Perception & Psychophysics
2004, 66 (2), 342–351

The visual and haptic perception of natural object shape

J. FARLEY NORMAN, HIDEKO F. NORMAN, and ANNA MARIE CLAYTON
Western Kentucky University, Bowling Green, Kentucky

JOANN LIANEKHAMMY
University of Kentucky, Lexington, Kentucky

and

GINA ZIELKE
University of Louisville, Louisville, Kentucky

In this study, we evaluated observers' ability to compare naturally shaped three-dimensional (3-D) objects, using their senses of vision and touch. In one experiment, the observers haptically manipulated 1 object and then indicated which of 12 visible objects possessed the same shape. In the second experiment, pairs of objects were presented, and the observers indicated whether their 3-D shape was the *same* or *different.* The 2 objects were presented either unimodally (vision–vision or haptic–haptic) or cross-modally (vision–haptic or haptic–vision). In both experiments, the observers were able to compare 3-D shape across modalities with reasonably high levels of accuracy. In Experiment 1, for example, the observers' matching performance rose to 72% correct (chance performance was 8.3%) after five experimental sessions. In Experiment 2, small (but significant) differences in performance were obtained between the unimodal vision–vision condition and the two cross-modal conditions. Taken together, the results suggest that vision and touch have functionally overlapping, but not necessarily equivalent, representations of 3-D shape.

Many researchers in the fields of both vision and haptics have examined the accuracy and precision of human observers' abilities to perceive local aspects of three-dimensional (3-D) object shape (e.g., Bülthoff & Mallot, 1988; de Vries, Kappers, & Koenderink, 1993; Gibson, 1950; Koenderink, Kappers, Todd, Norman, & Phillips, 1996; Norman & Raines, 2002; Norman & Todd, 1996, 1998; Norman, Todd, & Phillips, 1995; Pont, Kappers, & Koenderink, 1997, 1999; Tittle, Norman, Perotti, & Phillips, 1998). In contrast, there have been fewer studies in which the perception of overall (or global) 3-D shape has been examined, especially for objects that have complex, or *naturalistic*, global shapes, in which there are variations in surface orientation and curvature (Gibson, 1962, 1963; Lakatos & Marks, 1999; Norman, Dawson, & Raines, 2000; Norman, Todd, & Orban, in press). In an intriguing set of publications, Gibson (1962, 1963, 1966) informally described a cross-modal visual–haptic global-shape–matching task and concluded (Gibson, 1963) that "the ordinary observer, after very little practice, can distinguish among the tangible objects and match them to their visible replicas with little error" (p. 6). This conclusion, if valid, is important, since it directly suggests that the two different sensory modalities either share a high-level representation of shape or have representations of 3-D shape with similar enough formats for effective comparisons to take place (it is important at this point, however, to note that Gibson would not have approved of the usage of this *representational* terminology).

Gibson's (1962, 1963, 1966) experiments on cross-modal shape matching, conducted 40 years ago, are classics. Few people at that time investigated the perception of solid object shape, whether by vision or by touch. The issue of the *relationship* between vision and touch, however, had previously intrigued philosophers and scientists for many hundreds of years (see, e.g., Berkeley's *A New Theory of Vision*, 1709/1963). It is true that Gibson and his students (especially James Caviness) did study cross-modal shape matching. However, it is also unfortunately true that the actual data and results from these experiments were never published—Gibson refers to these studies only anecdotally in other journal articles and books. For example, in none of the relevant publications (1962, 1963, 1966) did Gibson ever report any quantitative data concerning the observers' performance for this task (no graphs, statistics, tables, etc.). Even if Gibson's conclusions re-

---

We thank G. John Andersen, Geoffrey P. Bingham, Hal A. Sedgwick, and Pat Cabe for their helpful comments on earlier drafts of the manuscript. Their suggestions led directly to the manipulations performed in Experiment 2. Correspondence concerning this article should be addressed to J. F. Norman, Department of Psychology, Western Kentucky University, Bowling Green, KY 42101 (e-mail: farley.norman@wku.edu).

Note—This article was accepted by the previous editorial team, headed by Neil Macmillan.

garding this task are valid, definitive published evidence is currently lacking. By delving into unpublished sources, it is possible, however, to find the data that Gibson used to support his belief in the "equivalence of the two modes of perception" (i.e., vision and touch; 1962, p. 489). First of all, Gibson (1963), when informally discussing the results obtained with the cross-modal shape-matching task, refers to a paper presentation at the Eastern Psychological Association meeting in 1962 (Caviness & Gibson, 1962). In his master's thesis, Caviness (1962, pp. 2–3) reviewed the results of this "preliminary experiment" and noted that their observers could match the shape of the haptically and visually presented objects with an accuracy of 89%. It is interesting to note that the 10 observers in this 1962 study performed only a total of 10 judgments each (one cross-modal shape-matching trial for each of the 10 stimulus objects). In his doctoral dissertation, Caviness (1964, pp. 29–34) then proceeded to replicate the results of the earlier study (now with a total of 2 cross-modal shape-matching judgments for each of the 10 stimulus objects) and obtained a second performance of 86% correct responses.

