

Running head: MULTIMODAL SIMILARITY AND CATEGORIZATION

Multimodal Similarity and Categorization of Novel, Three-Dimensional Objects

Theresa Cooke

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Frank Jäkel

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Christian Wallraven

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Heinrich H. Bühlhoff

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Theresa Cooke

Max Planck Institute for Biological Cybernetics

Spemannstr. 38

72076 Tübingen

Germany

e-mail: theresa.cooke@tuebingen.mpg.de

Abstract

Similarity has been proposed as a fundamental principle underlying mental object representations and capable of supporting cognitive-level tasks such as categorization. However, much of the research has considered connections between similarity and categorization for tasks performed using a single perceptual modality. Considering similarity and categorization within a multimodal context opens up a number of important questions: Are the similarities between objects the same when they are perceived using different modalities or using more than one modality at a time? Is similarity still able to explain categorization performance when objects are experienced multimodally? In this study, we addressed these questions by having subjects explore novel, 3D objects which varied parametrically in shape and texture using vision alone, touch alone, or touch and vision together. Subjects then performed a pair-wise similarity rating task and a free sorting categorization task. Multidimensional scaling (MDS) analysis of similarity data revealed that a single underlying perceptual map whose dimensions corresponded to shape and texture could explain visual, haptic, and bimodal similarity ratings. However, the relative dimension weights varied according to modality: shape dominated texture when objects were seen, whereas shape and texture were roughly equally important in the haptic and bimodal conditions. Some evidence was found for a multimodal connection between similarity and categorization: the probability of category membership increased with similarity while the probability of a category boundary being placed between two stimuli decreased with similarity. In addition, dimension weights varied according to modality in the same way for both tasks. The study also demonstrates the usefulness of 3D printing technology and MDS techniques in the study of visuohaptic object processing.

Multimodal Similarity and Categorization of Novel, Three-Dimensional Objects

The question of whether similarity can provide a theoretical basis for general categorization behaviour has been a source of heated debate in the field of cognitive psychology (Goldstone, 1994; Hahn & Ramscar, 2001). Critics of this idea have argued that the notion of similarity is vague and context-dependent, that it cannot explain category coherence, and that it does not account for the important role of theoretical knowledge in categorization decisions (Murphy & Medin, 1985). Nonetheless, similarity has served as the basis for a number of influential models of categorization (Rosch & Mervis, 1975; Medin & Schaffer, 1978; Nosofsky, 1992), which have been particularly successful in explaining classification of perceptual stimuli, including novel, 3D objects (Edelman, 1999). However, much of this work has been carried out within the context of perception involving a single modality, usually vision. Considering similarity and categorization within a multimodal context opens up a number of important questions: Are the similarities between objects the same when they are perceived using different modalities or by more than one modality at a time? Is similarity still able to explain categorization performance when objects are experienced multimodally?

In a preliminary study (Cooke, Steinke, Wallraven, & Bühlhoff, 2005), we showed how multidimensional scaling (MDS) techniques can be used to quantify differences in perceptual similarities when objects are perceived using touch and vision. In that study, subjects saw or touched novel, 3D objects which varied parametrically in shape and texture and then rated the similarity between object pairs. Using similarity as a psychological distance measure, MDS was used to visualize stimuli as points in multidimensional perceptual spaces, as for example in Shepard and Cermak (1973); Garbin (1988); Hollins, Faldowski, Rao, and Young (1993). We found that the relative importance of shape and texture in these perceptual spaces differed according to modality:

shape alone sufficed to represent the stimuli when perceived visually, while shape and texture were both required when the stimuli were perceived haptically.

In the present study, we extend this line of research by adding a second task, free sorting categorization, and including a condition in which objects are simultaneously both seen and touched. The categorization task was included in order to test whether a connection between similarity and categorization could be established within this multimodal context. The bimodal condition was added in order to assess whether multimodal similarity and categorization would be dominated by one specific modality. At first glance, vision appears to be the most likely candidate. Vision is traditionally considered to be the "dominant" modality (Rock & Victor, 1964). Furthermore, object shape has been shown to play a special role in category formation (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Landau & Leyton, 1999) and shape is thought to be a particularly salient feature for vision (Klatzky, Lederman, & Reed, 1987). On the other hand, recent studies have challenged the notion of ubiquitous visual capture and have argued in favour of weighted averaging models (Guest & Spence, 2003; Ernst & Bühlhoff, 2004).

The results of this study show an effect of modality on the relative importance of object properties for both similarity and categorization tasks. In the bimodal condition, shape and texture were weighted roughly evenly for both tasks, rejecting the visual capture hypothesis. The probability of objects being grouped together in a category increased with similarity, while the probability of a category boundary being placed between two stimuli decreased with similarity. In addition, the relative importance of dimension weights for similarity and categorization tasks varied in the same way as a function of modality. The connection between similarity and categorization in the context of visuohaptic object processing is discussed in light of these findings.

Methods

This section describes the stimulus set, the psychophysical tasks, and the analysis techniques used in this study.

Stimuli

A family of 25 novel, 3D objects (Figure 1) was designed using the graphics package 3D Studio Max (3DS). The "base object" in the family (Figure 1, object 1) consists of three parts connected to a centre sphere, specifying its macrogeometrical structure ("shape") and a displacement map applied to this 3D mesh, specifying its microgeometrical structure ("texture"). The remaining family members were created by parametrically varying the macrogeometrical and/or microgeometrical smoothness of the base object. Macrogeometrical smoothing was accomplished by applying a mesh relaxation operator which locally averages angles in the mesh in five linearly increasing steps (before application of the texture displacement). Microgeometrical smoothing was performed by linearly decreasing the amount of mesh displacement allowed by the application of the texture map in five steps. It is important to understand that the specific values of these parameters are only meaningful within 3DS. In addition, one cannot assume that equidistant changes in a software parameter yield perceptually equidistant changes in object properties.

