1 Introduction

The congenitally totally blind are able to identify raised-line drawings of common objects and make recognisable raised-line drawings explored haptically, and corrected themselves without feedback. The self-correction suggests that participants can assess the extent to which the referents they suggest as possible identifications fit the haptic pictures. Indeed, when we asked subjects to identify haptic pictures, and to judge how well the referents they mentioned fitted the pictures, their fit judgments predicted the accuracy of their suggestions. Also, when one group of subjects offered the suggestions and another group assessed the fit of the suggestions to the pictures, the fit judgments predicted the accuracy of the suggestions. Further, good fit predicted successful recognition memory. In addition, both high and low fit judgments were made confidently, so the range of confidence judgments was smaller than the range of fit judgments. Finally, visual judgments of fit by one group predicted the level of success of the suggestions from another (haptic) group. In sum, subjects assess their suggested identifications appropriately, most likely on the basis of object shape criteria, outlined surface edges, and use of a vantage point.

Abstract. D'Angiulli et al (1998 Scandinavian Journal of Psychology 39 187–190) found blind and sighted (blindfolded) children identified common objects in raised-outline drawings explored haptically, and corrected themselves without feedback. The self-correction suggests that participants can assess the extent to which the referents they suggest as possible identifications fit the haptic pictures. Indeed, when we asked subjects to identify haptic pictures, and to judge how well the referents they mentioned fitted the pictures, their fit judgments predicted the accuracy of their suggestions. Also, when one group of subjects offered the suggestions and another group assessed the fit of the suggestions to the pictures, the fit judgments predicted the accuracy of the suggestions. Further, good fit predicted successful recognition memory. In addition, both high and low fit judgments were made confidently, so the range of confidence judgments was smaller than the range of fit judgments. Finally, visual judgments of fit by one group predicted the level of success of the suggestions from another (haptic) group. In sum, subjects assess their suggested identifications appropriately, most likely on the basis of object shape criteria, outlined surface edges, and use of a vantage point.

D'Angiulli et al (1998) found blind and sighted (blindfolded) children identified common objects in raised-outline drawings explored haptically, and corrected themselves without feedback. The self-correction suggests that participants can assess the extent to which the referents they suggest as possible identifications fit the haptic pictures. Indeed, when we asked subjects to identify haptic pictures, and to judge how well the referents they mentioned fitted the pictures, their fit judgments predicted the accuracy of their suggestions. Also, when one group of subjects offered the suggestions and another group assessed the fit of the suggestions to the pictures, the fit judgments predicted the accuracy of the suggestions. Further, good fit predicted successful recognition memory. In addition, both high and low fit judgments were made confidently, so the range of confidence judgments was smaller than the range of fit judgments. Finally, visual judgments of fit by one group predicted the level of success of the suggestions from another (haptic) group. In sum, subjects assess their suggested identifications appropriately, most likely on the basis of object shape criteria, outlined surface edges, and use of a vantage point.

1 Introduction

The congenitally totally blind are able to identify raised-line drawings of common objects and make recognisable raised-line drawings (Kennedy 1993; Kennedy and Merkas 2000). D'Angiulli et al (1998) report that both blind and sighted children identified raised-line drawings. The pictures most often identified by one group were the pictures most frequently identified by the other. Given a second opportunity to identify the pictures, both the blind and the sighted often changed their suggestions when wrong on the initial trial, but retained the suggested labels when right.

Here, some implications of the D'Angiulli et al (1998) findings are put to the test. In a nutshell, the implications are as follows. To change when wrong and stay when right, subjects must judge the extent to which a suggestion fits the display. By ‘fit’ we mean the degree of correspondence between perceived structure and deduced identity (e.g. “the lower border of this roughly rectangular display contains two semicircular parts, and so it could depict a car”). Fit judgments of this kind should predict the likelihood of being correct, since they are based on objective shapes. If the fit judgments are based on objective shapes, person A’s fit judgments of B’s labels should correlate with B’s fit judgments, and should also predict the likelihood that the label is correct. If a subject sticks with a label, it is likely that he or she is confident it deserves a high fit judgment, but also if the person changes the suggested label it is likely that he or she is confident that it deserves a low fit judgment. Also, if the basis of the judgment is objective shape criteria, fit judgments should be similar in haptic and visual inspection.

These implications were tested in experiments on sighted subjects unfamiliar with haptic pictures. By way of introduction, first we will describe haptic pictures— or ‘tactile’ pictures (Eriksson 1998)—and positive and negative views on their identifiability.

The haptic pictures we tested were raised-line drawings of common objects such as a car, a person, a cup, and a telephone. These pictures are made with a raised-line kit (Kennedy 1997), which consists of a board with a face coated with rubber on which
a textured plastic sheet is placed. Raised lines are produced when an ordinary ballpoint pen is pressed on the sheet and drawn over it. The surface puckers and forms a raised line that can be felt from the side from which the pressure is applied. Blind and blindfolded participants can identify pictures in raised-line drawings, albeit after a longer exploration time than would be needed in visual inspection, and at a lower accuracy than vision would achieve (Kennedy 1974; Lederman et al 1990; Millar 1994).