From the results of the cross-modal shape-matching studies described by Caviness and Gibson (1962) and Caviness (1962, 1964), we know that, at least for some objects, observers can match with reasonable accuracy the shape of an object known only by touch to that of an object known only by vision. However, many questions still remain. The pioneering experiments of Caviness and Gibson and of Caviness used manmade "sculpted" objects (which Gibson called "feelies"). Would similar results be obtained for naturally shaped 3-D objects? In addition, we know nothing about the nature of the mistakes made by the observers in the studies by Caviness and Gibson and by Caviness. When their observers matched a haptically presented object to a visually presented object with a different shape, what factors led to the observers' confusion

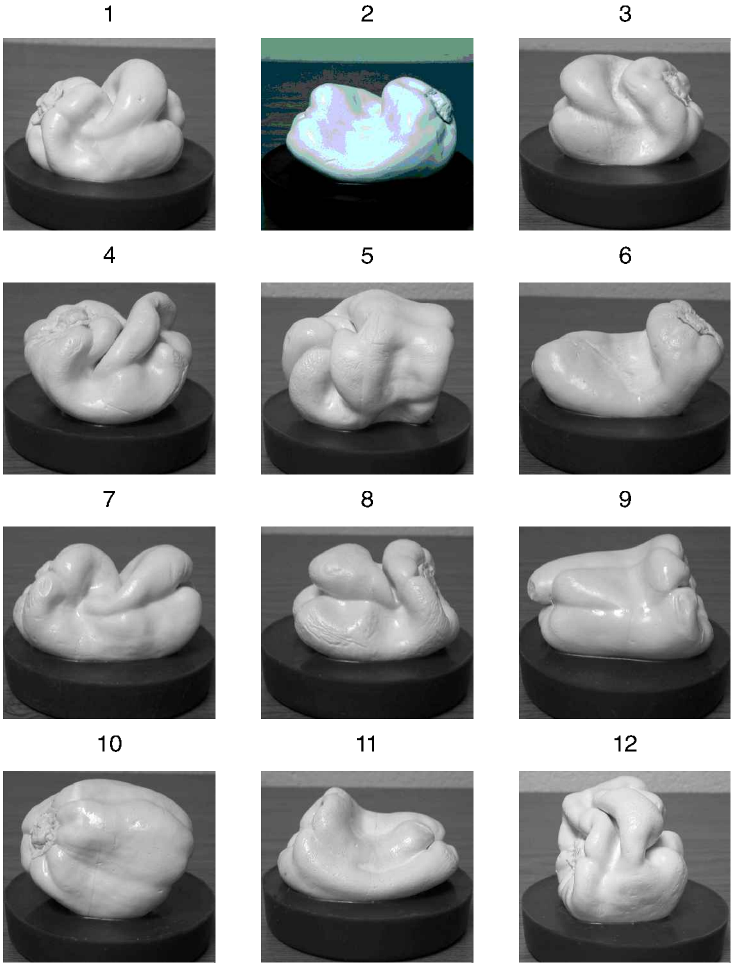

Figure 1. Photographs of the 12 objects (bell peppers, *Capsicum annuum*) used as experimental stimuli.

and error? What types of systematic mistakes do observers make in this type of cross-modal matching task? How do learning processes affect the accuracy and precision of cross-modal shape matching? We do not yet know, because in the experiments of Caviness and Gibson and of Caviness (1964), the observers performed only 10–20 matching judgments in a single experimental session (thus, there were only one or two trials for each of the 10 stimulus objects). One purpose of our experiments was to answer such questions by performing Gibson and Caviness's cross-modal shape-matching task for a set of naturally shaped 3-D objects. In the first experiment, we used the cross-modal shape-matching task and examined how the observers' performance changed over time. We also manipulated the amount of haptic exploration to determine how much time is required for observers to obtain an accurate haptic estimate of 3-D shape (in the experiments of Caviness and Gibson, their observers were always allowed an unlimited amount of time to make their judgments). In the second experiment, we used a *same/different* shape discrimination task and compared cross-modal discrimination performance with unimodal visual and unimodal haptic discrimination performance. By comparing the observers' performance across these conditions, we could evaluate whether there is any *loss*, or reduction, in efficiency when observers compare object shape across, as opposed to within, sensory modalities.

EXPERIMENT 1

Method

Stimulus displays. Two solid copies of 12 bell peppers (24 total) were made out of C-1506 rigid urethane casting compound (Smooth-on, Inc.). To create the "positive" replicas, liquid "plastic" C-1506 was poured into "negative" rubber molds (Evergreen 30, Smooth-on, Inc.) that were themselves created from a set of 12 natural objects (bell peppers, *Capsicum annuum*). The average volume of the stimulus objects was 350 cm^3 ($\sigma = 61.0$); the objects thus had an average diameter of 8.7 cm. Photographs of the 12 objects are shown in Figure 1.

Procedure. The basic procedure was for each observer to actively feel, or haptically *explore*, one of the 12 objects for a given amount of time (either 3, 5, 7, 9, or 15 sec), look at all 12 of the objects (spread out on a tabletop in front of the observer), and then select which visible object had the same shape as the one they had explored by touch. They were allowed to view the entire set of 12 objects during the haptic exploration (the same procedure as that described in Gibson, 1963). No feedback was ever given to the observers regarding their performance. The 12 visible objects were arranged in a semicircular pattern on the table and were all located at a distance of 60 cm from the observer. In haptically exploring an object, the observers reached under the table and behind a black curtain in order to touch it and were thus unable to see the object they were feeling. The observers viewed the visible objects binocularly, permitting stereopsis. They were allowed to move their heads (but not their bodies) from side to side, enabling the use of motion parallax. The plastic surfaces of the objects were shiny; therefore, specular highlights, as well as shading, were present. The viewing distance was a close 60 cm, allowing the use of oculomotor cues, such as accommodation and convergence. The viewing situation was, therefore, essentially full cue, so that there was a multiplicity of sources of information available to support the effective visual perception of the object's 3-D shape.

Each observer participated in five sessions, consisting of 24 trials each (120 trials total). The start of each new session began immediately upon the termination of the previous session. Within each session, each object (1–12) was presented twice, in random order. Any given observer was allowed 3, 5, 7, 9, or 15 sec to actively touch and explore an object prior to making a matching judgment (i.e., different groups of observers were given more or less time to perceive the object's shape by touch).

Observers. The observers were 50 students, faculty, and staff at Western Kentucky University. All of the observers were naive with regard to the purpose of the experiment. All the observers had normal or corrected-to-normal visual acuity.

