Max–Planck–Institut
für biologische Kybernetik
Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 080

# Viewpoint dependence in visual and haptic object recognition

Fiona N. Newell[1], Marc O. Ernst[2], Bosco S. Tjan[3]
& Heinrich H. Bülthoff[4]

March 2000

Fiona Newell is now at:
Department of Psychology
Áras an Phiarsaigh
University of Dublin
Trinity College, Dublin 2, Ireland
Bosco Tjan is now at:
NEC Research Institute Princeton, USA

[1] AG Bülthoff, E–mail: fiona.newell@tcd.ie
[2] AG Bülthoff, E–mail: marc.ernst@tuebingen.mpg.de
[3] AG Bülthoff, E–mail: tjan@research.nj.nec.com
[4] AG Bülthoff, E–mail: heinrich.buelthoff@tuebingen.mpg.de

---

# Viewpoint dependence in visual and haptic object recognition

*Fiona N. Newell, Marc O. Ernst, Bosco S. Tjan & Heinrich H. Bülthoff*

**Abstract.** On the whole, we recognise objects best when we see them from a familiar view and worse from views that were previously occluded from sight. Unexpectedly, we found haptic object recognition to also be viewpoint-specific, even though hand movements were unrestricted. This 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 recognising it, we found that recognition performance was best when the objects were also rotated back-to-front between learning and recognition. Our data indicate that the visual system recognises the front view of the objects best whereas the hand recognises the objects best from the back.

## 1 Introduction

We explore and navigate through our environment mainly using sight and touch. In order to guide actions and interact with objects, information acquired from the visual and the haptic systems must converge to form a coherent percept. What is the nature of the representation underlying each sensory system that facilitates 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 this paper, 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 and Edelman, 1992). The idea that visual recognition performance is viewpoint-specific has been well established (Jolicoeur, 1985; Palmer, Rosch and Chase, 1981). For unfamiliar objects, recognition performance is best when the objects are shown in views in which they were learned (Rock and DiVita, 1987; Edelman and Bülthoff, 1992). Even a familiar object (e.g., a dog) is recognised more efficiently when it is seen in the most typical (e.g., upright) position (Jolicoeur, 1985, Newell and Findlay, 1997). When objects are recognised independently of view it is generally because all views of the objects are familiar (Tarr and Bülthoff, 1995), or that the object contains very distinct parts (Biederman, 1987). View-dependent visual recognition performance has been found in humans and other primates (Logothetis, Pauls, Bülthoff, and Poggio, 1994). Neurophysiological studies have also revealed cells in inferior temporal cortex that are maximally tuned to specific views of objects (Logothetis, Pauls and 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 and Bülthoff, 1995).

To date, however, the nature of object recognition in the haptic system has received relatively less attention (Lederman and Klatzky, 1987; Lederman, Klatzky, Chataway and Summers, 1990). Some researchers have suggested that there are large representational similarities between the visual and haptic systems (Easton, Green and Srinivas, 1997). Indeed many of these studies have reported finding 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 and 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 only see an object from one view such that certain features of an object can be occluded from sight. This can lead to a general dependency on viewpoint. In contrast, when we handle 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 omni-directional and not viewpoint-specific. For the purpose of recognition it has been noted that the hands typically follow the contour of a 3-dimensional object until the object is recognised (Lederman and Klatzky, 1987). According to Lederman and Klatzky, this contour following exploration strategy was deemed necessary for haptic object recognition. What is not clear, however, is whether contour following in haptic object recognition necessarily leads to an omni- directional representation of objects in the haptic system.

Arguments for a viewpoint-specific haptic representation of objects often come 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 to visually identify the objects (Goodale, Milner, Jakobson and Carey, 1991; Goodale, Meenan, Bülthoff, Nicolle, Murphy and Racicot, 1994). Even for tasks in which visual information is not available, such as Braille reading or playing musical instruments, orientation and position information can, in theory, be attached to an omni-directional, viewpoint-independent haptic representation (Millar, 1997; Kennedy, 1993). Consequently, a viewpoint-specific haptic representation is not needed. If the haptic representation of objects is indeed omni-directional 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 this is not what we found. We conducted two experiments in which the effects of viewpoint were measured for visual, haptic and cross-modal object recognition.