Once an object is created in 3DS, it can either be rendered into a 2D image or printed into a solid 3D model. Printing is performed by a rapid prototyping machine (Dimension 3D Printer, Stratasys, Minneapolis, USA). The manufacturing process involves a head which deposits filaments of heated plastic such that the model is built up layer by layer. The final result is a hard, white, opaque, plastic model. In our case, models weighed about 40 g each and measured 9.0 ± 0.1 cm wide, 8.3 ± 0.2 cm high, and 3.7 ± 0.1 cm deep. It took two to four hours to print each object. The same set of 3D

models was used in all experiments described below.

Similarity Rating Task

Thirty naive subjects (16 men, 14 women) were paid 8 Euros per hour to participate in the experiments. The first task was to rate the similarities between pairs of objects on a scale between 1 (low similarity) and 7 (high similarity). Ten subjects explored the objects visually only, ten subjects explored the objects haptically only, and ten subjects explored the objects using both modalities simultaneously. All subjects in the haptic and bimodal conditions were right-handed. The same experimental setup was used in all conditions (Figure 2). Subjects used a chin rest placed 30 cm away from the stand on which the objects were presented; the height of the chin rest was set such that the centre of the object was aligned with the line of sight. An opaque curtain hung between the subjects and the stand and could be slid back and forth along a rod to hide or reveal the objects. A piece of black sheet metal (30 cm x 30 cm) was mounted on the back of the stand so that when the curtain was open, subjects saw the object on a black background. A set of grooves and a section of rubber tubing on the mount piece ensured that the objects were securely held in place in exactly the same upright position on every trial.

In the visual condition, the experimenter placed the first object on the stand, slid the curtain over to reveal the object, waited for 3 to 5 seconds, slid the curtain back to cover the object, replaced the first object by a second object, uncovered the second object, and waited for the subject's response. In the haptic condition, the curtain was left in place while subjects explored the objects. Subjects were instructed to follow the contour of the objects with their right index finger. The contour-following procedure was selected because it has been shown to allow haptic extraction of both global shape and local texture properties (Lederman & Klatzky, 1993). Before the experiment, subjects practiced the procedure until they could trace the contour comfortably in 5 seconds or less. Subjects

in the bimodal condition performed the same contour-following procedure while viewing the objects. In the visual condition, the presentation time of the first stimulus was set by the experimenter, whereas in the haptic and bimodal conditions it depended to a certain degree upon the subject's exploratory movements. The presentation time of the second stimulus was determined by the timing of the subject's response in all three conditions. In general, we strove to maintain presentation times of 3 to 5 seconds for all conditions.

The experiment consisted of three blocks of 325 randomized trials (each object was compared once with itself and once with every other object resulting in $25 + (25 \cdot 24) / 2 = 325$ trials) and the order of appearance of stimuli was randomized over blocks. Each trial took about 20 to 30 seconds and the similarity ratings experiment ran for approximately two hours per day for five consecutive days. The experiment began with a number of practice trials to help subjects become accustomed to the task. After the practice session, subjects were asked to write down their criteria for each value on the rating scale (e.g., "I say 7 whenever the objects are exactly the same"). On each of the subsequent days of the experiment, subjects were asked to read what they had written to ensure consistency over the course of the experiment.

Debriefing Questionnaire

Immediately after the similarity ratings experiment, subjects filled out a form in which they were asked to describe the objects ("How do the objects look?" in the visual condition, "How do the objects feel?" in the haptic condition, and both questions in the bimodal condition), to explain how they had performed the similarity judgments, and to describe how they would group the stimuli into categories.

Free Sorting Categorization Task

After having filled out the questionnaire, subjects performed a free sorting categorization task. We chose a free sorting task because of its relative simplicity and the

ecological relevance of spontaneous categorization, also referred to as category construction (Milton & Wills, 2004). We designed the task to make it as close to the similarity task as possible. Using the same setup, stimuli were shown one at a time in random order and subjects explored the stimuli using the same exploratory procedure which they had used before. They were asked to assign a category number to each object, using the groups they had described in their questionnaire responses. The stimuli were repeatedly cycled through until the subject assigned the same category number to each object twice in a row.

Analysis Techniques

Analysis of similarity data. A multidimensional scaling (MDS) technique was used to analyze the similarity data. MDS techniques take pair-wise proximities data for a set of objects (human similarity ratings in this case) and return the coordinates of the objects in a multidimensional space which best explains the proximity data. We used the individual differences weighted Euclidean distance model implemented as part of the ALSCAL MDS package in SPSS (Young & Harris, 2003; Carroll & Chang, 1970), with proximity data considered to be ordinal measurements (i.e., non-metric) and untying of tied proximities allowed. This particular technique was chosen because it allows for comparison of individual subject data and because the dimensions of the resulting spaces are uniquely specified, allowing for clearer interpretation (Borg & Groenen, 1997; Cox & Cox, 2001). The algorithm takes as input a set of individual subject similarity data and, for a specified dimensionality, returns a single underlying stimulus configuration together with a set of subject-specific weights. The weights specify how the underlying configuration should be scaled along each dimension to best fit each subject's similarity data. In addition, the SPSS implementation provides a goodness-of-fit measure, Young's S1 Stress, which is the normalized difference between the fitted distances and the observed proximities.

It is important to note that the weighted individual differences MDS model we used carries with it the following assumptions: 1) that the appropriate metric for the psychological similarity space is Euclidean¹ and 2) that each set of individual subject data included in the analysis can be modeled by linear stretching of the centroid configuration, as specified by the individual subject weights. If these assumptions hold true, one expects low stress values for the overall MDS solution. Although establishing a threshold for acceptable values of stress is notoriously controversial, Monte Carlo studies suggest that stress values below 0.2 are indicative of an output configuration which provides a good fit to the similarity data (Cox & Cox, 2001).

