Research Article

VIEWPOINT DEPENDENCE IN VISUAL AND HAPTIC OBJECT RECOGNITION

Fiona N. Newell,[1] Marc O. Ernst,[2] Bosco S. Tjan,[3] and Heinrich H. Bülthoff[2]

[1]University of Durham, Durham, England; [2]Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany; and [3]NEC Research Institute

Abstract—*On the whole, people recognize objects best when they see the objects from a familiar view and worse when they see the objects from views that were previously occluded from sight. Unexpectedly, we found haptic object recognition to be viewpoint-specific as well, even though hand movements were unrestricted. This viewpoint dependence was due to the hands preferring the back "view" of the objects. Furthermore, when the sensory modalities (visual vs. haptic) differed between learning an object and recognizing it, recognition performance was best when the objects were rotated back-to-front between learning and recognition. Our data indicate that the visual system recognizes the front view of objects best, whereas the hand recognizes objects best from the back.*

People explore and navigate through their environment mainly using sight and touch. In order to guide actions and interactions with objects, information acquired from the visual and the haptic systems must converge to form a coherent percept. What might be the nature of the representations underlying each sensory system, in order to allow this convergence? If the visual and haptic representations of an object are qualitatively different, a translation process must be involved for the two systems to communicate. The presence of a translator implies that moving information between the visual and haptic systems can be inefficient. In the experiments reported in this article, we studied the nature of object representation in each sensory system and the interaction between these systems. Specifically, we considered whether representations in each system are either dependent on or invariant to viewpoint.

Recognition occurs when the percept of an object matches a stored representation in memory (Bülthoff & Edelman, 1992). The idea that visual recognition performance is viewpoint-specific has been well established (Jolicoeur, 1985; Palmer, Rosch, & Chase, 1981). For unfamiliar objects, recognition performance is best when the objects are shown in views in which they were learned (Edelman & Bülthoff, 1992; Rock & DiVita, 1987). Even a familiar object (e.g., a dog) is recognized more efficiently when it is seen in the most typical (e.g., upright) position (Jolicoeur, 1985; Newell & Findlay, 1997). When objects are recognized independently of view, it is generally because all views of the objects are familiar (Tarr & Bülthoff, 1995) or the object contains very distinct parts (Biederman, 1987). View-dependent visual recognition performance has been found in humans and other primates (Logothetis, Pauls, Bülthoff, & Poggio, 1994). Also, neurophysiological studies have revealed cells in inferior temporal cortex that are maximally tuned to specific views of objects (Logothetis, Pauls, & Poggio, 1995). The findings from these and other studies have led researchers to speculate that objects in visual memory are represented in a view-specific manner (Tarr & Bülthoff, 1995).

To date, however, the nature of object recognition in the haptic system has received relatively less attention (Lederman & Klatzky, 1987; Lederman, Klatzky, Chataway, & Summers, 1990). Some researchers have suggested that there are large representational similarities between the visual and haptic systems (Easton, Green, & Srinivas, 1997). Indeed, many of these studies have reported good cross-modal recognition performance using implicit measures, suggesting that object representations are easily shared between the haptic and visual systems (Easton et al., 1997; Reales & Ballesteros, 1999). The question arises, however, as to whether these representations are mediated in a view-dependent or view-invariant manner.[1]

At any one time, the eyes can see an object from only one view, so that certain features of an object can be occluded from sight. This can lead to a general dependency on viewpoint. In contrast, when one handles an object, the thumbs and fingers contact it from different sides simultaneously (Gibson, 1962). It therefore seems intuitive that the haptic representation of objects would be omnidirectional and not viewpoint-specific. It has been noted that for the purpose of recognition, the hands typically follow the contour of a three-dimensional object until the object is recognized (Lederman & Klatzky, 1987). According to Lederman and Klatzky, this contour-following exploration strategy is necessary for haptic object recognition. What is not clear, however, is whether contour following in haptic object recognition necessarily leads to an omnidirectional representation of objects in the haptic system.