There are radically different views about the impressions given by a haptic picture. As we entertain a possible referent (a car? a truck? a bus?) do we get a useful sense of how well or badly the picture depicts its topic? A negative view is that raised pictures are simply not suited to haptics. A more optimistic view is that they are assessed effectively. The literature on haptic pictures indicates that both views deserve consideration (Arnheim 1990; Eriksson 1998; Heller 2000; Holmes et al 1998; Hopkins 2000; Jansson and Holmes, in press; Lederman et al 1990; Lopes 1997; Millar 1994, 2000).

Consider a very negative view. The perceiver may find information about figure-ground relations, orientation, occlusion, the observer's vantage point, and three dimensionality in haptic line pictures very difficult to appreciate (Eriksson 1998; Hopkins 2000). The lines may seem highly ambiguous since outline can have several figure-ground effects (Kennedy 1974), standing for occluding edges of curved and flat surfaces, as well as convex and concave corners, wires, and cracks. Perceivers may have trouble integrating information across a 2-D haptic line display if each line stands for any of several kinds of borders of a surface. If haptics provides poorly integrated form impressions from outline pictures (Arnheim 1990; Attnave and Benson 1969), it must produce very different form impressions from outlines than in vision, since outline drawings work exceedingly well in vision. We might judge any haptic picture as unsatisfactory, by the high standards our vision has given us. The outcome could be that high and low identification rates are unrelated to judgments of the fit of the picture to its referent. Is this a picture of a dagger or a tree, or a hand, or a garden claw, or an octopus that we feel before us? It might seem to fit any of these, but all equally poorly.

A haptic picture might be compared to an object glimpsed a bit at a time, perhaps through highly irregular or frosted glass (Heller et al 1999), in the sense that only a few features at most, such as a few intersections of lines, or some rough proportions of some selected parts, might be used in support of an attempted identification. Someone trying to identify the picture may use just a few of the picture's features to offer a guess about a possible referent, but have little reason to choose that referent over many others.

One reason for thinking haptic pictures give this kind of relatively imprecise information about their referents is that studies on haptic pictures have sometimes yielded recognition rates of only about 10% (Heller 1989; Kennedy 1974;Lederman et al 1990; Millar 1994; Pathak and Pring 1989; Pring and Walker 1993), though in some studies the rate is over 60% (Katz 1946/1991; Kennedy 1997; Shimizu et al 1993).

A more optimistic view is that haptic examination of pictures provides reliable and useful information. Overall rates in a study can be low, but some pictures may be highly recognisable. While several possibilities may come to mind when exploring a haptic picture, over time the participant may assess all of the lines in a haptic picture, and determine astutely that many or all fit a possible referent. Each part of the picture may correspond to a significant part of a possible referent. Each part may seem in proportion to every other part; relative lengths of legs, arms, and body may seem in keeping with each other. The order of the shapes may be apt, with the head, body, and legs in the proper relative locations, for example. The outlines may readily stand for surface borders, as they do for vision, since through haptics we obtain information about surfaces, and make out edges, corners, wires, and cracks. Like sight, haptics allows the perceiver to detect some objects covering others, or interposed in front of
more-distant targets to which we are reaching, and thus vantage point effects in pictures may be intelligible to some extent (Kennedy, in press).

Not only may a perceiver decide a picture is a “good fit” to a given object, he or she may be able to assess confidently that another referent would offer a partial fit. Only a fraction of the picture’s lines may seem apt. Some of the shape may seem inappropriate, the proportions off, or the parts not to be in the proper order. The perceiver could estimate the proportion of features that fit, and use this to decide the likelihood the suggested identification is correct.

The positive and negative views of haptic pictures have implications for memory for haptic pictures. Pictures with parts we fail to discern may attract low fit judgments and be hard to recall or recognise on a later occasion compared with ones with lines that provide fully coherent information. The lower the fit judgment, the more likely some distinctive features (Gibson 1969) are not recognised. These might trigger a new response on a second trial.

The value of fit judgments was tested in a series of experiments on: fit predicting successful identification and the frequency with which suggestions are repeated on a later trial (experiment 1); fit as an aid to memory (experiment 2); consistency across subjects (experiment 3); confidence in low fit judgments (experiment 4); and the close relation between visual and haptic fit judgments (experiment 5).

2 Experiment 1

D’Angiulli et al (1998) found that blind children and blindfolded sighted children repeated successful identifications on a second trial (95% of the correct responses and only 3% of the incorrect suggestions in the case of the sighted, 100% versus 24% in the case of the blind). Likewise, in our experiment 1 blindfolded, sighted adults explored pictures twice, trying to identify as many pictures as possible on both trials. On both trials they also estimated the fit of the pictures to the labels which they volunteered. We asked if judgments of fit predict accuracy of identification on trial 1 and trial 2, and repetition of responses across trials.