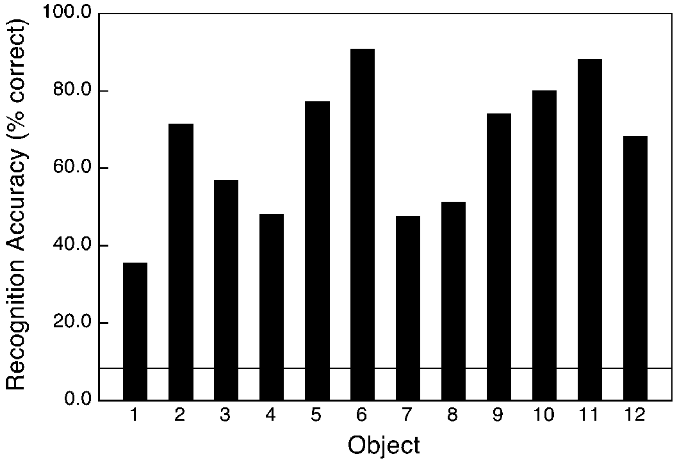

Figure 2. Overall recognition accuracies obtained in Experiment 1 for each of the 12 stimulus objects. The solid horizontal line represents chance performance (8.3% correct).

Results and Discussion

The overall results regarding recognition accuracy are shown in Figure 2. There were significant differences in recognizability [$\chi^2(11) = 259.5, p < .0001$]. In particular, Object 1 was the least recognizable, whereas Object 6 was the easiest to identify. All recognition accuracies were significantly greater than chance performance, which was 8.3% [$\chi^2(11) = 3641.4$, 4337.1, 6248.5, 5910.7, and 6119.5 for the 3-, 5-, 7-, 9-, and 15-sec groups, respectively; all ps $< .0001$].

The recognition results, plotted as a function of haptic exploration time and session number (effects of experience with the objects) are shown in Figure 3. There was a clear, although modest, effect of exploration time. Performance improved 31.6% from 3 to 7 sec and then asymptoted at longer haptic exploration times. There was also a small effect of experience over sessions; the improvement in performance occurred primarily between Sessions 1 and 2. These main effects were confirmed by a two-way split-plot analysis of variance [ANOVA; exploration time, $F(4,45) = 5.9, p < .001$; experience, $F(4,180) = 30.2, p < .0001$]. The interaction between exploration time and experience was not significant [$F(16,180) = 0.8, p > .05$]—that is, the effects of increases in experience were the same for all five groups of observers.

Since we recorded, on every trial, not only whether the observer was correct or incorrect in their matching judgment, but also his or her exact response, it is possible to go much further in analyzing the observers' data. On every trial, we actually acquired a pair of objects as data: the actual object that was presented haptically and the visible object that was chosen by the observer in making his or her shape-matching judgment. The grand confusion matrix (combining the results of all five groups of observers) is shown in Table 1. One can see immediately, for example, why the observers' recognition performance for Object 1 was so low (see Figure 2); it was highly confusable with Object 3. The observers frequently confused Object 1 with Object 3, but never with Object 6, 11, or 12. Similar confusions (and nonconfusions) are evident in Table 1 for all 12 stimulus objects. Figure 4 graphically plots the magnitudes of the confusions evident in Table 1 for all 66 pairs of objects (66 combinations of objects, 12 things taken 2 at a time; i.e., Pair 1 represents Objects 1 and 2, Pair 2 represents Objects 1 and 3, Pair 3 represents Objects 1 and 4, ... Pair 66 represents Objects 11 and 12). In particular, Figure 4 plots the observers' discrimination accuracies (high discrimination accuracy for a particular pair of objects indicates low *confusability*, whereas low discrimination accuracy indicates high *confusability*) for all 66 pairs of objects, using $-\ln \eta$,[1] which is a measure of discriminability derived from Luce's (1963) choice theory. Note that, in Figure 4, there is a striking homogeneity within the pattern of data points, indicating a large variability in the discrimination accuracies. Some pairs of objects are highly discriminable (not confusable, with high $-\ln \eta$ scores), whereas other pairs of objects are much less discriminable (more confusable, with low $-\ln \eta$ scores).

The data plotted in Figure 4 represent overall discrimination accuracies, where the data were collapsed across the five experimental sessions. The presence of $-\ln \eta$ values near 6.0 indicates that those pairs of objects were highly discriminable across all five sessions, whereas the pairs of objects with $-\ln \eta$ values near 1.0 were much less discriminable across all five experimental sessions. What about the pairs of objects for which there were intermediate discrimination accuracies? How did discrimination performance for those pairs change over the experimental sessions, as the observers gained additional experience with the objects? It is important to remember that in the experiments of Caviness and Gibson (1962) and Caviness (1964), each observer made only one or two matching

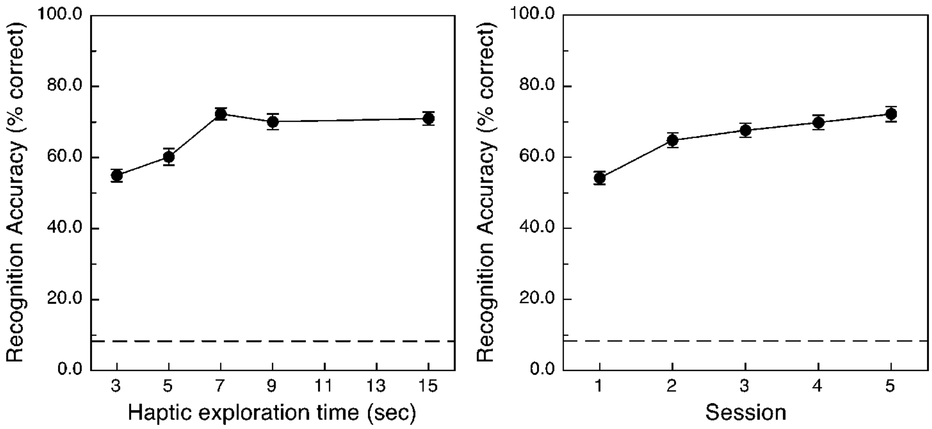

Figure 3. Plots of recognition accuracy in Experiment 1 as a function of haptic exploration time (left panel) and of increasing experience and practice with the visual–haptic shape-matching task (right panel). The error bars indicate ±1 standard error of the mean.