## 2 Experiment 1

### 2.1 Method

**Participants**

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

**Apparatus**

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

---

[1] In previous studies on cross-modal recognition of 3-D objects the nature of the input to each of the senses was often not controlled. Generally, visual presentation consisted of viewing the object from a single viewpoint, whereas in the haptic condition the objects, not fixed in space, were freely manipulated by the hands (Lederman et al., 1990; Easton et al, 1997). Although these and other studies proposed that cross-modal recognition is mediated by abstract, structural descriptions we would argue that such studies can not directly address the issue of the nature of object representation in either sensory system because the sensory input was not equivalent.

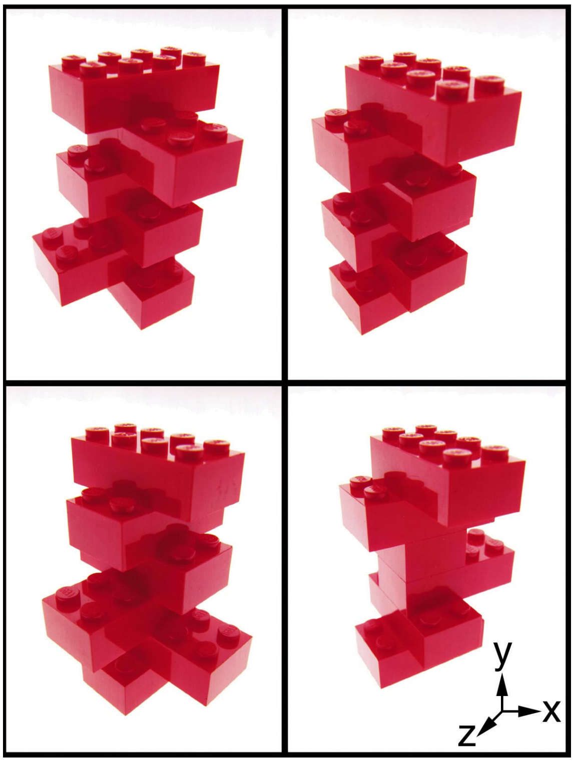

Figure 1: Examples of the objects used in our recognition experiments, and the co-ordinate system.

## Design

Thirty-two uniquely configured objects were used as stimuli in each experiment. The experiments were similarly designed and were based on a 3-way, repeated measures design with learning modality (vision or haptics), transfer at test (within or across modalities), and view change at test (0° or 180°), as factors. In each experiment there were four experimental blocks, referring to 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 counter-balanced 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 participants, each for 30 seconds in the visual-learning condition, or 1 minute for the haptic-learning condition. These presentation durations were chosen to yield equivalent recognition accuracy within each modality as determined by a pilot experiment. During test, participants were presented with four target objects and 4 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. Objects were presented to a subject by fixing it on a 15-cm-tall stand on the table. Presentation was done behind a curtain in the haptic condition.

## Procedure

There were four separate blocks in each 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 object during haptic exploration and their head during visual exploration, thus all surfaces of the objects could be perceived by each modality. However, they were instructed not to move the object or to walk around it. 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 given object was either from the learning set or was a distractor object. Recognition was tested either in the same modality as learning or in the other modality (16). All possible combinations were tested in four blocks: Visual learning - visual testing (visual-visual); haptic learning - haptic testing (haptic-haptic); visual learning - haptic testing (visual-haptic); and haptic learning - 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 Experiment 1 and 3, we tested rotations about the Y- (vertical) axis, whereas in Experiment 2 we used rotations about the X- and Z-axes (see Figure 1).

## 2.2 Results and Discussion

Figure 2 shows the percentage of correct responses (%CR) made to the learned objects for all four learning and testing conditions. Performance was above chance throughout the experiment, [$Z = 3.6, p < 0.01$]. We ran a 3-way 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 [$F(1, 25) = 1.205$, n.s.], Learning [$F < 1$] or Viewpoint [$F < 1$]. However, we found a significant interaction between Transfer and Viewpoint [$F(1, 25) = 6.755, p < 0.05$]. Simple effects analyses revealed an almost significant effects of viewpoint within modalities [$F(1, 25) = 3.261$, $p = 0.083$] and a significant effect of viewpoint across modalities [$F(1, 25) = 4.416, p < 0.05$].