2.1 Method

2.1.1 Participants. Twelve university students (mean age = 23.1 years, SD = 7.0 years; seven women) from an introductory class in Psychology participated. Like all the subjects in the experiments here, they were recruited at the University of Toronto at Scarborough, had no previous experience with raised-line drawings, had normal vision, and participated for course credit or as part of a class project. All were tested individually.

2.1.2 Materials. The materials used in experiment 1 were twelve raised-line drawings produced with Swedish raised-line drawing kits (from SIH Laromedel, Solna, Sweden). The pictures were derived from ones used by D’Angiulli et al (1998), Heller (1989), Kennedy (1974), and Richardson et al (1998).

Eight pictures were used on trial 1: hammer, key, person, scissors, swan, toothbrush, fork, and umbrella. The same pictures were used on trial 2, plus two extras, a car and a sailboat. These served to show not all the pictures on trial 2 had appeared previously, and to discourage the participants from simply recalling what had been said on trial 1. A raised-line drawing of a star was used for a demonstration. A picture of a tree was used for practice. Figure 1 shows the pictures and their long dimension (horizontally or vertically). The largest was 22.5 cm on its large dimension and the smallest was 15.5 cm. This is 50%–100% bigger than the pictures used by Kennedy (1974) and Heller (1989) and 30% larger than the figures used by D’Angiulli et al (1998), to increase the likelihood that subjects would detect small features of the displays and their proportions. The presentation orders of the eight pictures were randomised across participants and differed on trial 1 and trial 2 for each participant.
2.1.3 Procedure. Participants were tested individually. A raised-line star was shown to the participants visually, and they were invited to explore it tactually. The participants were blindfolded, given practice on a raised-line drawing of a tree, and asked to identify raised-line drawings of common objects in their normal orientation, one picture at a time, with a maximum of 120 s per picture, and were told a response should be offered even if not sure. (Responses were often offered after 60 s and before 120 s.) Subjects were to estimate how closely the pictures fit their responses on a 7-point scale (1 low, and 7 a perfect fit). No feedback about accuracy was given for any response.

Between trial 1 and trial 2 participants had a 3 min break. At the start of trial 2, participants were told that some of the pictures might be ones they had been presented with before, some might be new, and again they should try to identify as many pictures as possible. On trial 2 two distractors were added: a car and a sailboat, as the first and fifth pictures respectively. All of the eight pictures shown on trial 1 had a copy, used on trial 2, to avoid participants noting minor irrelevant features of the picture sheets on trial 1 and using them on trial 2.

2.2 Results and discussion

We first consider accuracy, and then fit.

On trial 1, the mean number of correctly identified pictures per participant was 4.9 (SD = 1.9) or 61%. This was a high recognition rate compared with several previous studies (D'Angiulli et al 1998; Heller 1989; Kennedy 1974, 2000). The large pictures used here may be easier to explore for detail than pictures in previous studies.
On trial 2, the mean number of correctly identified ‘old’ pictures per participant was 5.5 (SD = 1.6), ie 69%, a slight increase over trial 1. Including the distractors, the mean number of correctly identified pictures per participant was 7.1 (SD = 1.7), 71%.

Data from one participant who identified all the pictures on both trials were removed from the analyses to allow comparisons of correctly and incorrectly identified pictures.

As shown in table 1, on trial 1, counting only the eleven participants who offered erroneous responses, the mean fit judgment for the correctly identified pictures per participant was higher than that for the incorrectly identified pictures (paired t-test \( t_{10} = 4.77, p < 0.001 \)). On trial 2, the mean fit judgment for the correctly identified eight ‘old’ pictures per participant was higher than the mean fit judgment for the incorrectly identified ‘old’ pictures (\( t_{10} = 4.85, p < 0.001 \)), and for all ten figures (eight ‘old’, two ‘new’) the mean fit judgment per participant for the correctly identified pictures was higher than that for the incorrectly identified pictures (paired t-test \( t_{10} = 6.12, p < 0.001 \)). Evidently, correct identifications had higher fit judgments on trials 1 and 2.

**Table 1.** Mean fit judgment per participant of correctly identified pictures and incorrectly identified pictures: experiment 1.

<table>
<thead>
<tr>
<th>Pictures identified</th>
<th>Trial 1 (eight pictures)</th>
<th>Trial 2 (eight pictures)</th>
<th>Trial 2 (ten pictures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M )</td>
<td>5.6</td>
<td>5.8</td>
<td>5.8</td>
</tr>
<tr>
<td>( SD )</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Incorrectly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M )</td>
<td>3.6</td>
<td>3.8</td>
<td>3.7</td>
</tr>
<tr>
<td>( SD )</td>
<td>1.2</td>
<td>1.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>

When comparing fit judgments of repeated responses and changed ones, data from another participant were dropped from further analysis because this participant repeated 100% of her responses. For the remaining ten participants, the mean percentage of correct responses repeated per participant (97%, SD = 6%), was higher than the mean percentage of incorrect responses repeated per participant (51%, SD = 38%; \( t_9 = 3.78, p < 0.004 \)).