Table 1
Results of Experiment 1: Overall Confusion Matrix

	Stimulus Object											
Response	1	2	3	4	5	6	7	8	9	10	11	12
1	177	4	76	32	8	0	42	26	1	14	1	16
2	17	357	24	11	0	20	62	28	42	3	23	0
3	162	16	284	67	2	1	95	114	1	18	1	14
4	16	0	11	240	11	0	3	12	0	5	0	15
5	1	0	0	8	386	0	2	3	0	16	0	47
6	0	41	0	3	0	453	1	0	19	0	30	0
7	75	17	41	42	9	1	238	35	6	7	0	1
8	48	1	61	72	15	0	43	256	3	13	0	24
9	1	17	0	2	4	4	8	19	370	20	5	2
10	3	0	2	7	25	0	1	3	3	400	0	37
11	0	47	0	2	1	21	4	1	55	0	440	3
12	0	0	1	14	39	0	1	3	0	4	0	341

judgments for each of their 10 stimulus objects; thus, they would be unable to evaluate how their observers' matching performances changed over time as a result of experience. Figure 5 plots, for all 66 pairs of objects, the ratio of the observers' discrimination accuracies for Sessions 5 and 1, thus showing how much our observers' ability to discriminate or distinguish those objects changed over time from Session 1 to Session 5. Remember that no improvement would be indicated by a ratio of 1.0. From an inspection of Figure 5, it can be seen that for more than one half of the 66 pairs of objects (56%), there were discrimination accuracy ratios of 1.25 or less. Most pairs of objects thus had similar discriminabilities across all experimental sessions, and for those pairs, there was little effect of experience. This is why the overall effects of experience shown in Figure 3 were relatively modest.

Figure 6 shows representative pairs of objects for which significant improvements in discrimination performance across sessions were (left panel) and were not (right panel) exhibited. It is important to keep in mind that Figure 5 shows that for most pairs of objects, discriminability did not improve markedly over time. Let us first consider the pairs of objects depicted in the right panel. What properties do those pairs of objects have in common? Note that Objects 2 and 7 (data depicted in right panel, Figure 6, refer also to Figure 1) have global shape similarities: They both have a prominent *trough* in the middle (a similar global similarity occurs for Objects 1 and 3; the trough in Object 3 is facing toward the observer). Objects 4 and 7 (right panel) both have a prominent *protrusion* on their right sides. Note that the pairs of objects for which there was essentially no improvement in the observers' performance across sessions all had global shape similarities, although the objects' specific shapes (local magnitudes of surface depth, orientation, and/or curvature) were quantitatively quite different. In contrast, what object properties were responsible for the increases in performance that were demonstrated by the pairs of objects depicted in the left panel of Figure 6? Objects 3 and 8 both have prominent troughs (like Objects 1 and 3 or Objects 2 and 7) and were thus very confusable (low $-\ln \eta$ scores) in Session 1. However, note (see Figure 1) that the widths of the troughs were quite different for Objects 3 and 8. Over time, the observers' discrimination performance for Objects 3 and 8 improved as a result of discovering this difference in trough width. As an additional example, consider Objects 6 and 11. They are also similar globally: They both had large, relatively flat areas, as compared with the other objects. Because of this, they were initially confused with each other in Session 1, but this confusion was much reduced by Session 5, probably because, although they were both relatively flat, Object 11's *flat area* was noticeably more curved than that of Object 6. It seems clear from the results shown in Figure 6 that any objects that have strong global similarities will initially be confused to some extent (even if the objects have innumerable quantitative metrical differences in their physical shape). Whether that initial confusion is removed or ameliorated by experience depends on whether there are enough additional prominent

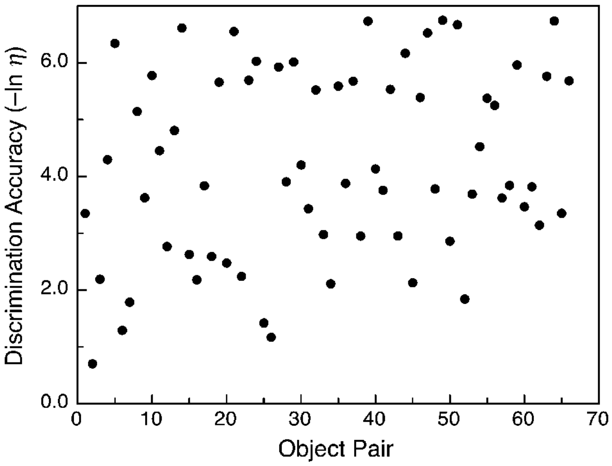

Figure 4. Overall discrimination accuracies for the 66 pairs of objects in Experiment 1. A $-\ln \eta$ value of zero represents chance performance, whereas values near 5.0 and 6.0 represent essentially perfect discrimination performance. The sequence of object pair numbers from 1 to 66 represents objects 1–2, 1–3, 1–4, . . . , 1–12, 2–3, 2–4, . . . , 2–12, 3–4, 3–5, . . . , 11–12.