Analysis of categorization data. In the free sorting task, our observations consisted of: 1) the category membership which the subject assigned to each stimulus, 2) the total number of categories created by the subject, and 3) the number of repetitions of the free sorting task that was required before the subject provided the same categorization twice in a row. In addition to these raw measures, we calculated a measure of the relative importance of texture as compared to shape in the categorization task. The measure we chose relies on the assumption that subjects perceived the shape-based and texture-based adjacencies in the stimulus map, which indeed turned out to be the case (see Results). Subjects' free sorting categories were superimposed on the stimulus map (Figure 1). When this is done, the boundaries between categories cross a certain set of adjacencies in the map. Each boundary separates two neighbours either on the basis of a difference in shape or on the basis of a difference in texture. We used the proportion of separations based on differences in texture as our measure of the relative importance of texture compared to shape for categorization.

Analysis of the relationship between similarity and categorization. We computed two basic measures of correlation between similarity and categorization data. First, for each value on the similarity scale (1-7), we created groups of stimulus pairs which had received

that particular rating. We then divided each group into two subgroups according to whether subjects had placed the two objects into the same category or into different categories. For the second measure, we selected those pairs of stimuli which are neighbours along either the shape or texture dimension and computed the probability of subjects setting a category boundary between these neighbours as a function of their perceptual similarity.

Analysis of subject questionnaires. Questionnaires were read and scored independently by three judges (two authors and one external judge). We evaluated which object properties subjects mentioned when 1) describing the object, 2) describing how they made similarity judgments, and 3) describing how they would categorize the objects. For each of these points, we evaluated whether the subject had made at least one reference to the attributes shape, texture, material properties, colour or temperature. We additionally distinguished between references to global shape and part shape. A reference to part shape was defined as the explicit mention of one of the objects' parts ("leg", "centre ball"), while a reference to global shape was defined as the use of holistic shape term such as "star-like" or a reference to part configuration, such as "three ends which extend from a ball-shaped center". When subjects mentioned a configuration of parts, we counted this as both a reference to global and part shape.

Results and Discussion

This section presents the results of the similarity rating and free sorting categorization tasks, the debriefing questionnaires, as well as results obtained by comparing data from the similarity and categorization tasks.

Results of Similarity Rating Task

Number of underlying perceptual representations. A critical question for this study is whether allowing for separate, *modality-specific* stimulus representations provides a better

explanation of the data than a single, multimodal representation which combines information from touch and vision. Recall that the weighted individual difference MDS model makes the following assumption: the data under consideration can be explained by a single underlying map and a set of dimension weights which are individually-adjusted for each subject (see Methods). The better this assumption holds true for a given data set, the lower the MDS stress will be. Here, we use this idea to evaluate whether 1) a single, multimodal representation with individual weights or 2) three, modality-specific representations provide better fits to our data by comparing stress values from 1) a global MDS computed over all similarity data and 2) three separate MDS solutions, one for each set of modality-specific similarity data. The stress for a two-dimensional MDS solution using grouped similarity data from all modality conditions was 0.167. Using modality-specific similarity data sets, stress values were 0.157 (visual), 0.168 (haptic), and 0.160 (bimodal). Since stress values were below 0.2 in all cases, all two-dimensional solutions provided good fits to the respective sets of similarity data (see Methods). The similarity amongst these stress values indicates that positing a single, multimodal representation of the stimuli provides an equally good explanation for our data as positing three separate representations.

Dimensionality of the underlying perceptual representation. Two perceptual dimensions were recovered in all MDS analyses, i.e., increasing the dimensionality of the spaces did not produce substantial decreases in stress. This shows that subjects recovered a two-dimensional representation using visual, haptic, and visuohaptic exploration. Subjects' descriptions of how they made similarity judgments (Figure 7) confirmed that they perceived the two dimensions as "shape" and "texture". Although several additional object properties (such as material, colour, and temperature) were mentioned when describing the objects, only shape and texture properties were mentioned when subjects explained how they made similarity judgments.

Topology of the underlying perceptual representation. The stimulus configuration resulting from the two-dimensional MDS solution computed over all similarity data is shown in Figure 3. This topology was qualitatively very similar to the topologies recovered with modality-specific MDS analyses. Ordinal adjacency relationships between the stimuli were preserved, demonstrating that subjects were able to recover the *ordering* of shape and texture variations in the stimuli, i.e., the adjacency relationships between neighbours in the map. This was also true of maps derived from separate analyses of individual subject similarities: there was perfect recovery of the ordinal relationships in 22/30 cases and recovery with one exception in 5/30 cases. As we have previously discussed (Cooke et al., 2005), this is a non-trivial task given the high dimensionality of the measurement spaces involved. To fully appreciate this, one need only consider the difficulty of recovering these relationships using computational methods, which we demonstrated in the aforementioned study.

It is also of note that the perceptual distances between stimuli, reflected in the MDS map, deviate from the distances between stimuli in the space defined by the manipulation of software parameters, in which stimuli lie on a rectangular grid (see Methods). This provides an important reminder that the *perceptual* distances between stimuli cannot be assumed to vary linearly with the distances defined by manipulating parameters in a software program, an issue often neglected in the growing number of studies involving parametrically-controlled stimuli (e.g., stimuli created using morphing techniques).