The accuracy of identification of the distractors on trial 2 was 71% for the car and 93% for the sailboat, high rates of identification. Eight participants correctly identified both distractors. For the remaining four participants, and considering the two distractors only, the mean fit judgment for the correct responses per participant was 5.75 (SD = 1.5), higher than the mean fit judgment for the incorrect responses per participant (2.75, SD = 1.7; \( t_3 = 4.24, p < 0.05 \)), again indicating that fit judgments predict identification rate.

Repetition of responses was also related to fit judgments. The mean fit judgment per participant of trial 1 responses that were repeated on trial 2 (\( M = 5.3, SD = 0.7 \)) was higher than the mean fit judgment per participant of trial 1 responses that were changed (\( M = 2.9, SD = 1.4 \); \( t_9 = 2.9, p < 0.02 \)).

The mean number of changed responses per participant was 2.2 (SD = 1.4). On trial 1, the mean fit judgment per participant for the responses that later changed was 2.9 (SD = 1.4) but on trial 2 the mean fit judgment per participant for new responses increased to 4.1 (SD = 1.8; \( t_9 = 2.2, p < 0.055 \)). Despite the small number of responses (22) on which the analysis is based, this is marginally significant, indicating that participants were likely dissatisfied with some of their trial 1 responses and were seeking a higher fit.
In sum, high fit judgments accompany responses likely to be correct. Responses with high fit judgments are likely to be repeated, and those with low fits are likely to be changed to suggestions that offer a more satisfactory fit, and higher accuracy. The results indicate that subjects can assess whether a referent (say “scissors”) fits a given figure well, with all the shapes having proper convex and concave curves and proportions, and can also determine that another (say “butterfly”, offered by one subject) fits some of the shape, but has some anomalies (such as very narrow wings at the top). Participants monitor their own performance.

3 Experiment 2

Experiment 2 tested the relation of judged fit to recognition memory.

A low fit judgment may indicate that the subject did not organise the display into a single coherent unit. If a picture of a horse is correctly identified, then the participant can remember this picture as a single chunk, efficiently, since details do not have to be memorised separately. But if a picture of a dragon is mistakenly called a horse, it will not be a single unit. It may seem to have odd nostrils, extra protuberances on the head whose function is unclear, surplus curves in the body, an unusual tail, unexpected proportions, and legs and hooves drawn strangely.

In the participant’s effort to remember ill-fitting pictures, extra parts and their positions would present difficulties. Some of the features may be nonspecific—“odd shapes, possibly hooves”, “ill-proportioned body”, and “some protuberances”. Nonspecific features could make identification difficult.

In experiment 2 the pictures from experiment 1 were used in a recognition memory task. However, to avoid ceiling effects, the task in experiment 2 was made challenging. All the pictures in a small set can be easily remembered. Multiplying the number of pictures several-fold was not appropriate given the time it takes to explore a picture tactually. Therefore the task was made more challenging in four ways. First, we added several new pictures to the initial trial. Second, we reduced picture exposure times, especially in the recognition-memory task. Third, we presented several new distractor pictures in the recognition-memory test. Fourth, we presented the displays in the recognition-memory task in different orientations.

3.1 Method

3.1.1 Participants. Twelve students (mean age $\pm 21.2$ years, SD $\pm 2.0$ years; seven women) participated in experiment 2.

3.1.2 Materials. The twelve pictures used in trial 1 were: hammer, key, person, scissors, swan, toothbrush, fork, umbrella, cup, sailboat, table, and telephone. The twelve pictures were used, inverted, in trial 2 for a recognition-memory test, with six distractors (car, star, Christmas tree, triangle, house, coat hanger). A raised-line drawing of an apple was used for a demonstration. Figure 2 shows the displays, additional to those in figure 1, that we used. As in experiment 1, all of the pictures in trial 1 had two versions, one for trial 1 and one for trial 2.

3.1.3 Procedure. Participants were tested individually as in experiment 1 except that the exploration time was 90 s per picture. On trial 2 the participants were told that some of the pictures they were going to explore might be new, some might be old, and there were four possible orientations for the pictures—normal orientation, inversion (being rotated 180° in the plane), being turned 90° clockwise, and being turned 90° counter-clockwise. All four orientations were demonstrated visually with the drawing of the apple. On trial 2, all of the previous twelve pictures were shown inverted, mixed with six distractors that were presented in the four possible orientations. For all subjects, the car turned 90° clockwise from upright served as picture 2; the star as number 5; the Christmas
tree, inverted, as 9; the triangle as 11; the house turned 90° counterclockwise from upright as 13; and the coat hanger, inverted, as picture 16. The order of the twelve ‘old’ pictures was randomised so that no participant received the same order as any other, and the orders of the twelve pictures differed on trial 1 and trial 2 for each participant. On trial 2, exploration time was cut to 45 s per picture to avoid ceiling effects.