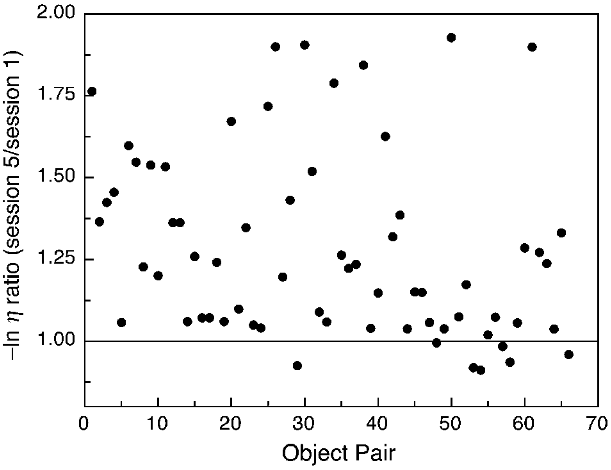

Figure 5. For each of the 66 pairs of objects in Experiment 1, the ratios of –ln η values for Sessions 1 and 5 (Session 5/Session 1) are plotted. Thus, no change in discrimination accuracy from Session 1 to Session 5 would be indicated by a ratio of 1.0. Ratios higher than 1.0 indicate improvement in discrimination accuracy as a result of increasing experience over sessions, whereas ratios less than 1.0 indicate a deterioration in performance over sessions. The object pair numbers are the same as those used in Figure 4.

local features, such as widths of trough gaps (or bumps) or large differences in surface curvature.

EXPERIMENT 2

The results of the first experiment showed that observers can indeed match objects known only by touch with other objects known only by vision at performance levels that are many times that of chance. This indicates that the two sensory modalities have representations of shape that are at least similar, if not identical. Performance is not perfect, however, and observers often confuse haptically presented objects with visual objects that are different but share important global shape similarities. In order to place further constraints on the relationship between vision and touch, it is necessary to compare the performance of cross-modal shape judgments (haptic–vision and vision–haptic) with that of similar unimodal judgments (haptic–haptic and vision–vision). If cross-modal shape judgments can be performed as accurately as the unimodal shape judgments, it is possible that the observers' judgments are being based on a high-level representation of shape toward which both vision and haptics contribute. An alternative possibility in this case is that the representations for vision and touch are separate entities but effective comparisons can be made between them. If, however, performance for cross-modal comparisons of 3-D shape is significantly worse than that obtained for both unimodal vision–vision and haptic–haptic comparisons, that would suggest a number of possibilities. One possibility is that the visual and the haptic representations may be distinct entities with similar formats (i.e., the 3-D object properties that are explicitly represented or detected by vision and active touch are similar or identical) but the process by which these representations can be compared is not completely effective or efficient. A distinctly different possibility is that the 3-D object properties encoded within the visual and the haptic representations are themselves different (i.e., each sensory modality is detecting and encoding different attributes of the ob-

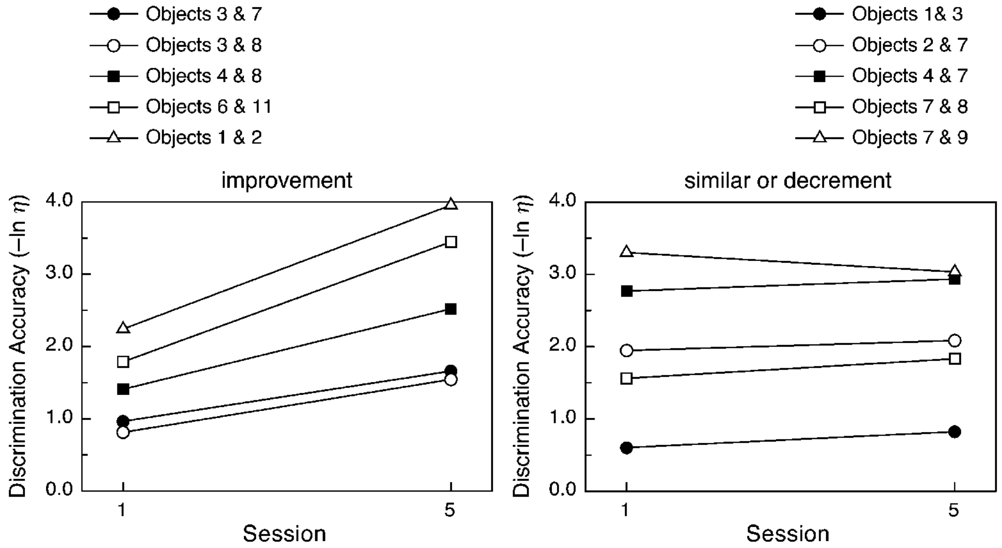

Figure 6. Example pairs of objects for whom discrimination accuracy in Experiment 1 either did (left panel) or did not (right panel) improve from Session 1 to Session 5.

jects' 3-D structures to a lesser or greater extent). The purpose of Experiment 2 was to investigate these issues.

Method

Stimulus displays. The stimulus objects were the same as those in Experiment 1 (see Figure 1).

Procedure. In this experiment, a *same/different* response task was used. The observers were presented with two objects sequentially and were required to indicate whether the two objects possessed the same shape or were different in shape. The objects were presented for a duration of 3 sec each, with a 3-sec interstimulus interval (these timing parameters were identical to those used in a similar experiment performed by Caviness, 1964). There were four experimental conditions: (1) vision–vision (both objects were presented visually), (2) haptic–haptic (both objects were presented haptically), (3) haptic– vision (the first object was presented haptically, and the second object was presented visually), and (4) vision–haptic (the first object was presented visually, and the second object was presented haptically). The objects in the haptic conditions were presented to the observers in the same manner as that used in Experiment 1 (i.e., the observers felt, or haptically explored, the object(s) behind an occluding black curtain). In the conditions with vision, the objects were placed onto a turntable that rotated the objects in depth around a vertical axis at a constant angular speed of 30 rpm (given that the objects were presented for 3 sec, this speed of rotation would allow the objects to make 1.5 complete revolutions). Given that in the haptic conditions, the observers could actively rotate and explore the object with their hands, it was felt desirable to have the visual object be viewed from changing perspectives as well, to make the patterns of proximal stimulation as similar as possible. In all conditions (haptics and/or vision), the object in each temporal interval was presented to the observers in a randomly chosen orientation. The rotation axis of the visual objects was located at a viewing distance of 50 cm, and the visual objects were viewed at the observers' eye heights. The viewing situation was full cue, as was the case in Experiment 1. No feedback was ever given to the observers regarding their performance.