Arguments for a viewpoint-specific haptic representation of objects often arise from the observation that common motor tasks, such as grasping or manipulating objects, require information about an object's position and orientation relative to the observer. However, such motor tasks are usually guided by vision, and therefore all the necessary information is available visually. That is, it is generally unnecessary to invoke any haptic representation or even visual identification of the objects (Goodale et al., 1994; Goodale, Milner, Jakobson, & Carey, 1991). Even for tasks in which visual information is not available, such as reading Braille or playing musical instruments, orientation and position information can, in theory, be attached to an omnidirectional, viewpoint-independent haptic representation (Kennedy, 1993; Millar, 1997). Consequently, a viewpoint-specific

---

Address correspondence to Fiona Newell, Department of Psychology, Áras an Phairsaigh, University of Dublin, Trinity College, Dublin 2, Ireland; e-mail: fnewell@tcd.ie.

[1] In previous studies on cross-modal recognition of three-dimensional objects, the nature of the input across both senses was often not balanced. Generally, visual presentation allowed viewing the object from a single viewpoint, whereas in the haptic condition, the objects were not fixed in space and were freely manipulated by the hands (Easton et al., 1997; Lederman et al., 1990). Although these and other researchers proposed that cross-modal recognition is mediated by abstract, structural descriptions, we argue that such studies cannot directly address the issue of the nature of object representation in either sensory system because the sensory input to the two systems was not equivalent.

haptic representation is not needed. If the haptic representation of objects is indeed omnidirectional, and if the task is object recognition, which does not require reporting an object's orientation and position, we should expect performance to be viewpoint-independent. To our surprise, however, this is not what we found. We conducted three experiments in which the effects of viewpoint were measured for visual, haptic, and cross-modal object recognition.

EXPERIMENT 1

Method

Participants

Twenty-six undergraduate students from the Department of Psychology, University of Durham, in Durham, England, participated in the experiment for pay. Sixteen of the participants were female. Participants' ages ranged from 19 to 30 years old.

Apparatus

We used a set of unfamiliar objects made from six identical red LEGO bricks. Each object was constructed in a unique configuration of these bricks. Therefore, only the shape—and not the color, texture, or weight—of the objects played a role in visual and haptic recognition (see Klatzky, Lederman, & Reed, 1987). Figure 1 shows several examples of the test objects. Each object was placed on a stand so that the object was in a fixed orientation with respect to the observer.

Design

Thirty-two uniquely configured objects were used as stimuli. The experiment was based on a three-way, repeated measures design with learning modality (visual or haptic), transfer at test (within or across modalities), and view change at test (0° or 180°) as factors. The experiment contained four experimental blocks, defined by the modalities under which the objects were learned and tested. The objects were distributed, at random, across the experimental blocks (resulting in a set of eight objects per block). Each object set was counterbalanced across blocks.

Within each block, an object was randomly assigned as either a target or a distractor; thus, there were four target objects and four distractors. During learning, only the four target objects were presented to the participants, each for 30 s in the visual learning condition or 1 min for the haptic learning condition. These presentation durations were chosen to yield equivalent recognition accuracy across the two modalities as determined by a pilot experiment. During test, participants were presented with four target objects and four distractor objects in random order. Presentation time during test was unlimited. There were 12 test trials in each block: We repeated 4 trials in order to prevent participants from guessing by elimination. The results from these repeated trials were discarded.

For all conditions, participants remained seated in front of a table. No restriction on hand or head movement was imposed. Each object was presented to a subject by fixing it on a 15-cm-tall stand on the table. Presentation was behind a curtain in the haptic condition.