Modality-based analysis of dimension weights. The subject weights provided by the individually-weighted MDS can be visualized as vectors connecting each subject to the origin of a two-dimensional weight space. Because the sum of squared weights is constrained to a constant value, we simply calculated the angle between each subject vector and the texture axis of this weight space as a single variable which represents the relative weighting of shape and texture dimensions (Figure 4). The mean weights for the

ten subjects in the visual condition ($M = 0.16\pi$ radians, $SE = 0.01\pi$ radians) and for the ten subjects in the haptic condition ($M = 0.27\pi$ radians, $SE = 0.02\pi$ radians) were found to be significantly different ($t[14]=6.0$, $p<0.001$). The mean weights for the ten subjects in the visual condition and the ten subjects in the bimodal condition ($M = 0.25\pi$ radians, $SE = 0.03\pi$ radians) were also found to be significantly different ($t[10.6]=2.9$, $p=0.01$) using a two-sample, two-tailed t-test for independent samples with unequal variances. No significant difference in mean weights was found for the haptic and bimodal conditions ($t[13.8]=0.57$, $p>0.1$). These results show a clear effect of modality on the relative weighting of stimulus dimensions for similarity judgments. Note that the mean tradeoff value in the bimodal condition ($M = 0.25\pi$ radians) lay between the values obtained in the unimodal conditions (0.16π radians for vision, 0.27π radians for touch), although statistically speaking, there was no difference between the bimodal and haptic weights.

Next, we tested the hypothesis that the weights came from a distribution with a mean of 0.25π radians, representing equal importance for shape and texture properties in the subjects' similarity judgments. A single-sample t-test rejected the hypothesis for the visual condition ($t[9]=10$, $p<0.001$), but not in the haptic ($t[9]=1.0$, $p>0.1$) or bimodal conditions ($t[9]=0.1$, $p>0.1$). Taken together, these results show that shape dominated texture when similarity judgments were performed visually, while shape and texture were equally important when similarity ratings were performed either haptically or bimodally.

The individual data (Figure 4, right) shows that all visual subjects were indeed shape-dominated, while haptic subjects were quite evenly distributed around 0.25π radians (equal shape and texture weight). In the bimodal condition, 6/10 subjects weighted shape and texture quite evenly. However, two subjects weighted shape much more heavily than texture, and two subjects weighted texture much more heavily than shape. One explanation for the wide range of weights observed overall in the bimodal condition is that the involvement of two modalities requires an integration of information

from touch and vision to occur and that this integration process varies across subjects. For example, the reliability of shape/texture estimates may vary from subject to subject (e.g., as a function of relative expertise or familiarity with a given modality). Another possibility is that the integration process leads to a conflict between shape/texture weights dictated by the haptic and visual systems and that subjects attempt to resolve the conflict by making a conscious decision about the relative feature weights.

Results of Free Sorting Categorization Task

General task measures. Figure 5 (left) shows that subjects in all modality conditions performed the free sorting task with a very high degree of consistency, rarely requiring more than the minimum number of two iterations through the stimuli in order to provide the same categorization twice (visual condition: $M = 2.1$ iterations, $SE = 0.2$ iterations; bimodal condition: $M = 2.3$ iterations, $SE = 0.3$ iterations; haptic condition: $M = 2.4$ iterations, $SE = 0.3$ iterations). As shown in Figure 5 (right), there was a noticeable effect of modality on the number of categories created by subjects: subjects in the bimodal condition created more categories ($M = 8.3$ groups, $SE = 1.6$ groups) than subjects in unimodal conditions (visual: $M = 4.7$ groups, $SE = 0.7$ groups, haptic: $M = 5$ groups, $SE = 0.5$ groups). This could be due to a combinatorial effect of having redundant or conflicting information available from the two modalities. We also observed a tendency for subjects to use dimension-based rules to construct their categories: half of our thirty subjects appeared to use rules along a single dimension, while eight subjects combined rules along both shape and texture axes; categories constructed by the remaining seven subjects could not be well-described using combinations of unidimensional rules.

Modality-based analysis of dimension weights. As a measure of the relative importance of texture compared to shape, we calculated the proportion of category boundaries based on texture differences between stimuli, shown in Figure 6 (see Methods).

The mean proportion of texture-based boundaries for the ten subjects in the visual condition ($M = 26\%$, $SE = 10\%$) and for the ten subjects in the haptic condition ($M = 55\%$, $SE = 8\%$) was found to be significantly different ($t[17.5]=2.2$, $p=0.04$; two-sample, two-tailed t-test for independent samples with unequal variances), indicating that subjects in the visual condition relied more on shape in their categorization decisions than subjects in the haptic condition. No significant difference was found in between the mean proportion of texture-based boundaries used in the visual condition and in the bimodal condition ($M = 38\%$, $SE = 8\%$) nor was a significant difference found between the bimodal and haptic conditions (visual-bimodal: $t[17.3]=0.98$, $p>0.1$; bimodal-haptic: $t[18] = 1.4$, $p>0.1$). Thus, there was no statistical evidence in favour of more "visual-like" or more "haptic-like" use of stimulus dimensions in bimodal categorization, however the mean weight for bimodal categorization (38% texture-based boundaries) lay between the values obtained in the unimodal conditions (26% for vision, 55% for touch).

Next, we tested the hypothesis that weights came from a distribution with a mean of 50%, i.e., that subjects based category boundaries equally often on shape and texture differences. A single-sample t-test rejected this hypothesis for the visual condition ($t[9]=2$, $p=0.04$), but not for the haptic ($t[9]=0.6$, $p>0.1$) or bimodal conditions ($t[9]=1.4$, $p>0.1$). Strikingly, this is the same pattern which we observed for the similarity tradeoff values: shape dominated texture for visual categorization while shape and texture were roughly evenly weighted for both haptic and bimodal categorization. There was a large amount of individual variation in the relative importance of shape/texture for categorization across all modalities (Figure 6). However, 7/10 haptic subjects weighted texture as heavily or more heavily than shape, 9/10 visual subjects weighted shape as or more heavily than texture, and 6/10 bimodal subjects exhibited fairly equal weighting of the two properties, a pattern which supports the outcome of the t-test. The remaining variation could be due to the measure we computed, to the design of the free sorting task, or to intrinsic

modality effects. Further studies involving different categorization tasks are needed to disentangle these factors.