The participants were instructed to identify each picture and to estimate the fit of the picture to their response, ignoring the orientation of the picture. In addition, they were asked to answer the recognition-memory question—‘‘have you had the picture before?’’—once again ignoring any orientation change that might have occurred.

3.2 Results and discussion
Pictures recognised in trial 2 as having been present in trial 1 attracted higher fit judgments ($M = 5.4$, $SD = 0.7$) than those that failed to be recognised ($M = 2.5$, $SD = 1.1$; $t_{11} = 8.24$, $p < 0.001$), an outcome that supports the key hypothesis of experiment 2, namely that fit predicts recognition memory. Subjects correctly said “yes” (this was presented on trial 1) to a majority (mean 8.75, range 5–11) of the twelve repeated pictures, and they correctly responded “no” to almost all distractors. (Ten said “no” to all six distractors, but one subject said “yes” to 1 display, and one said “yes” to 2.)

In matters of identification, the results of experiment 2 were similar to those of experiment 1. On trial 1, the mean number of pictures correctly identified was 63% per participant, much like the 61% on trial 1 of experiment 1, and the mean estimate of fit per participant for the correctly identified pictures ($M = 6.0$, $SD = 0.7$), was higher than that for the incorrectly identified pictures ($M = 2.8$, $SD = 0.8$; $t_{11} = 8.63$, $p < 0.001$). Also, 89% ($SD = 17\%$) of the correct responses per participant on trial 1 were repeated, whereas only 40% ($SD = 23\%$) of the incorrect responses on trial 1 per participant
were repeated ($t_{11} = 5.70, p < 0.001$) in accord with the results of experiment 1 and D’Angiulli et al (1998). The mean fit per participant for the identifications on trial 1 that were repeated on trial 2 ($M = 5.7, SD = 0.6$) was higher than the mean fit of the identifications on trial 1 that were changed on trial 2 ($M = 2.6, SD = 1.0; t_{11} = 9.95, p < 0.001$).

In summary, recognition memory was found to be better for high-fit pictures. In addition, despite the inclusion of memory tasks, shorter exploration times, and orientation changes, experiment 2 confirmed that fit judgments predict both identification accuracy and the likelihood of repetition. We suggest that low fit pictures may have some lines for which subjects do not settle on a referent, some shapes that do not match proportions of the referent the subject offers, and other lines that seem out of place in the referent elected by the subject, making recognition memory difficult. If a display is labeled a “telephone” all the parts correspond. Labeled a “typewriter” (a suggestion from one subject) the proportions are odd, the upper portion droops on both sides for unexplained reasons, the number of keys is rather few, and the curled line to the left has no clear role. It is likely that the more all parts of the display correspond to the suggested referent, the more it is accessible in our memory task.

4 Experiment 3

Experiment 3 tested whether one person’s fit judgments would correlate with another’s. The two judgments should be related since both ratings are based on physical correspondence between lines and possible referents, an objective matter that can be discerned through exploration of the shapes in the tactile display. If one subject deems a picture might be showing a car since it has a rectangular form above two semicircles, the second subject can assess the extent to which the shapes are indeed relevant to a car, and have the proper proportions. The subject might call an outline of a toothbrush a “saw” (as one did) but consider the long straight handle ill-suited to a saw. The next subject, asked about the fit of the picture to the referent “saw”, might also consider that only a fraction of the picture corresponds to a saw.

4.1 Method
4.1.1 Participants. Twelve students (mean age $\bar{x} = 24.2$ years, $SD = 8.2$ years; nine women) participated.

4.1.2 Materials. The raised-line drawings were those used on trial 1 of experiment 1.

4.1.3 Procedure. Each of the twelve participants was paired with one participant from experiment 1. The participants were given the example of a star, as in experiment 1. Then the participants were blindfolded, and asked to explore the eight pictures. The experimenter provided the participants with the identification responses to the pictures from their matched participants on trial 1 of experiment 1. As in experiment 1, the participants were given up to 120 s to explore each picture and were asked to judge the fit of each drawing to the response.

4.2 Results and discussion

We correlated the judgments of pairs of subjects, one from experiment 1, the other a yoked subject from the new group. If Sam from experiment 1 identified picture 1 as a swan, with a high fit of 7, and picture 2 (say the key) as a trumpet, but with a low fit of 3, did his match, Janet from experiment 3, offer similar ratings for these identifications? And so on for twelve pairs. The results were that the median Pearson correlation coefficient of the twelve pairs of participants was 0.6 (the sixth highest, with the seventh highest being 0.71) and the range was 0.04 to 0.9, all positive ($p < 0.0001$, two-tailed binomial), with six of the twelve coefficients significant beyond $p = 0.05$ ($p < 0.0001$, two-tailed binomial, 6, $N = 12, p = 0.05$).
In sum, the fit judgments of response–picture combinations from two groups were positively correlated. When one subject calls a display a “cup” and gives it a high fit judgment, the matched subject does so too. Presumably both subjects find the lines in the display correspond well to the referent. When one subject calls the picture of the toothbrush a “hammer for pounding meat”, as one did, both he and the matched subject rate it a low fit to the referent, a likely reason being that the hammer head only protrudes on one side of the handle, and it should protrude on both.