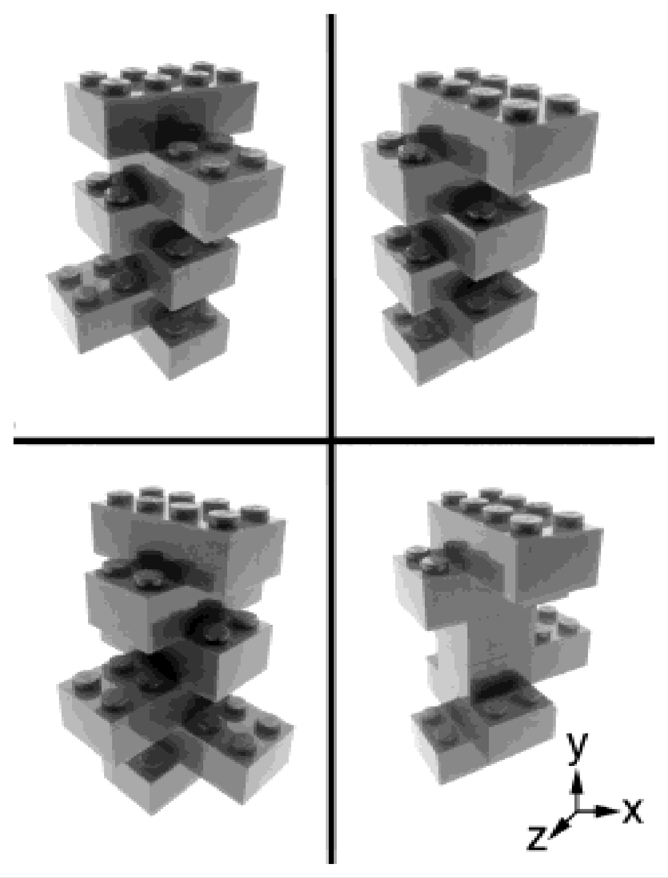

Fig. 1. Examples of the objects used in our recognition experiments. The coordinate system is shown in the lower right. In Experiments 1 and 3, we tested rotations about the *y*-axis (vertical), whereas in Experiment 2, we used rotations about the *x*- and *z*-axes.

Procedure

There were four separate blocks in the experiment, and within each block participants were required to learn four target objects in a sequential order, either visually or haptically using both hands. Participants were not given any explicit instructions on how to learn the objects. They were free to move their hands around the objects during haptic exploration and their head during visual exploration; thus, all surfaces of the objects could be perceived whichever modality was used for exploration. However, participants were instructed not to move the objects or to walk around them. During the subsequent test session (which immediately followed the learning session), four new objects were added to the set of the four learned objects. Participants were instructed to decide if each object presented either was from the learning set or was a distractor object.

Recognition was tested either in the same modality as learning or in the other modality. All possible combinations were tested in the four blocks: visual learning and visual testing (visual-visual), haptic learning and haptic testing (haptic-haptic), visual learning and haptic testing (visual-haptic), and haptic learning and visual testing (haptic-visual). Furthermore, participants were informed that each target object could appear either in the same position as it was learned (0° view) or rotated by 180° around a predefined axis of rotation. In this experiment, we tested rotations about the *y*-axis (vertical; see Fig. 1).