Results of Subject Questionnaires

Figure 7 shows the frequency with which subjects mentioned various object features when describing the objects (left column) and when explaining how they performed similarity ratings (centre column) and free sorting (right column). When subjects were asked to *describe* the objects, the frequency with which they mentioned various object features depended on modality. Shape was mentioned by 10/10 subjects in the visual condition (V), 9/10 subjects in the visuohaptic condition (VH), and 3/10 subjects in the haptic condition (H); texture was mentioned by 5/10 subjects (V), 10/10 subjects (VH), and 9/10 subjects (H); material properties were mentioned by 1/10 subjects (V), 2/10 subjects (VH), and 5/10 subjects (H); colour was mentioned by 4/10 subjects (V), 2/10 subjects (VH), and 0/10 subjects (H).

Interestingly, although subjects described the objects using a variety of features, they only mentioned shape and texture when asked to explain how they had performed the similarity and categorization tasks. For haptic subjects, there was a particularly striking difference in that 10/10 subjects mentioned shape for the similarity task and 9/10 subjects mentioned shape for the categorization task, even though only 3/10 mentioned it when describing the objects. One explanation could be that spontaneous description better reflects intrinsic modality feature biases (i.e., texture and material properties for haptics), whereas descriptions of features for similarity and categorization are more strongly influenced by the experimental task and stimulus set.

To help identify which aspects of the stimulus geometry play a role in the perceptual dimension "shape", we separated the subjects' references to shape into two categories: part shape and global shape (see Methods). Global shape (Figure 8, left) was

mentioned more often by visual than by haptic subjects for all tasks (description: 9/10 (V), 1/10 (H); similarity: 9/10 (V), 4/10 (H); categorization: 7/10 (V), 5/10 (H)). This could be explained by differing amounts of effort involved in extracting global shape in the two conditions: in particular, the contour following procedure causes haptic extraction of global shape to be slow and memory-intensive. Allowing subjects to enclose the objects, for example, would have provided a quicker, albeit cruder estimate of global shape (Lederman & Klatzky, 1993) and might have increased the mention of global form. Interestingly, bimodal subjects' mention of global shape falls between the values for visual and haptic conditions (description: 7/10 (VH); similarity: 7/10 (VH); categorization: 5/10 (VH)). Part shape (Figure 8, right) was consistently mentioned by subjects in all modality conditions when describing how categorization was performed (8/10 (V), 9/10 (VH), 8/10 (H)), which is consistent with the recognized importance of part information in basic level categorization (Tversky & Hemenway, 1984) and with the importance of part information in haptic categorization by blind and sighted children (Morrongiello, Humphrey, Timney, Choi, & Rocca, 1994).

Results on the Connection between Similarity and Categorization

Figure 9 (left) shows the number of stimulus pairs which were placed into either the same or different categories, sorted by similarity. The highest frequency of stimulus pairs being placed in different categories occurs for similarity ratings of 3 and 4, while the highest frequency of stimulus pairs being placed in the same category occurs for similarity ratings of 6. Note that the decrease in same-category occurrences for a similarity value of 7 and the decrease in different-category occurrences for similarity values of 1 and 2 are due to the relative infrequency of these similarity values; the *relative proportion* of different-category to same-category occurrences is indeed a monotonically decreasing function of similarity.

Figure 9 (right) shows the mean probability of subjects setting a category boundary between shape or texture neighbours in the stimulus map as a function of their perceptual similarity. Note that the fact that subjects consistently recover ordinal shape and texture structure in the stimulus set justifies the use of these adjacency relationships in the calculation of this measure. Here, we observe that the probability of subjects violating such an adjacency relationship decreases monotonically as a function of similarity. Again, the large amount of variance in the measure at a similarity value of 1 is due to the fact that adjacent stimulus pairs were rarely rated with a similarity of 1.

A further connection between similarity and categorization is that modality had the same effect on the relative weight of stimulus dimensions in both tasks. Shape was weighted more heavily than texture when similarity and categorization were performed visually, whereas shape and texture were weighted roughly evenly when the tasks were performed either haptically or bimodally (Figures 4 and 6).

General Discussion

Modality-Dependent Weighting of Stimulus Dimensions

We found that the relative importance of shape and texture for judging similarities between objects and for creating categories of objects varied systematically according to modality. For both similarity and categorization, shape dominated texture in the visual condition, while texture and shape were evenly weighted in the haptic condition. In the bimodal condition, we found texture and shape to be weighted evenly on average for both our similarity and categorization tasks. This finding is in agreement with Klatzky et al. (1987), in which subjects were instructed to sort wafers varying in shape, texture, size, and hardness based on their similarity. Substance dimensions (hardness and texture) were most salient after haptic exploration while saliency was evenly distributed across dimensions when subjects used both haptic and vision. When subjects were explicitly

instructed to use visual imagery to compare the objects after haptic exploration, shape became overwhelmingly dominant.

The results of this study replicate our previous results on unimodal shape and texture weights in similarity judgments (Cooke et al., 2005); together, these studies provide clear evidence that the perceptual modality used to interact with objects has an effect on object representations. The fact that we obtained the same pattern of similarity-based weights as in our previous study despite differences in the experimental conditions (e.g., stimuli in the previous visual condition were 2D images presented on a computer monitor, with shorter presentation times; haptic stimuli lay flat on a table instead of upright) indicates that the weight pattern we obtained is robust against these variations. Interestingly, Lakatos and Marks (1999) reported that local shape initially played an important role relative to global shape in haptic similarity ratings of 3D objects but that the importance of local shape decreased when exploration time was increased. Variables such as exploratory procedure, exploration time, and viewpoint need to be systematically manipulated in order to characterize the sensitivity of modality weights to such factors.