5 Experiment 4
Like high fit judgments, low fit judgments may be reliable and confident. A participant may feel a picture and say, “I think it might be a rabbit, because it has rabbit’s ears coming from an oval head, but the mouth is odd, if it is a rabbit, and the rest of the body seems quite misshapen for a rabbit, so the fit is just 2 (on the 7-point scale, 7 high). I am fairly confident about this fit judgment, so my confidence judgment on a 7-point scale (7 high) for this fit judgment of 2 is 5”. Here, one would have a low fit judgment (2), but a fairly high confidence judgment (5) about this low fit judgment.

Poor fit is based on an assessment of the correspondence between physical features of the picture (perceived structure) and physical features of a possible referent (deduced identity). It is not just a matter of low confidence. Physically, the shapes and order of the lines do or do not match the referent’s parts. Both high fit and low fit are physical facts the participant is checking. Hence, if the labels given by one participant to haptic pictures are given to a second participant, and fit judgments for the picture–label pairs are required, followed by confidence judgments about the fit judgments, then the fit judgments from this second participant may have high or low ratings, depending on the accuracy of the label, but the confidence judgments should have a smaller range and the difference in range should be due to the low fit judgments being coupled with relatively high confidence judgments. These two hypotheses were tested in experiment 4.

Two pictures per participant from experiment 1 were selected. One had been given the highest fit by the participant and the other the lowest fit. The two pictures plus their labels were then given to a fresh participant. Following experiment 3, the participants in experiment 4 should offer a high and low fit judgment to the two picture–label combinations, respectively. However, a large gap between fit judgments should be accompanied by a relatively small gap in confidence judgments about the fit judgments.

5.1 Method
5.1.1 Participants. Twelve students (mean age = 24.0 years, SD = 7.5 years; eight women) participated.

5.1.2 Materials. Twelve pairs of raised-line drawings were selected, from results from trial 1 in experiment 1, one pair for each participant in experiment 4.

5.1.3 Procedure. All participants in experiment 4 were given two drawings and their attempted identifications by a matched participant on trial 1 in experiment 1. Each pair contained a picture–response combination given the highest fit by the experiment 1 participant and one given the lowest fit. (In the event of a tie, the first drawings given those levels of fit were used.) The participants in experiment 4 gave their own fit judgments to the drawing–response combinations (1–7, 7 high), one at a time. And they also offered confidence judgments about the fit judgments (1–7, 7 high).

For example, for participant 1 of the current experiment, the picture that received the highest fit judgment from the matched participant in experiment 1 (trial 1) was the one called an “umbrella” (fit judgment of 7, for the picture of an umbrella) and the first given the lowest fit judgment was the one called a “duck” (fit judgment of 4, for the picture of a swan), respectively. So we showed these two pictures, i.e. the umbrella...
and the swan, to participant 1 in this experiment, and asked for fit judgments for the “umbrella” and “duck”, and confidence judgments for the fit judgments.

Participants were blindfolded and shown the two pictures one at a time. The order of the two pictures was the same as that for experiment 1. For each picture, subjects were given a maximum of 60 s to explore and respond.

5.2 Results and discussion
The difference between fit judgments for the two pictures per participant ($M = 3.0$, $SD = 1.9$) was twice that of the difference between the confidence judgments ($M = 1.5$, $SD = 1.5$; $t_{11} = 3.32$, $p < 0.007$).

For every participant, the drawing-and-response combination that received the higher fit judgment was also the one that received the higher confidence judgment (or the participant was equally confident of both fit judgments). Although the mean of the high fit judgments ($M = 6.6$, $SD = 0.4$) was slightly higher than the mean of the matched confidence judgments ($M = 6.4$, $SD = 0.8$), the difference was not significant.

On the other hand, the mean of the low fit judgments per participant ($M = 3.6$, $SD = 1.8$) was significantly lower than the mean of the matched confidence judgments ($M = 4.9$, $SD = 1.5$; $t_{11} = 2.35$, $p < 0.04$). Therefore, the difference between fit and confidence judgments is due to relatively high confidence ratings for low fit judgments. When a display is called “a swan”, the subject may deem all the features of the display correspond to the referent, and rate it a good fit confidently. When the display is called “a boot” (as the swan was once), many of the shapes may not correspond well to the referent. The “ankle” may seem very thin, and unnaturally curved, for example, and the “toe” of the boot swollen. The subject may deem this to be at best a low fit, and do so confidently.