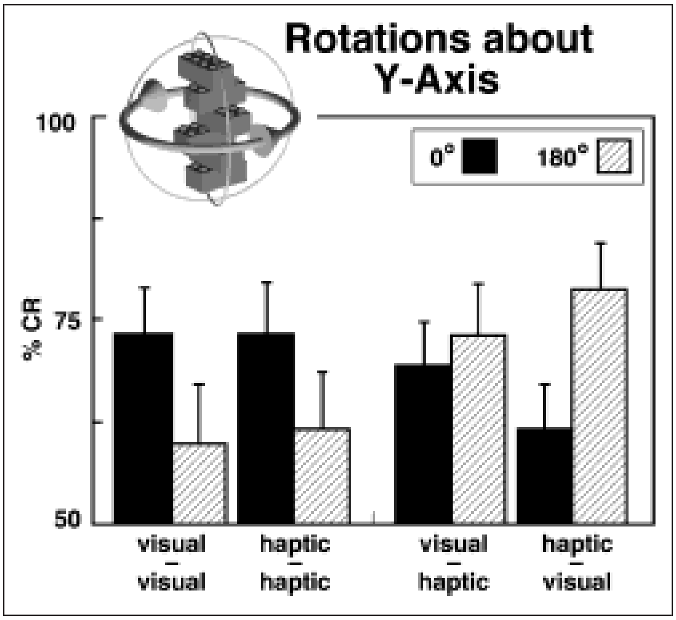

Fig. 2. Recognition performance (percentage of correct responses, or %CR) in Experiment 1. Learning and testing were conducted either within the same modality (visual-visual or haptic-haptic) or across the modalities (visual learning and haptic testing, visual-haptic, or haptic learning and visual testing, haptic-visual). The view of the objects either changed between learning and test (180°) or did not change (0°). Error bars denote the standard errors of the mean across participants.

Results and Discussion

Figure 2 shows the percentage of correct responses made to the learned objects for all four learning-and-testing conditions. Performance was above chance throughout the experiment, $Z = 3.6$, $p < .01$. We ran a three-way analysis of variance (ANOVA) on the number of correct responses to the targets, using transfer (within or across modality), learning modality (visual or haptic), and viewpoint (0° or 180°) as factors. We found no significant main effect for transfer, $F(1, 25) = 1.205$; learning, $F < 1$; or viewpoint, $F < 1$. However, we found a significant interaction between transfer and viewpoint, $F(1, 25) = 6.755$, $p < .05$. Simple effects analyses revealed an almost significant effect of viewpoint within modalities, $F(1, 25) = 3.261$, $p = .083$, and a significant effect of viewpoint across modalities, $F(1, 25) = 4.416$, $p < .05$.

When the task was conducted within modalities (visual-visual and haptic-haptic), recognition performance was about 75% correct when there was no change in view (0°). When the object was rotated by 180° around the vertical axis, the recognition performance was almost significantly reduced to around 60%. This result shows that recognition within both the visual and the haptic domains depends on the view of the object, suggesting that object representation is not omnidirectional in either domain. More interestingly, the opposite effect of rotation was found when there was a change of modality between learning and testing. Recognition performance was better ($p < .05$) when the test object was rotated about the vertical axis (180°) than when it remained in the learning position (0°). Furthermore, there was no main effect on performance due to a change in modality per se, suggesting that no significant loss resulted when shape information had to be transferred between modalities.

We noticed that when the participants explored an object during the haptic learning or testing sessions, the fingers of both hands typically felt the back of the object, whereas only the thumbs contacted the front. We surmised that information integrated across the fingers might yield a better representation of the surface of the object than information gathered from the thumbs alone. This would be analogous to the visual system having a better representation of the front, and therefore visible, surface of an object (Fig. 3). In sum, the back side of a hand-sized object is often more accessible than the front to the haptic system, whereas the front is more accessible than the back to the visual system. When an object is learned in one modality and tested in another, information must be transferred between the two. Performance should be better when both modalities sense the same surface. We achieved this in the cross-modality condition by rotating the object 180° about the vertical axis.

We conducted a second experiment to test our prediction that recognition performance in the cross-modal conditions would be better when there was an exchange of the front and back surfaces of an object between learning and testing sessions, compared with when the front and back surfaces were not exchanged. We studied rotation about two additional axes: One axis of rotation involved an exchange of the front and back surfaces, and the other axis of rotation did not.

