accommodate both object and scene recognition (e.g., Christou & Bülthoff, 1999; Diwadkar & McNamara, 1997).

AUTHOR NOTE

Preparation of this article and the research reported in it were supported in part by the 973 Program of Chinese Ministry of Science and Technology (2006CB303101) and by a grant from the Natural Sciences and Engineering Research Council of Canada to W.M. and National Institute of Mental Health Grant 2-R01-MH57868 to T.P.M. We thank Michael Brown, Thomas Schmidt, and two anonymous reviewers for their helpful comments on a previous version of the manuscript. Please address correspondence to W. Mou, Department of Psychology, University of Alberta, Edmonton, AB, T6G 2E9 Canada (e-mail: wmou@ualberta.ca).

REFERENCES

- BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147. doi:10.1037/0033-295X.94.2.115
- BIEDERMAN, I., & GERHARDSTEIN, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, *19*, 1162-1182. doi:10.1037/0096-1523.19.6.1162
- BURGESS, N., SPIERS, H. J., & PALEOLOGOU, E. (2004). Orientational manoeuvres in the dark: Dissociating allocentric and egocentric influences on spatial memory. *Cognition*, *94*, 149-166. doi:10.1016/j.cognition.2004.01.001
- CHRISTOU, C. G., & BÜLTHOFF, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, *27*, 996-1007.
- COOPER, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*, 20-43. doi:10.1016/0010-0285(75)90003-1

- DIWADKAR, V. A., & MCNAMARA, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, **8**, 302-307. doi:10.1111/j.1467-9280.1997.tb00442.x
- FOSTER, D. H., & GILSON, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society B*, **269**, 1939-1947. doi:10.1098/rspb.2002.2119
- HAYWARD, W. G. (2003). After the viewpoint debate: Where next in object recognition? *Trends in Cognitive Sciences*, **7**, 425-427. doi:10.1016/j.tics.2003.08.004
- HINTON, G. E. (1979). Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*, **3**, 231-250. doi:10.1016/S0364-0213(79)80008-7
- HINTON, G. E., & PARSONS, L. M. (1988). Scene-based and viewer-centered representations for comparing shapes. *Cognition*, **30**, 1-35. doi:10.1016/0010-0277%2888%2990002-9
- JOLICŒUR, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology*, **42**, 461-478.
- MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of spatial-organization of three-dimensional shapes. *Proceedings of the Royal Society B*, **200**, 269-294.
- MOU, W., & MCNAMARA, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 162-170. doi:10.1037/0278-7393.28.1.162
- MOU, W., FAN, Y., MCNAMARA, T. P., & OWEN, C. B. (2008). Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition*, **106**, 750-769. doi:10.1016/j.cognition.2007.04.009
- MOU, W., MCNAMARA, T. P., VALIQUETTE, C. M., & RUMP, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 142-157. doi:10.1037/0278-7393.30.1.142
- PALMER, S. E. (1989). Reference frames in the perception of shape and orientation. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 121-163). Hillsdale, NJ: Erlbaum.
- PANI, J. R., & DUPREE, D. (1994). Spatial reference systems in the comprehension of rotational motion. *Perception*, **23**, 929-946. doi:10.1068/p230929
- ROCK, I. (1973). *Orientation and form*. New York: Academic Press.
- SCHMIDT, T., & LEE, E. Y. (2006). Spatial memory organized by environmental geometry. *Spatial Cognition & Computation*, **6**, 347-369. doi:10.1207/s15427633scc0604_4
- SHELTON, A. L., & MCNAMARA, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, **4**, 102-106.
- TARR, M. J., & PINKER, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282. doi:10.1016/0010-0285%2889%2990009-1
- TARR, M. J., & PINKER, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, **1**, 253-256. doi:10.1111/j.1467-9280.1990.tb00209.x
- ULLMAN, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254. doi:10.1016/0010-0277%2889%2990036-X

NOTES

1. Intrinsic orientation is used here to indicate an orientation perceived or identified by a viewer for a specific scene, rather than to imply an orientation determined by the intrinsic property of the scene; for example, symmetric axis. An intrinsic orientation can be identified with different kinds of cues, including viewpoints, intrinsic properties of the scene, local or global contexts, and so on.
2. The results when *d*-prime is used as a dependent variable are the same as those when accuracy of the response to the target shapes is used as a dependent variable in both Experiments 1 and 2.

(Manuscript received October 6, 2008;
revision accepted for publication January 17, 2009.)