6 Experiment 5
If haptic exploration obtains inadequate form information from pictures and, in particular, if it extracts different information from outlines than vision, visual criteria should be unrelated to haptic judgments of fit. However, if fit judgments assess the correspondence between the shapes in the picture and the shape of the relevant referent, and criteria used by vision are relevant to haptic pictures, then visual fit judgments would be akin to haptic ones.

One subject using haptics identified the drawing of a person as “a flower in a pot”, another said “a flower”, and one identified it as a drawing of a stool. These referents have some features relevant to the pattern in the display, subjects using vision may admit. However, many subjects using haptics identified the drawing as showing a person and subjects using vision would surely readily agree. But, would the low and high fit ratings of the match to “flower” and “person” from a subject using vision be like those from subjects using haptics? If the ratings are based on the correspondence of the display’s structure to the suggested referent then similar fit ratings should result.

6.1 Method
6.1.1 Participants. Twelve students (mean age = 23.6 years, $SD = 6.8$ years; seven women) participated.

6.1.2 Materials. The raised-line drawings used on trial 1 in experiment 1 were used in experiment 5.

6.1.3 Procedure. Participants were shown the demonstration picture of a star visually while the scale for the fit judgment was explained. Then the eight test pictures used on trial 1 of experiment 1 were shown to the matching participants in the current experiment. Each participant was assigned a participant from experiment 1. Each received the same order of pictures as the participant in experiment 1 to whom he or
she was matched. The participants were asked to judge, visually, the fit of the picture to the responses given by the participants from experiment 1, using the 7-point scale (7 high). Each participant responded within seconds.

6.2 Results and discussion
Each of the twelve participants in experiment 5 offered a visual fit judgment and the result was a pair of judgments—a tactual fit judgment from experiment 1 and a visual fit judgment from experiment 5. Hence, there were eight pairs of judgments from each pair of participants to be correlated. The median Pearson correlation coefficient for the twelve pairs of participants was 0.56 (sixth highest, with seventh highest being 0.62, and range 0.07 to 0.96), all positive ($p < 0.0001$, two-tailed binomial), and seven of the correlations had a probability that was lower than $p < 0.20$ ($p < 0.0001$, two-tailed binomial, $7, N = 12, p = 0.20$).

In sum, the judged fit of the response-picture pairs from one group exploring tactually predicts the judged fit from a second group examining visually. If a figure was called a “key” by the subject using haptics, and given a high fit rating, presumably its lines, shapes, and proportions were taken to be in good correspondence with the referent. The sighted subject should also regard the correspondence highly. On this account, when the display was called a “running shoe” by a subject exploring it haptically, it attracted a low fit rating, likely to be because the “heel” of the “shoe” is swollen out of proportion, and curved where it should be flat along the ground, the ‘toe’ has a point rather than a curve, etc. It is likely that the low ratings for the picture-plus-label pair from a subject asked to provide a visual check were given because the same failures of correspondence were evident.

7 General discussion
D’Angiulli et al (1998) found blind and sighted children corrected themselves when identifying pictures a second time, without feedback. Participants judged the picture–response pairings by their own standards, and their standards are accurate and similar across subjects. Here, we have offered a basis for this conclusion, and a series of experiments testing its implications. We suggest that the participants assessed a physical correspondence between the referent they mentioned and the way the picture suggested it. The higher the rating, the more likely the identification is to be correct, we suggest. The correspondence relation depends on shapes of lines being in keeping with shapes of parts of objects, matching proportions, and an apt order of shapes, accessed similarly in vision and haptics.

We found that high fit judgments accompany responses likely to be correct and repeated. Recognition memory is better for a picture that was part of a high-fit pairing. Confidence judgments are relatively high for low fit judgments. The fit of response–picture pairs judged visually predicts the fit given with haptics.

While it is likely that a high fit judgment indicates that each line in a picture corresponds to a part of the suggested referent, a lower fit judgment may indicate that only a few of the display’s lines correspond to features of the referent, that proportions are unusual, or that the order of the parts was odd. For example, the ears of a rabbit might seem to be attached to the neck. The fewer features of the display that can be made to correspond with parts of the referent, the less likely the suggested identification is to be correct.

In many respects vision and haptics are rather different. However, in some key aspects related to outline pictures they may have many characteristics in common. One major difference is that the observer’s vision can survey pictorial displays noticeably faster than haptics. However, principles such as good continuation and proximity may operate similarly in the two senses (Heller 2000; Kennedy 1974). Also, subjects using either vision
or haptics may use lines in outline drawings for similar referents (Kennedy 2000, in press),
namely edges of surfaces forming corners or occlusions. Consequently, many of the
principles which govern fit judgments of pictures may be similar in the two modal-
ities.