EXPERIMENT 2

Method

The procedure and the materials in Experiment 2 were identical to those of Experiment 1 with one important exception. In this experiment, we tested two orientation changes different from the rotation tested in Experiment 1. In Experiment 2, in separate sessions, the target objects were rotated around the *x*-axis (i.e., the horizontal axis) and around the *z*-axis (i.e., the depth axis). Rotations around the *x*-axis involve an exchange of front and back surfaces. The recognition of targets rotated around the *x*-axis was tested in one session, which we refer to as Experiment 2a. The front and back of the targets

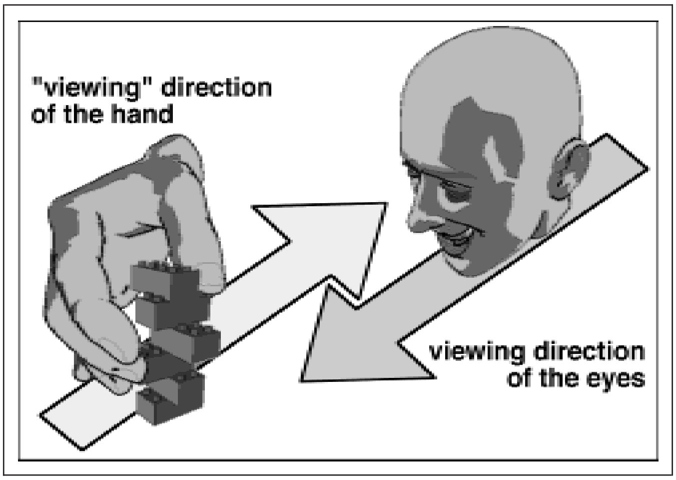

Fig. 3. Schematic diagram of our model, which suggests that information integrated across the fingers is analogous to seeing an object from behind.

remained in the same position with rotations around the *z*-axis. The recognition of targets rotated around the *z*-axis was tested in the other session, Experiment 2b. Thirty-two undergraduate students from the Eberhard-Karls University of Tübingen, Germany, participated in both sessions of the experiment for pay. Target objects and distractor objects were counterbalanced across sessions. The order of the testing session was randomized across the participants. Participants received a self-timed break between sessions.

Results and Discussion

Figure 4 shows the recognition performance for the objects rotated around these two axes. The results confirmed our prediction. Performance was above chance in both Experiment 2a, $Z = 5.32$, $p < .01$, and Experiment 2b, $Z = 5.22$, $p < .01$. We conducted separate three-way ANOVAs on the correct responses made in the two sessions of the experiment. In Experiment 2a, we found no significant main effects for transfer, $F < 1$; for learning modality, $F < 1$; or for viewpoint, $F < 1$. However, we found a significant interaction between transfer and viewpoint, $F(1, 35) = 12.530$, $p < .01$. Simple effects analyses on the interaction revealed a significant effect of viewpoint within modalities, $F(1, 35) = 9.833$, $p < .01$, and a significant effect of viewpoint across modalities, $F(1, 35) = 5.736$, $p < .05$.

In Experiment 2b, we found no main effects for transfer, $F < 1$, or for learning modality, $F < 1$, but we found a significant main effect of viewpoint, $F(1, 35) = 8.448$, $p < .01$. There were no interactions between the factors.

In each session, recognition within each modality was viewpoint-dependent; that is, recognition performance was reduced if a test object was rotated relative to its learning view. When recognition was tested across modalities, performance improved with rotation around the horizontal (*x*) axis (i.e., when a front-back exchange of surfaces was present), but not with rotation around the depth (*z*) axis (i.e., when a front-back exchange of surfaces was absent). Thus, the results of this experiment support our prediction that different viewpoints allow better integration of information across the visual and haptic modalities than the same viewpoint does.

Our findings so far suggest that the hand prefers to explore the back surface of the object for recognition purposes. However, the question remains as to whether the back surface was preferred because of the biomechanical design of the hand, or if any surface can be used for recognition by the hand. In Experiment 3, we restricted hand exploration to one surface in the learning session, either the front or the back surface, and tested recognition performance with each of these surfaces. We expected that if any surface of the object can be used for recognition, we would find equally good recognition performance with the fronts and backs of the objects. Alternatively, if the back surface is the preferred haptic “view” of an object, then recognition performance would be relatively better for the backs than the fronts of the objects.