Each observer participated in a single experimental session consisting of 120 trials (10 trials for each of the 12 stimulus objects). In the 10 trials for a given object, it was paired either with itself, creating *same* trials, or with randomly chosen second objects, creating *different* trials. Half of the 120 trials were *same* trials, and half were *different* trials. The order of the experimental stimuli (12 objects, *same* vs. *different* trials, etc.) was determined randomly for each individual observer.

Observers. The observers were 56 students at Western Kentucky University (14 observers for each experimental condition). None had participated in Experiment 1. The observers were naive with regards to the purpose of the experiment. All the observers had normal or corrected-to-normal visual acuity.

Results and Discussion

The observers' overall discrimination performances for the four main experimental conditions (vision–vision, haptic–haptic, vision–haptic, and haptic–vision) are shown in Figure 7. The results are plotted in terms of d' rather than $-\ln \eta$, because a choice theory measure of sensitivity for the *same/different* task has not yet been developed (for a yes/no discrimination task, d' and $-\ln \eta$ are almost perfectly correlated; see Figure 1.1 of Macmillan & Creelman, 1991). The observers' performances across the unimodal and cross-modal conditions were similar (in terms of percentage correct, the lowest discrimination accuracy was 80.1%, whereas the highest was 87.7%). Nevertheless, there were significant differences between the various conditions. A one-way between-subjects ANOVA performed on the d' scores revealed the presence of significant differences [$F(3,52) = 4.7, p < .01$]. The results of a Fisher LSD post hoc test revealed that there were significant differences ($p < .05$) between the unimodal vision–vision condition and the two cross-modal conditions (vision–haptic and haptic–vision). There were no other significant pairwise differences.

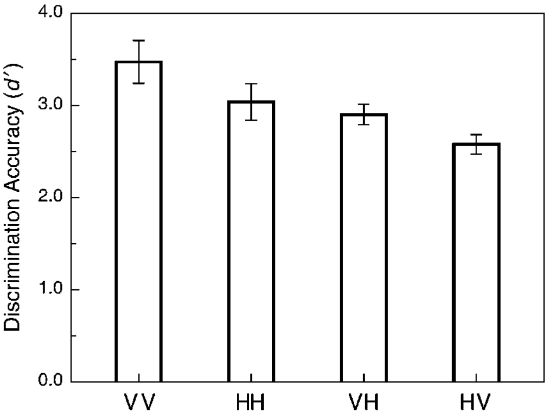

Figure 7. Results of Experiment 2. Discrimination performance is plotted as a function of the various uni- and cross-modal conditions. The error bars indicate ±1 standard error of the mean. V, visual; H, haptic.

During the collection of the data, it readily became evident that the observers' performances for the *same* and the *different* trials were not equivalent. Furthermore, how the observers performed on *same* and *different* trials seemed to vary systematically with the particular experimental condition. To highlight these effects, we performed a two-way split-plot ANOVA with one between-subjects factor (the various unimodal and cross-modal conditions) and one within-subjects factor (the type of trial: *same* vs. *different* trials). Both main effects were significant [for differences among the uni- and the cross-modal conditions, $F(3,52) = 5.0, p < .01$; for type of trial, $F(1,52) = 5.3$, $p < .03$]. In addition, the interaction between modality conditions and type of trial (*same* vs. *different*) was significant [$F(3,52) = 6.3, p = .001$]. The nature of this interaction can be seen in Figure 8. It is obvious that in the two unimodal conditions (vision–vision and haptic–haptic), the observers' performances were substantially higher on *different* trials. When the two objects were presented within a particular modality, there were very few mistakes when the objects were, in fact, different in 3-D shape. The errors in these conditions typically occurred in the *same* trials and were, thus, probably due to failures of perceptual constancy (since the objects in each temporal interval were placed in new random orientations). However, the observers' patterns of errors were quite different for the cross-modal conditions (vision–haptic and haptic–vision). In these cross-modal conditions, performances were either similar for *same* and *different* trials (haptic–vision) or

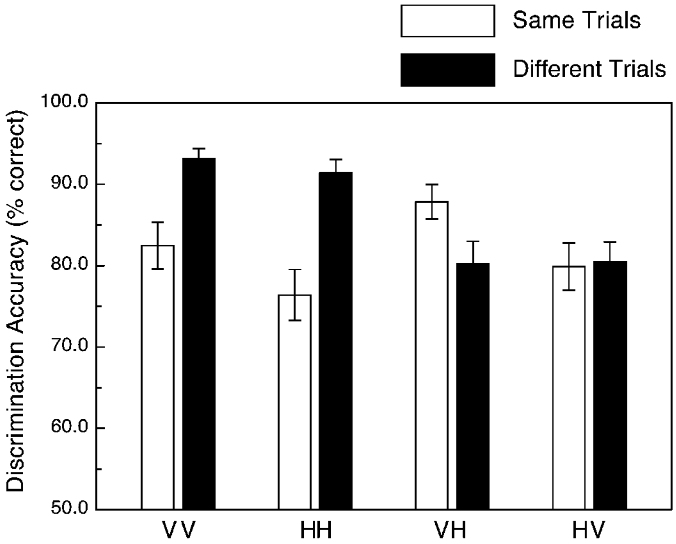

Figure 8. A graphical depiction illustrating the observers' differential pattern of performance on *same* and *different* trials as a function of the uni- and cross-modal conditions employed in Experiment 2. The error bars indicate ±1 standard error of the mean. V, visual; H, haptic.

worse for different trials (vision–haptic). In the vision–haptic condition, performances were best for the *same* trials, and the mistakes typically occurred when the objects were, in fact, different in terms of 3-D shape. In other words, the observers in this condition tended to confuse *different* objects as being the same, probably because of similarities in the two objects' global shape (see the results of Experiment 1; Figure 6, right panel).