Though subjects using vision or haptics may rely on the same criteria in dealing
with many pictures, this may not be true for all pictures of course (Hopkins 2000;
Lopes 1997). There are many outstanding questions about the best ways to depict
objects in both vision and haptics. Further, there may be some principles that apply
more readily to vision than to haptics, eg some aspects of perspective convergence
and foreshortening (Heller et al 1995; Hopkins 2000; Kennedy, in press). In this regard,

Our purpose here has been relatively modest: to show that participants have a
useful ability to assess the relative fit of referents they deem relevant to haptic pictures.

Contemporary theory of haptic picture perception holds that it is not necessary to
have been sighted for a person to recognise objects in pictures (Kennedy 2000). The
most radical version of this theory would be that the modality through which we
obtain pictures of the world does not matter. This is surely exaggerated. Taken literally,

Since the role of projection to a vantage point in haptic pictures is controversial,
we took the opportunity a posteriori to distinguish between two kinds of pictures in
experiment 2, which contained the largest sample of pictures.

There were some flattish objects in the stimuli used in trial 1 of experiment 2. The
fork, hammer, key, scissors, swan, and toothbrush “can make a recognisable imprint
on a flat surface because they have little depth [and few] overlapping parts” (Kennedy
1974, page 151). These six stimuli can be called “imprints”. The other six stimuli in trial 1
of experiment 2, namely the person, cup, umbrella, sailboat, table, and telephone, are
more three-dimensional figures. “Three-dimensional figures have overlapping parts
and result in projective distortions” (Kennedy 1974, page 151). The top of a table is a
rectangle but is “projecting a parallelogram” (Kennedy 1974, page 152). Accordingly,
these stimuli can be called “projections”. If features such as overlap of occluding edges,
with foreground and background surfaces, mean little to haptics, projections should
be difficult to identify. However, if observers using haptics appreciate overlap and
parallel projection of forms readily, there should be little difference between imprints
and projections.

On trial 1 of experiment 2, the mean identification rate for the six imprints per parti-
cipant was 62% (SD = 18%), while the mean identification rate for the six projections
per participant was 64% (SD = 16%), evidently not significantly different. Also on trial 1 of experiment 2, the mean fit judgment for the correctly identified projections per participant was 5.9 (SD = 0.8), and mean fit judgment for the correctly identified imprints per participant was close at 6.1 (SD = 0.7; \( t_{11} = 1.02, p < 0.34 \)). Kennedy (1974) and Lederman et al (1990) reported similar findings testing blind participants. We speculate that the blind and the sighted understand some aspects of parallel projection in haptic pictures relatively straightforwardly (Eriksson 1998). The reason may be that parallel projection presents many aspects of object size without changing the sizes of far and near sides of objects.

Hopkins (2000), Kennedy (1993), and Lopes (1997) discussed the possibility that convergence may arise in some aspects of haptics. If we reach out to the four corners of a table from a vantage point in the middle of one side we find there is a smaller angle between the directions of the far corners, compared with the directions of the near corners. The far corners may subtend 90° and the near corners close to 180°. However, the relation between haptic directions and haptic pictures remains unsettled, and the debate about perspective and haptic pictures is far from concluded.

The moderate theory we noted above implies that untrained blind people should be able to recognise haptic pictures with some kinds of projection. In this moderate theory, there should be a role for the participant’s own ability to assess a picture for many possible referents, and then to sort out which referent fits the picture best. The present studies, together with D’Angiulli et al (1998) and D’Angiulli and Kennedy (2000, 2001), indicate it may be that many people, children and adults, blind or sighted, can judge the fit of their responses to haptic pictures. Fit judgments may be especially helpful in distinguishing between several ways of showing an object (Kennedy 1993; Kennedy and Merkas 2000). An interesting possibility is that fit judgments may be sensitive to subtle variations, so that two projections of an object or two pictures readily identified as pictures of horses, with similar recognition rates, may differ in fit judgments, indicating that one is a better representation for some purposes than the other.

In summary, fit judgments of haptic pictures are related usefully to identification in several ways, including accuracy, repetition, remembering, confidence, and visual judgments.

Acknowledgments. We thank Barry Hughes and two other Perception reviewers for helpful comments on a draft of this paper.

References
Arnheim R, 1990 “Perceptual aspects of art for the blind” Journal of Aesthetic Education 24 57 – 65
Gibson E, 1969 Principles of Perceptual Learning and Development (New York: Appleton-Century-Crofts)
Heller M A, 1989 “Picture and pattern perception in the sighted and the blind: The advantage of the late blind” Perception 18 379 – 389


Kennedy J M, in press “Touch and the observer’s vantage point”, in *Unfolding Perceptual Continua* Ed. L Albertazzi (Amsterdam: Benjamins) pp 181–204


Lopes D M M, 1997 “Art media and the sense modalities: Tactile pictures” *Philosophical Quarterly* 47 425–440


Richardson B, Symmons M, Kennedy J M, 1998 “Findings with the tactile display system”, poster presented at the Conference on Representation and Blindness, San Marino, May