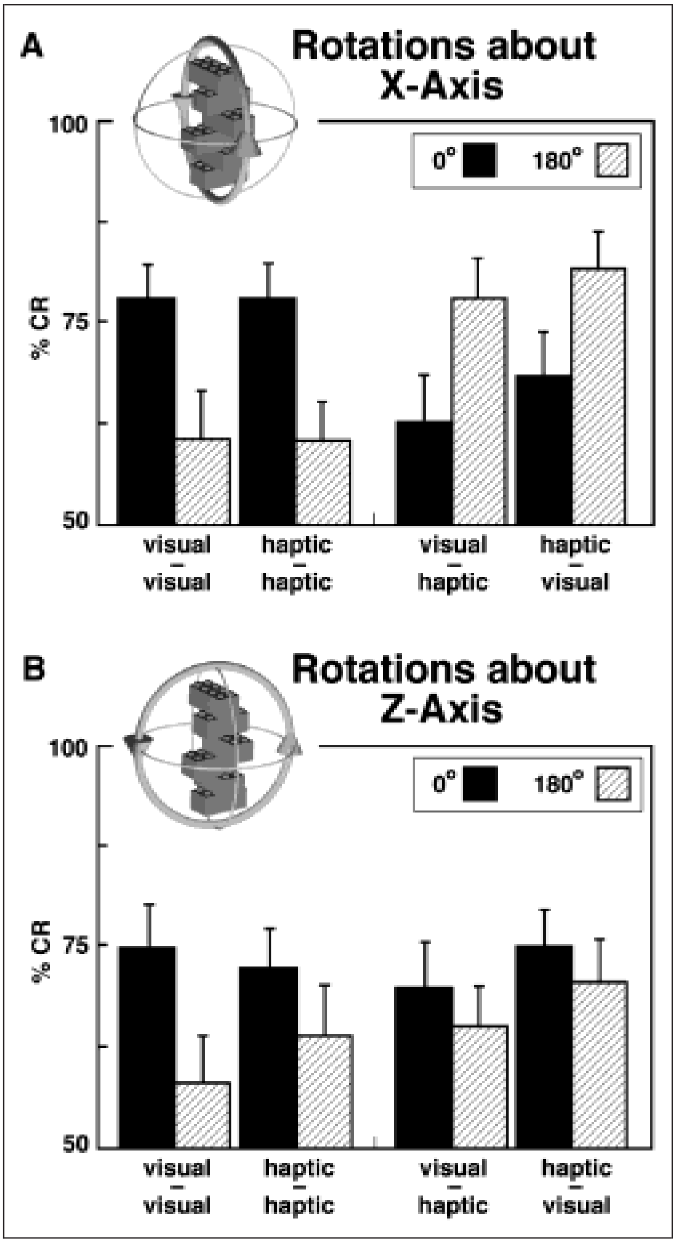

Fig. 4. Recognition performance (percentage of correct responses, or %CR) in Experiment 2. Learning and testing were conducted either within the same modality (visual-visual or haptic-haptic) or across the modalities (visual learning and haptic testing, visual-haptic, or haptic learning and visual testing, haptic-visual). In Experiment 2a (a), the objects were rotated about the horizontal (*x*) axis, which involved an exchange of the front and back of the objects. In Experiment 2b (b), the objects were rotated about the depth (*z*) axis, which did not involve an exchange of the front and back of the objects.

EXPERIMENT 3

Method

The procedure and the materials in Experiment 3 were identical to those of the within-modality, haptic condition of Experiment 1; however, in this experiment, a cover on each object restricted finger exploration to one surface of the object. Furthermore, in this experiment we ran two separate blocks, in which participants learned the objects from either the front or the back surface. Participants were allowed to explore the objects freely with the constraint that the thumbs could not be used during exploration. In one block, participants learned all of the four target objects from the front, and in the other block, all targets were learned from the back surface. The order of the blocks was counterbalanced across participants. Target objects and distractor objects were counterbalanced across participants. As in the previous experiments, participants were informed that during the test session the target objects could be presented in either the same position as in the learning session or rotated by 180° about the *y*-axis. The cover was used throughout the entire experiment (i.e., during learning and testing). The orientation of the objects remained fixed within the cover; therefore, the object and the cover were treated together as an entire object during rotation. Thirty-two undergraduate students from the Eberhard-Karls University of Tübingen, Germany, participated in both blocks of the experiment for pay.