Convergence of Stimulus Representations

Positing a single, multimodal stimulus representation with modality-dependent weights provided the same goodness-of-fit to our similarity data as three, modality-specific representations. This suggests that the modalities make use of similar (or even perhaps common) object representations for the purposes of judging similarity. The idea that object information coming from touch and vision converges or at least overlaps in a multimodal object representation agrees with evidence from a number of visuohaptic processing studies using brain imaging techniques (e.g., Amedi, Malach, Hendler, Peled, & Zohary, 2001; Amedi, Jacobson, Hendler, Malach, & Zohary, 2002; James et al., 2002;

Pietrini et al., 2004; Forti & Humphreys, 2005), and psychophysics (e.g., Easton, Greene, & Srinivas, 1997; Easton, Srinivas, & Greene, 1997; Reales & Ballesteros, 1999; Norman, Norman, Clayton, Lianekhammy, & Zielke, 2004). Elucidating the computational principles which govern multimodal integration is an important area of current research (Ernst & Bühlhoff, 2004). Early studies of visuohaptic integration proposed that vision simply dominated touch when both modalities were available (Rock & Victor, 1964). In this study, we did not find evidence for visual capture for similarity and categorization tasks. Instead, our results in the bimodal condition appear to be more compatible with weighted averaging models of multisensory integration. In one such model (Ernst & Banks, 2002), the bimodal estimates of stimulus properties are weighted by the reliability of the unimodal estimates. A variant of our experiment in which the reliability of unimodal estimates is manipulated in the bimodal condition (e.g., by having subjects wear gloves, blurring the visual stimulus, or showing different stimuli in haptic and visual conditions) could be used to test whether this model is capable of predicting integration effects for similarity judgments and category construction.

Connection Between Similarity and Categorization in a Multimodal Setting

In this study, we were able to establish a connection between similarity and categorization: similarity was lower for pairs which subjects placed in different categories and higher for pairs which subjects placed in the same category. In addition, when we made use of the fact that subjects perceived nearest-neighbour adjacencies in the stimulus set (e.g., between two objects which differ only in terms of one step along either the shape or texture dimensions), the probability of crossing such an adjacency with a category boundary decreased as a function of the perceptual similarity between the objects. We also found that the relative weight of shape and texture varied with modality in the same way for our similarity and categorization tasks. One explanation for this could be that

modality-specific biases affect both tasks (or the representations upon which they operate) in a uniform fashion. Modality-specific biases towards features can arise due to a number of factors, including the relative discriminability of features, the relative reliability of feature value estimates, directing of attention, past experience, and ecological validity (Lederman, Summers, & Klatzky, 1996; Guest & Spence, 2003; Ernst & Bühlhoff, 2004). The effects of modality bias in determining the relative weights of features in object representations may co-exist or compete with the effects of top-down category learning (Nosofsky, 1986; Goldstone & Steyvers, 2001; Sigala & Logothetis, 2002).

Despite the connection we found between performance in similarity and categorization tasks, we were not able to demonstrate a strong relationship between the two. We hypothesized that a strong connection between similarity and categorization might enable us to predict subjects' free sorting categories using the clusters of stimuli in their individual similarity-based stimulus spaces. However, this proved to be more difficult than expected. In several cases, subjects categorized the stimuli based on rules which corresponded to unidimensional decision boundaries along shape or texture levels. Although these decision boundaries were compatible with the configurations recovered from the subjects' similarity data in a certain number of cases, they *clearly contradicted* subjects' similarity-based stimulus representations in several other cases. This result was surprising to us given the large amount of evidence that perceptual categorization is intrinsically related to perceptual similarity (Goldstone, 1994; Hahn & Ramscar, 2001).

The tendency for subjects to sort according to unidimensional rules as opposed to similarity could have been an artifact of the free sorting task, as pointed out by one of our reviewers. A number of studies in the cognitive psychology literature (e.g., Imai & Garner, 1965; Medin, Wattenmaker, & Hampson, 1987; Ahn & Medin, 1992), have reported the use of single-feature rules in free sorting tasks. Interestingly, Regehr and Brooks (1995) found that when stimuli were presented all at once (the traditional "array procedure" for

free sorting), subjects used unidimensional rules, but when stimuli were presented sequentially and matched to standards of Category A and B (present at all times), subjects suddenly began sorting according to similarity. A recent study showed that the match-to-standards procedure only led to similarity-based sorting for a perceptually simple stimulus set (a sequence of line drawings of basic geometrical shapes) whose dimensions were spatially separated (Milton & Wills, 2004). For a more perceptually complex stimulus set (schematic butterflies) with spatially co-located features, subjects again resorted to unidimensional rules. These results are consistent with our findings considering that our 3D object stimuli are "perceptually complex" and vary in terms of shape and texture, two features which are spatially co-located. However, it is important to note that our task was neither a match-to-sample nor a classical array procedure; rather, subjects had to construct their categories and then sequentially assign stimuli to them. This may have imposed a significant working memory load which is not present in the other tasks. One study of memory-based category construction found that although sensitivity to similarity relationship was observed in perceptual sorting, subjects preferred to sort according to single dimensions in memory-based tasks (Wattenmaker, 1992). Thus, the memory requirements of our task may have been another factor which encouraged the use of unidimensional rules. A final factor could be that our stimuli only varied along two dimensions; it has been shown that subjects tend to classify using similarity when objects vary simultaneously along many dimensions, but prefer unidimensional rules when objects vary along fewer dimensions (Smith, 1981). Further studies are needed to disentangle the effects of stimulus dimensions and task design on category construction and to determine whether, under certain circumstances, categorization can be predicted from similarity in a multimodal setting.