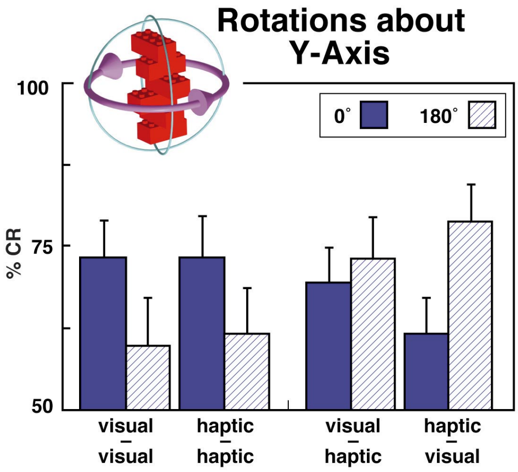

Figure 2: Recognition performance (% correct responses) of 26 individuals on the target objects that have their view either changed between learning and test (180°) or not changed (0°). Learning and testing were conducted either within the same visual (visual-visual) or haptic (haptic-haptic) modalities, or across the modalities (visual learning and haptic testing (visual-haptic), or haptic learning and visual testing (haptic-visual)). Error bars denote the standard errors of the mean across participants.

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 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 ($p < 0.05$) when the test object was rotated about the vertical axis (180°) than when it remained in the learning position (0°). Lastly, 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 is analogous to the visual system having a better representation of the front, and therefore visible, surface of an object (Figure 3). In sum, the backside of a hand-size object is often more accessible to the haptic system whereas the front is more accessible 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 sensed 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 predication that recognition performance in the cross-modal conditions would be best

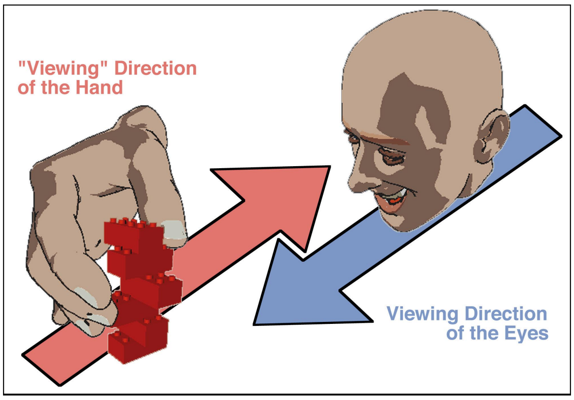

Figure 3: Schematic diagram of our model, which suggests that information integrated across the fingers is analogous to seeing an object from behind.

whenever there is an exchange of the front and back surfaces of an object between learning and testing sessions. We studied rotation about two additional axes: one axis of rotation involves an exchange of the front and back surfaces and the other axis of rotation does not.

# 3 Experiment 2

## 3.1 Method

The procedure and the materials were identical to those of Experiment 1 with one important exception. In this experiment, we tested two different orientation changes of the objects than the rotation tested in Experiment 1. Here, 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 session A. The front and back remain in the same position with rotations around the Z-axis. The recognition of targets rotated around the Z-axis was tested in session B. 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 randomised across participants. Participants received a self-timed break between sessions.

## 3.2 Results and Discussion

Figure 4 shows the recognition performance of the objects rotated around these two axes, which confirms our prediction. Performance was above chance throughout the experiment, [in Experiment 2(a), $Z = 5.32, p < 0.01$ and in Experiment 2(b), $Z = 5.22, p < 0.01$]. We conducted separate 3-way ANOVAs on the correct responses made in sessions A and B of the experiment. In Experiment 2 (A) 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 factors [$F(1, 35) = 12.530, p < 0.01$]. Simple effects analyses revealed a significant effect of viewpoint within modalities [$F(1, 35) = 9.833$, $p < 0.01$] and a significant effect of viewpoint across modalities [$F(1, 35) = 5.736, p < 0.05$].