Results and Discussion

Figure 5 shows the recognition performance across the different learning and test conditions. Performance was above chance throughout the experiment, $Z = 4.0$, $p < .01$. We conducted a two-way ANOVA on the correct responses, with the learning condition (front or back surface) and the test condition (0° or 180° rotation) as factors. We found a main effect of learning condition, $F(1, 31) = 8.68$, $p < .01$: More correct responses were made when objects were learned from the back than when they were learned from the front. There was no effect of rotation at test, $F(1, 31) < 1$, and no interaction between the factors, $F(1, 31) < 1$.

Our results suggest that although objects can be recognized by the hand from either the front or the back surface, performance is better for objects learned from the back surface than those learned from the front. We also note that there was no effect of rotation at test. This is not surprising because the hand can use both the front and the back surface for haptic object recognition.

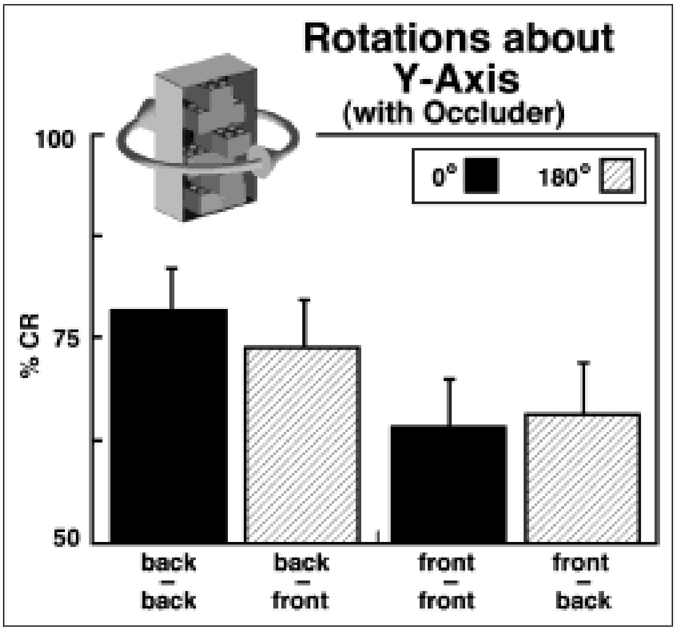

Fig. 5. Recognition performance (percentage of correct responses, or %CR) in Experiment 3. Objects were learned haptically from either the surface facing away from the participant (i.e., the back surface) or the surface facing the participant (the front surface). During test, the objects were explored from either the front or the back, and either the objects were presented in the same orientation as during learning (0° rotation) or the learned surface was rotated to face the opposite direction (180° rotation).

Our findings could be due to some combination of the biomechanical constraints of the hands and the nature of the stimuli used in our experiments; that is, perhaps it was more comfortable to learn these particular objects from the back than the front surface. Indeed, when each participant had finished the experiment, we conducted an informal inquiry about the participant's haptic exploration of the objects, and most of our participants reported that they found the back of the objects easier to explore. Clearly, there might be situations in which the front surface is the preferred view for haptic exploration; this might be true, for example, of face recognition by blind persons, although to our knowledge this has not been tested empirically. Nevertheless, in these experiments, viewpoint dependency in haptic recognition was due to the back surface providing a better representation of the objects than the front surface.