GENERAL DISCUSSION

The present study adopted the visual–haptic shape-matching methodology developed in the classic experiments by Gibson (1962, 1963), but the essential interest in whether vision and touch exhibit important similarities for the perception and representation of object shape is far older. Past philosophers and scientists have come to quite different conclusions. Berkeley (1709/1963), for example, believed that the perception of an object's shape by touch and that by vision were essentially dissimilar. Any perceived similarity between the shape of two objects, one perceived by touch and one by vision, would occur only after long experience; the perceptions per se initiated by the senses were quite different. Berkeley said that "no man, at first receiving his sight, would know there was any agreement between this or that particular object of his sight, and any object of touch he had been already acquainted with" (p. 60). In contrast, Lucretius Carus (~58 B.C./1950) pointed out an essential identicalness in the perception of shape by touch and vision, "since shape examined by our hands within the dark is known to be the same as that by eyes perceived within the light and lustrous day, both touch and sight must be by one like cause aroused" (p. 143).

Gibson's (1962, 1963) conclusions from his experiments conducted in the 1960s were more like those of Lucretius than those of Berkeley. He concluded (in 1962) that vision and touch "seem to register the *same* information and to yield the *same* phenomenal experiences" (p. 490; italics are Gibson's). Later research has confirmed that haptic object recognition, like that of vision, can indeed be performed rapidly and accurately. For example Klatzky, Lederman, and Metzger (1985) showed that haptic identification of a 3-D object typically occurs within 1–2 sec, and that 94% of the correct recognitions took place within 5 sec or less. In addition, there are now a wide variety of studies in which many different methods have been used (tests of viewpoint dependence/independence, cross-modal priming, multidimensional scaling, etc.) that document important similarities between vision and touch for object recognition and/or the perception of shape (Easton, Greene, & Srinivas, 1997; Garbin, 1990; Garbin & Bernstein, 1984; Newell, Ernst, Tjan, & Bülthoff, 2001; Reales & Ballesteros, 1999; see also Marks, 1978).

Both touch and vision are similar in that they are sensitive to an object's global, or overall, shape (for vision, see Norman et al., 2000, and Norman et al., in press; for touch, see Lakatos & Marks, 1999); thus, when two physically different objects' global shapes are similar, visual–haptic shape matching becomes more difficult, and significant confusions develop that do not necessarily diminish with additional experience (see the results shown in Figure 6, right panel). When there are *prominent* local shape differences between two objects, however, our results show that shape-matching judgments do improve across experimental sessions, as observers gain increasing amounts of experience and familiarity (Figure 6, left panel). However, there are definite limits to observers' visual and haptic sensitivities to local 3-D metric structure. For example, Norman and Todd (1996, 1998) showed that even with a multiplicity of sources of visual information about 3-D structure, such as stereoscopic disparity, motion, shading, and texture, their observers' Weber fractions for judging the magnitude of depth differences between different local parts of an object's surface ranged from 12% to 44%. Their observers' abilities to judge the magnitude of local depth differences between two regions were actually worse when there was a smooth 3-D surface connecting the two regions; performance improved when the two regions to be judged were disparate spots hanging (or floating) in "empty space." However human observers visually perceive the 3-D shape of a solid object, it does not appear to necessarily involve the formation of an accurate and dense representation of local 3-D structural information, such as surface depth. In a like vein, Pont et al. (1999) found that tactually obtained curvature thresholds varied substantially for the same set of objects, depending on the amount (length) of the surface the observers were allowed to tactually explore; the physical shapes did not change, but the perceived shapes did, depending on how the objects were touched (see also the "cookie-cutter" experiment of Gibson, 1962). In any event, if both vision and touch were exquisitely sensitive to the exact magnitudes of local surface depths, orientations, and/or curvatures, none of the objects used in the present study should necessarily have been confused with any of the others, since each object did have a *distinct* 3-D shape that was geometrically different from all the rest (see Figure 1). Apparently, the human ability to perceive differences in local metric structure between objects is inexact, since our observers often had difficulty in discriminating one object from another even with significant amounts of practice (see Figures 4–6).

The results of the present experiments show that there are many similarities in how the visual and the haptic systems detect 3-D object shape. As was mentioned in the previous paragraph, both systems are apparently sensitive to differences in the global aspects of 3-D object shape, and they both exhibit inaccuracies in the detection of local 3-D surface properties, such as depth and curvature. Perhaps it is because of this similarity in how the visual and the haptic systems respond to global aspects of object shape that it is possible for human observers to compare with reasonably high levels of accuracy the shape of an object known only by active touch with one known only by vision (Experiment 1). The results of Experiment 2 showed that the observers' ability to compare object shapes across sensory modalities could be just as good as when both objects were compared within a single modality (i.e., there was no significant difference between the unimodal haptic–haptic judgments and the cross-modal haptic–vision and vision–haptic judgments). At the same time, however, our observers did demonstrate a slight superiority for purely visual judgments (see Figure 7); there were significant differences between the performances obtained for the unimodal vision–vision and the cross–modal conditions.

Given our findings of significant amounts of *crosstalk* between the senses of vision and haptics in how they respond to 3-D object shape, it is especially interesting to consider some of the recent findings in neurophysiology. Using fMRI brain scanning, James et al. (2002; see also Amedi, Malach, Hendler, Peled, & Zohary, 2001) found that both visual and haptic exploration of 3-D object shape produced significant amounts of activation in the human extrastriate visual cortex. They concluded from their research that human observers possess "a common haptic and visual representation of object shape" (p. 1713). James et al. suggested that one candidate for this common haptic and visual representation is the middle occipital (MO) area, which lies within the lateral occipitotemporal complex. Our psychophysical results do not necessarily disagree with the conclusions of James et al.; indeed, we found that observers can effectively perform cross-modal 3-D shape comparisons. Our observers' performance could have been due to the operation of a single representation that was contributed to by both vision and haptics. On the other hand, our observers did make mistakes and were not perfectly accurate in their cross-modal shape comparisons, as one might have expected if their visual and haptic perceptions of 3-D shape were being mediated by a single common representation. The results of future research should be able to determine whether the MO area (or some other cortical area) does indeed constitute a single perceptual representation that simultaneously underlies both the visual and the haptic perception of 3-D object shape. At the moment, we can at least conclude that there are important similarities between vision and haptics. The results of our present experiments and those of James et al. would seem to confirm Gibson's (1979) statement that "seeing and touching are two ways of getting much the same information about the world" (p. 258).