In Experiment 2(B) we found no main effects for Transfer [$F < 1$], or for Learning modality [$F < 1$] but we found a significant main effect of Viewpoint [$F(1, 35) = 8.448, p < 0.01$]. There were no interactions between the factors.

In each case, we found that 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. where a front-back exchange of surfaces was present) but not around the depth (Z) axis (i.e. where a front-back exchange of surfaces was absent). Thus our results of this experiment support our prediction that different viewpoints mediate better integration of information across the visual and haptic modalities.

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 sur-

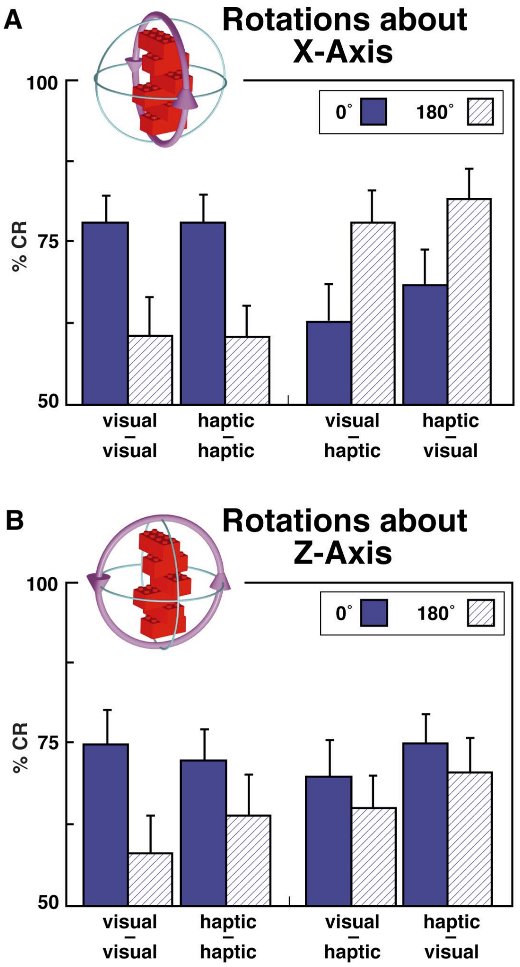

Figure 4: Recognition performance (% correct responses) of 36 individuals within and across modalities for objects rotated across different axes. In (A), the objects were rotated about the horizontal (X) axis, which involved an exchange of the front and back of the objects. In (B), the objects were rotated about the depth (Z) axis, which did not involve an exchange of the front and back of the objects.

face 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 to each of these surfaces. We would expect equally good recognition performance of the front and back of the objects if any surface of the object can be used for recognition. Alternatively, if the back surface is the preferred haptic 'view' of an object then recognition performance should be relatively better to the back than the front of the object.

## 4 Experiment 3

### 4.1 Method

The procedure and the materials were identical to those of the within-modality, haptic condition of Experiment 1, however, in this experiment, we used a cover on the objects which allowed us to restrict finger exploration to one surface of the object. Thus, in separate blocks in the experiment, 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 counter-balanced across participants. Target objects and distractor objects were counter-balanced 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 direction 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, both the object and the cover were treated as an entire object during rotation. Thirty-two undergraduate students from the Eberhard-Karls University of Tübingen, Germany participated in both sessions of the experiment for pay.

## 5 Results and Discussion

Figure 5 shows the recognition performance across the different learning and test conditions. Performance was above chance throughout the experiment [$Z = 4.0, p < 0.01$]. We conducted a two-way ANOVA on the correct responses, with the learning (front or back surface) and the test conditions (0° or 180° rotation) as factors. The learning or the test was either from the back surface of the object or the front surface. We found a main effect of learning, $[F(1,31) = 8.68, p < 0.01]$: There were more correct responses made when objects were learned from the back than the front. There was no effect of Rotation at test $[F(1,31) < 1]$, and no interaction between the factors $[F(1,31) < 1]$.