A New Approach to the Study of Visuohaptic Processing

This study makes use of a novel combination of computer graphics, 3D printing technology, and MDS techniques presented in Cooke et al. (2005). In recent years, studies of visual perception have profited from advances in computer graphics and virtual reality, but studies of haptic perception have been hampered by the lack of adequate haptic presentation devices and the paucity of techniques available to easily create artificial, controlled three-dimensional stimuli. For example, stimuli have been made by by precision-cutting (Klatzky et al., 1987; Lakatos & Marks, 1999), casting (Norman et al., 2004), moulding by hand (James et al., 2002; Forti & Humphreys, 2005), or manually assembling toy bricks (Newell, Ernst, Tjan, & Bühlhoff, 2001; Forti & Humphreys, 2005). The technique used here facilitates the production and reproduction of novel, 3D objects and allows for a high degree of control over object properties. In addition, the combination of parametrically-varying stimuli and MDS techniques allows for intuitive visualizations and quantification of relative differences in feature weights. Another advantage of this approach is that it can be used to generate stimulus maps and dimension weights using any kind of proximity data gathered on parametrically-varying stimuli. For example, maps and dimension weights generated by computational models of the visual system can be tested against those provided by human viewers, as we have demonstrated in Cooke et al. (2005), and the same could conceivably be done to evaluate computational models of the haptic system and/or models of visuohaptic perception. Given its broad potential applicability, the method offers a valuable tool for research in multisensory processing.

Summary and Outlook

This study provides clear evidence that the perceptual modality used to interact with objects affects the representations used for similarity judgments and categorization. The relative importance of shape and texture varied systematically according to modality

for both our similarity and categorization tasks: shape was more important than texture when tasks were performed using vision only, whereas texture and shape were roughly equally important when tasks were performed either haptically or bimodally. We were able to model these differences as a modality-dependent rescaling of a single map, suggesting similar or perhaps even common multimodal representations.

The study also demonstrates a connection between similarity and categorization within a multisensory context. The same basic modality effects on dimension weights were observed for similarity and categorization tasks; in addition, the probability of within-category membership increased with perceptual similarity, while the probability of a category boundary being placed between neighbouring stimuli decreased with similarity. The lack of a stronger connection between similarity space and category structure was discussed in relation to the free sorting task which may have encouraged the use of unidimensional rules; additional studies involving are required to test this hypothesis.

Further work is needed to generalize these results by applying the same methodology to new stimulus sets which vary, for example, in terms of part-whole configuration or in terms of the scales at which macrogeometrical and microgeometrical manipulations are applied. Systematic variation of exploratory procedures and viewpoint will also be important steps towards the goal of understanding the cognitive consequences of multisensory object perception.

References

- Ahn, W., & Medin, D. (1992). A two-stage model of category construction. *Cognitive Science*, 16(1), 81-121.
- Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cereb Cortex*, 12, 1202-1212.
- Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci*, 4(3), 324-330.
- Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling*. Springer.
- Carroll, J., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283-319.
- Cooke, T., Steinke, F., Wallraven, C., & Bülthoff, H. (2005). A similarity-based approach to perceptual feature validation. In *APGV 2005 - Symposium on Applied Perception in Graphics and Visualization*. ACM SIGGRAPH.
- Cox, T., & Cox, M. (2001). *Multidimensional scaling* (2nd ed.). Chapman & Hall.
- Easton, R., Greene, A., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-d patterns and 3-d objects. *Psychon Bull Rev*, 4(3), 403-410.
- Easton, R., Srinivas, K., & Greene, A. (1997). Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. *J Exp Psychol Learn Mem Cogn*, 23(1), 153-163.
- Edelman, S. (1999). *Representation and recognition in vision*. MIT Press.
- Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429-433.

- Ernst, M., & Bühlhoff, H. (2004). Merging the senses into a robust percept. *Trends Cogn Sci*, 8(4), 162–169.
- Forti, S., & Humphreys, G. (2005). Cross-modal visuo-tactile matching in a patient with a semantic disorder. *Neuropsychologia*, 43(11), 1568–1579.
- Garbin, C. (1988). Visual-haptic perceptual nonequivalence for shape information and its impact upon cross-modal performance. *J Exp Psychol Hum Percept Perform*, 14(4), 547-553.
- Garner, W. (1974). *The processing of information and structure*. Maryland: Lawrence Erlbaum Associates.
- Goldstone, R. (1994). The role of similarity in categorization: providing a groundwork. *Cognition*, 123, 125–157.
- Goldstone, R., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *J Exp Psychol Gen*, 130(1), 116-139.
- Guest, S., & Spence, C. (2003). Tactile dominance in speeded discrimination of textures. *Exp Brain Res*, 150(2), 201–207.
- Hahn, U., & Ramscar, M. (Eds.). (2001). *Similarity and categorization*. Oxford University Press.
- Hollins, M., Faldowski, R., Rao, S., & Young, F. (1993). Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Percept Psychophys*, 54(6), 687-705.
- Imai, S., & Garner, W. (1965). Discriminability and preference for attributes in free and constrained classification. *J Exp Psychol*, 69, 596–608.
- James, T., Humphrey, G., Gati, J., Servos, P., Menon, R., & Goodale, M. (2002). Haptic

study of three-dimensional objects activates extrastriate visual areas.

Neuropsychologia, 40, 1706-1714.

Klatzky, R., Lederman, S., & Reed, C. (1987). There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision. *J Exp Psychol Gen*, 116(4), 356-369.

Lakatos, S., & Marks, L. (1999). Haptic form perception: relative salience of local and global features. *Percept Psychophys*, 61(5), 895-908.

Landau, B., & Leyton, M. (1999). Perception, object kind, and object naming. *Spat Cogn Comput*, 1, 1-29.

Lederman, S., & Klatzky, R. (1993). Extracting object properties through haptic exploration. *Acta Psychol*, 84, 29-40.