REFERENCES

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, **4**, 324-330.

Berkeley, G. (1963). *A new theory of vision and other writings.* New York: Dutton. (Original work published 1709)

Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A*, **5**, 1749-1758.

Caviness, J. A. (1962). *The equivalence of visual and tactual stimulation for solid shape perception.* Unpublished master's thesis, Cornell University.

Caviness, J. A. (1964). *Visual and tactual perception of solid shape.* Unpublished doctoral dissertation, Cornell University.

Caviness, J. A., & Gibson, J. J. (1962, April). *The equivalence of visual and tactual stimulation for the perception of solid forms.* Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, NJ.

de Vries, S. C., Kappers, A. M. L., & Koenderink, J. J. (1993). Shape from stereo: A systematic approach using quadratic surfaces. *Perception & Psychophysics*, **53**, 71-80.

Easton, R. D., Greene, 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.

Garbin, C. P. (1990). Visual–touch perceptual equivalence for shape information in children and adults. *Perception & Psychophysics*, **48**, 271-279.

Garbin, C. P., & Bernstein, I. H. (1984). Visual and haptic perception of three-dimensional solid forms. *Perception & Psychophysics*, **36**, 104-110.

Gibson, J. J. (1950). The perception of visual surfaces. *American Journal of Psychology*, **63**, 367-384.

Gibson, J. J. (1962). Observations on active touch. *Psychological Review*, **69**, 477-491.

Gibson, J. J. (1963). The useful dimensions of sensitivity. *American Psychologist*, **18**, 1-15.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

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

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, **40**, 1706-1714.

Klatzky, R. L., Lederman, S. J., & Metzger, V. A. (1985). Identifying objects by touch: An "expert system." *Perception & Psychophysics*, **37**, 299-302.

Koenderink, J. J., Kappers, A. M. L., Todd, J. T., Norman, J. F., & Phillips, F. (1996). Surface range and attitude probing in stereoscopically presented dynamic scenes. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 869-878.

Lakatos, S., & Marks, L. E. (1999). Haptic form perception: Relative salience of local and global features. *Perception & Psychophysics*, **61**, 895-908.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103-189). New York: Wiley.

Lucretius Carus, T. (1950). *Of the nature of things* (W. E. Leonard, Trans.). New York: Dutton. (Original work published ~58 B.C.)

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Marks, L. E. (1978). *The unity of the senses: Interrelations among the modalities*. New York: Academic Press.

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, **12**, 37-42.

Norman, J. F., Dawson, T. E., & Raines, S. R. (2000). The perception and recognition of natural object shape from deforming and static shadows. *Perception*, **29**, 135-148.

Norman, J. F., & Raines, S. R. (2002). The perception and discrimination of local 3-D surface structure from deforming and disparate boundary contours. *Perception & Psychophysics*, **64**, 1145-1159.

Norman, J. F., & Todd, J. T. (1996). The discriminability of local surface structure. *Perception*, **25**, 381-398.

Norman, J. F., & Todd, J. T. (1998). Stereoscopic discrimination of interval and ordinal depth relations on smooth surfaces and in empty space. *Perception*, **27**, 257-272.

Norman, J. F., Todd, J. T., & Orban, G. (in press). Perception of 3D shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science*.

Norman, J. F., Todd, J. T., & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception & Psychophysics*, **57**, 629-636.

Pont, S. C., Kappers, A. M. L., & Koenderink, J. J. (1997). Haptic curvature discrimination at several regions of the hand. *Perception & Psychophysics*, **59**, 1225-1240.

Pont, S. C., Kappers, A. M. L., & Koenderink, J. J. (1999). Similar mechanisms underlie curvature comparison by static and dynamic touch. *Perception & Psychophysics*, **61**, 874-894.

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, & Cognition*, **25**, 644-663.

Tittle, J. S., Norman, J. F., Perotti, V. J., & Phillips, F. (1998). The perception of scale-dependent and scale-independent surface structure from binocular disparity, texture, and shading. *Perception*, **27**, 147-166.

NOTE

1. The $-\ln\eta$ values were calculated according to the following equation:

$$-\ln\eta = \frac{1}{2}\ln\frac{\left(N_{a,a}+\frac{1}{2}\right)\left(N_{b,b}+\frac{1}{2}\right)}{\left(N_{a,b}+\frac{1}{2}\right)\left(N_{b,a}+\frac{1}{2}\right)},$$

and where $N_{a,a}$ is the number of responses of a, given that a was the actual stimulus, $N_{b,b}$ is the number of responses of b, given that b was the actual stimulus. Likewise, $N_{a,b}$ is the number of responses of a, given that b was the actual stimulus, and $N_{b,a}$ is the number of responses of b, given that a was the actual stimulus.

(Manuscript received September 11, 2002;
revision accepted for publication August 14, 2003.)

Nominations for the Editorship of *Memory & Cognition*

Nominations are solicited for the editorship of *Memory & Cognition*. The term of the present editor, Colin M. MacLeod, expires at the end of 2005. The new editor will begin an official five-year term on January 1, 2006, and will begin to receive manuscripts January 1, 2005. The Publications Committee of the Psychonomic Society expects to appoint the new editor by September 2004.

Nominations (including self-nominations) should be submitted by May 1, 2004, to

Kathryn Bock
Chair, *Memory & Cognition* Search Committee
Beckman Institute, University of Illinois
405 North Mathews
Urbana, IL 61801
e-mail: kbock@psych.uiuc.edu