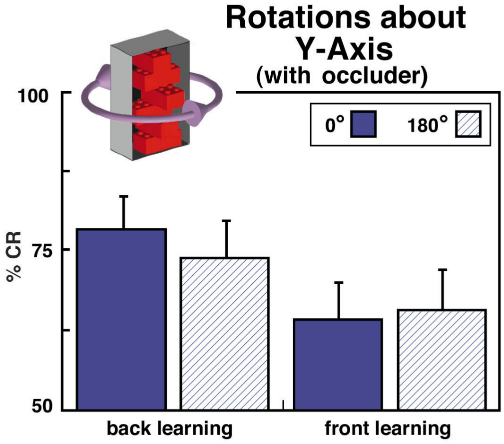

Figure 5: Recognition performance (% correct responses) of 32 individuals to objects which were learned haptically from either the surface facing away from the participant (i.e. the back surface) or from the surface facing the participant (the front surface). During test, the objects were either presented in the same orientation as learning, (0° rotation), or the learned surface was rotated to face the opposite direction (180° rotation).

Our results suggest that although objects can be recognised by the hand from either the front or the back surface, performance is better for objects that were learned from the back surface than those which were learned from the front. We also noted that there was no effect of rotation at the test. This is not surprising since the hand can use both the front and the back surface for haptic object recognition. 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, in that it was more comfortable to learn these particular objects from the back surface. Indeed, when each participant had finished the experiment we conducted an informal enquiry about their 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, such as face recognition by blind persons, although to our knowledge this has not been tested empirically. Nevertheless, here in our experiments, viewpoint dependency in haptic recognition was due to the back surface providing a better representation of the object than the front surface.

## 6 General Discussion

Two general conclusions can be drawn from our studies. First, similar to the visual system, haptic object representation is viewpoint-specific. From our results, a haptic 'view' is the back surface of the object, in that recognition performance was better to objects which were learned from the back. 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 minute for haptic conditions) and testing (unlimited). If haptic representations were not viewpoint-specific but omni-directional, our observers could have adopted an omni-directional exploration strategy, which would allow them to retain good performance regardless of view. Although it remains possible that the haptic representation of objects is omni-directional while haptic exploration is not, this possibility leaves one to wonder about the purpose of having a representation without the necessary sensory data to fully utilise it.

Our second conclusion is that the transfer of object information between the visual and haptic system is viewpoint-specific. We found no evidence for a more abstract representation mediating the transfer (c.f. 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 found to be best when the front view from the visual representation matched the back 'view' from the haptic representation (see also Shimojo et al. 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 axes. 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 an object facing the observer. For the haptic system, we have shown it to be the back of the objects used in our experiments. When an object is fixed in space, the exploration of the back of an 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. However, this is by no means the only strategy for haptic object recognition since 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 backside 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. 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 human participants who participated in our experiments.

Correspondence concerning this article should be sent to Fiona Newell, Department of Psychology, Áras an Phiarsaigh, University of Dublin, Trinity College, Dublin 2, Ireland. E-mail: fiona.newell@tcd.ie.

## References

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

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

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

Edelman, S. and 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., and 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., and 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 and Cognition, **13**, 289-303.

Klatzky, R., Lederman, S.J., and 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(4)**, 356-369.

Kennedy, J.M. (1993). Drawing and the Blind: Pictures to touch. Yale University Press.

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

Logothetis, N.J., Pauls, J., Bülthoff, H.H., and Poggio, T. (1994). View-dependent object recognition in monkeys. Current Biology, **4(5)**, 401-414.

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

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

Newell, F.N. and Findlay, J.M. (1997). Effects of depth rotation on object identification. Perception, **26(10)**, 1231-1257.

Palmer, S., Rosch,E., and Chase, P. (1985). Canonical perspective and the perception of objects. In H. Long and A.D. Baddeley, (Eds.). Attention and Performance IX, LEA, New Jersey.

Reales, J.M. and 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(3)**, 644 - 663.

Rock, I. and DiVita, J. (1987). A case of view-centred object perception. Cognitive Psychology, **19(2)**, 280-293.

Shimojo, S., Sasaki, M, Parsons, L.M., and Torii, S. (1989). Mirror reversal by blind subjects in cutaneous perception and motor production of letters and numbers. Perception and Psychophysics, **45 (2)**, 145-152.

Tarr, M.J. and 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.