GENERAL DISCUSSION

Two general conclusions can be drawn from our studies. First, much like visual object recognition, haptic object representation is viewpoint-specific. Our results suggest that a haptic view is the back surface of an object, in that recognition performance was better for objects learned from the back than for objects learned from the front. Our observers knew that an object's view could change between learning and testing, and were given sufficient time to explore the objects during both learning (1 min for haptic conditions) and testing (unlimited time). If haptic representations were not viewpoint-specific but omnidirectional, our observers could have adopted an omnidirectional exploration strategy, which would have allowed them to retain good performance regardless of view. Although it remains possible that the haptic representation of objects is omnidirectional whereas haptic exploration is not, we wonder what would be the purpose of having a representation without the necessary sensory data to fully utilize it.

Our second conclusion is that the transfer of object information between the visual and haptic systems is viewpoint-specific. We found no evidence for a more abstract representation mediating the transfer (cf. Easton et al., 1997; Reales & Ballesteros, 1999). Had a more abstract representation been used, recognition performance across modalities should have been less sensitive to the object's view than recognition performance within the modalities. This was not the case. Instead, performance was good if, and only if, the same side of an object was sensed during learning and testing, a characteristic that was equally true both within and across modalities. Furthermore, there was little or no cost of transfer when there was no change of the sensed surfaces between learning and testing. For cross-modal recognition, performance was better when the front view from the visual representation matched the back view from the haptic representation than when the views remained fixed across modalities (see also Shimojo, Sasaki, Parsons, & Torii, 1989). Consequently, performance was not significantly different between the within-modality condition without a change in view and the across-modalities condition with a 180° rotation in either the horizontal or the vertical axis. This suggests that no additional representation is needed to mediate the transfer.

For the visual system, the optimal view of an object for recognition is the side of the object facing the observer. For the haptic system, we have shown that the optimal view is the back of the objects used in our experiments. When an object is fixed in space, the exploration of the back of the object may be a natural strategy adopted by the hands. Given the biomechanical constraints of the hands, the back of certain objects may be more accessible to the haptic system than the front is. However, this is by no means the only strategy for haptic object recognition, because the biomechanics of the hand do not necessarily restrict the exploration of a hand-sized object to a single surface. To us, it seems more than just a coincidence that the hands complement the eyes by "seeing" the back side of an object.

Acknowledgments—Special thanks go to LEGO Germany for kindly donating the stimuli. We thank M.S. Banks, G.E. Legge, J.M. Findlay, and D. Kersten for comments on earlier drafts of this article. This research was supported by the Human Frontiers Science Program, the Durham University Research Foundation, and the Max-Planck Society. Informed consent was obtained from all persons who participated in our experiments.

REFERENCES

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

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

Easton, R.D., Green, A.J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin & Review*, *4*, 403–410.

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

Gibson, J.J. (1962). Observations on active touch. *Psychological Review*, *69*, 477–490.

Goodale, M.A., Meenan, J.P., Bülthoff, H.H., Nicolle, D.A., Murphy, K.J., & Racicot, C.I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*, *4*, 604–610.

Goodale, M.A., Milner, D.A., Jakobson, L.S., & Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*, 154–156.

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

Kennedy, J.M. (1993). *Drawing and the blind: Pictures to touch*. New Haven, CT: Yale University Press.

Klatzky, R., Lederman, S.J., & Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General*, *116*, 356–369.

Lederman, S.J., & Klatzky, R.L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, *19*, 342–368.

Lederman, S.J., Klatzky, R.L., Chataway, C., & Summers, C.D. (1990). Visual mediation and the haptic recognition of two dimensional pictures of common objects. *Perception & Psychophysics*, *47*, 54–64.

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

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

Millar, S. (1997). *Reading by touch*. London: Routledge.

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

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

Reales, J.M., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: Cross-modal priming depends on structural descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 644–663.

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

Shimojo, S., Sasaki, M., Parsons, L.M., & Torii, S. (1989). Mirror reversal by blind subjects in cutaneous perception and motor production of letters and numbers. *Perception & Psychophysics*, *45*, 145–152.

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

(RECEIVED 10/14/99; REVISION ACCEPTED 5/4/00)