Lederman, S., Summers, C., & Klatzky, R. (1996). Cognitive salience of haptic object properties: role of modality-encoding bias. *Perception*, 25, 983-998.

Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychol Rev*, 85(3), 207-238.

Medin, D., Wattenmaker, W., & Hampson, S. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognit Psychol*, 19(2), 242-279.

Milton, F., & Wills, A. (2004). The influence of stimulus properties on category construction. *J Exp Psychol Learn Mem Cogn*, 30(2), 407-415.

Morrongiello, B., Humphrey, G., Timney, B., Choi, J., & Rocca, P. (1994). Tactual object exploration and recognition in blind and sighted children. *Perception*, 23(7), 833-848.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychol Rev*, 92(3), 289-316.

- Newell, F., Ernst, M., Tjan, B., & Bühlhoff, H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychol Sci*, *12*(1), 37-42.
- Norman, J., Norman, H., Clayton, A., Lianekhammy, J., & Zielke, G. (2004). The visual and haptic perception of natural object shape. *Percept Psychophys*, *66*(2), 342-351.
- Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen*, *115*(1), 39-57.
- Nosofsky, R. (1992). Similarity scaling and cognitive process models. *Ann Rev Psychol*, *43*, 25-53.
- Pietrini, P., Furey, M., Ricciardi, E., Gobbini, M., Wu, W., Cohen, L., et al. (2004). Beyond sensory images: object-based representations in the human ventral pathway. *Proc Natl Acad Sci U S A*, *101*, 5658-5663.
- Reales, J., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: cross-modal priming depends on structural descriptions. *J Exp Psychol Learn Mem Cogn*, *25*(3), 644-663.
- Regehr, G., & Brooks, L. (1995). Category organization in free classification: the organizing effect of an array of stimuli. *J Exp Psychol Learn Mem Cogn*, *21*(1), 347-363.
- Rock, I., & Victor, J. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science*, *143*, 594-596.
- Rosch, E., & Mervis, C. (1975). Family resemblances: studies in the internal structures of categories. *Cognit Psychol*, *7*, 573-605.
- Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognit Psychol*, *8*, 382-439.

- Shepard, R., & Cermak, G. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognit Psychol*, 4, 351-377.
- Sigala, N., & Logothetis, N. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415, 318-320.
- Smith, L. (1981). Importance of the overall similarity of objects of adults' and children's classifications. *J Exp Psychol Hum Percept Perform*, 7(4), 811-824.
- Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *J Exp Psychol Gen*, 113(2), 169-197.
- Wattenmaker, W. (1992). Relational properties and memory-based category construction. *J Exp Psychol Learn Mem Cogn*, 18(5), 1125-1138.
- Young, F., & Harris, D. (2003). ALSCAL. In *SPSS 12.0 Command syntax reference* (p. 100-116). Chicago, IL: SPSS.

Author Note

The authors wish to thank Fiona Newell and Douglas Cunningham for helpful conversations, Michael Renner and Martin Breidt for technical assistance with stimulus production, as well as Karin Bierig and Bianca Arsene for their role in running the experiments. Informed consent was obtained from all subjects who participated in the experiments.

Footnotes

¹There is no general consensus on the most appropriate general psychological similarity metric for haptically-perceived stimuli. We were aware of the possibility that subjects' psychological metric could be non-Euclidean and tried to fit the similarity data using a city-block metric (Garner, 1974). We did not find a significant decrease in fit error compared to using a Euclidean metric and thus felt that the more intuitive Euclidean approach was preferable, especially given that our stimulus dimensions may not be strictly separable.

Figure Captions

Figure 1. Stimuli: Novel, 3D objects ordered according to shape (macrogeometry) and texture (microgeometry)

Figure 2. Experimental Setup: Top view of subject participating in the visuohaptic exploration condition. She follows the object's contour with her right index finger, repeats this for a second object, and then rates their similarity. The experimenter (shown on the right) records her response.

Figure 3. Perceptual Stimulus Map: Map derived by MDS using similarity data from all modality conditions.

Figure 4. Relative Weight of Shape and Texture in Similarity Judgments. Population mean (left) and individual subject data (right). Error bars represent standard error. The relative weight is calculated as the angle that the weight vector in MDS subject space makes relative to the texture axis. Overlapping individual data points are vertically shifted for better visualization. V = visual, VH = visuohaptic, H = haptic, * = significant difference, n.s. = not significant.

Figure 5. Performance of Free Sorting Categorization. Number of repetitions of free sorting task needed until all stimuli were categorized the same way twice in a row (left) and number of categories created by subjects in each modality condition (right). Error bars represent standard error.

Figure 6. Relative Weight of Shape and Texture in Categorization. Population mean (left) and individual subject data (right). Error bars represent standard error. The relative weight is calculated as the percentage of the total number of category boundaries selected by a subject separated stimuli on the basis of texture differences. Overlapping individual

data points are vertically shifted for better visualization. V = visual, VH = visuohaptic, H = haptic, * = significant difference, n.s. = not significant.

Figure 7. Verbal Mention of Object Properties for Object Descriptions, Similarity Judgments, and Categorization. Shape responses includes references to both global and part shape properties. Two haptic subjects also mentioned temperature when describing the objects. V = visual, VH = visuohaptic, H = haptic.

Figure 8. Verbal Mention of Global Shape (left) and Part Shape (right) for Object Descriptions, Similarity Judgments, and Categorization. V = visual, VH = visual-haptic, H = haptic.

Figure 9. Similarity & Categorization. Left: Histogram showing the number of stimulus pairs placed in the same category (grey) and the number of stimulus pairs placed in different categories (white) as a function of pairwise similarity. Right: Mean probability of subjects (all modality conditions, N = 30) placing a category boundary between *adjacent* stimulus pairs as a function of pairwise similarity. Error bars represent standard error.

















