Spatial Vision, Vol. 10, No. 3, pp. 201–236 (1996)
© VSP 1996.

Categorical relations in shape perception

JOHN E. HUMMEL* and BRIAN J. STANKIEWICZ

Department of Psychology, University of California, Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095-1563, USA

Received 8 July 1996; revised 2 August 1996; accepted 14 August 1996

Abstract—Many researchers have proposed that objects are perceived as *structural descriptions*, which specify the configuration of an object's features (or parts) in terms of their categorical relations to one another. Others have proposed that objects are perceived as *views*, which specify the configuration of an object's features in terms of their coordinates, in particular 2D views. This paper presents five experiments testing these competing accounts of the perception of the configuration of an object's features. Subjects learned to recognize a set of target objects and were tested for their ability to distinguish them from various distractors that differed either in their categorical relations or their coordinates. Subjects were consistently more likely to confuse both 2D and 3D objects that were similar in their parts' relations to each other than to confuse objects similar in their parts' coordinates (in any reference frame). This effect persisted when subjects were allowed to view the objects as long as they wished and when they were explicitly trained to distinguish them from the distractors. These findings suggest that we perceive an object's features in terms of their categorical relations to one another. A preliminary model of the findings is presented.

1. RELATIONS IN SHAPE PERCEPTION

A substantial body of work in both psychology and computer science is addressed to understanding the representations underlying human shape perception and object recognition. A representation of object shape is characterized by at least three independent attributes (see Palmer, 1978; Tarr, 1995): a reference frame (e.g. viewer- or object-centered); a collection of primitive elements or 'features' (e.g. Gabor wavelets, lines and vertices, volumetric parts, etc.); and a set of relations for specifying the primitives' configuration inside the reference frame (e.g. relative to one another or relative to the origin of the reference frame). The reference frames and primitives serving human shape perception have been the subject of a substantial body of research (see Biederman, 1987; Biederman and Cooper, 1991; Quinlan, 1991; Tarr,

*Correspondence should be addressed to John Hummel at the University of California, Los Angeles, Department of Psychology, Franz Hall, 405 Hilgard Ave., Los Angeles, CA 90095-1563. E-mail: jhummel@psych.ucla.edu.

1995), but very little is known about how we perceive the spatial relations among an object's features or parts. Understanding relations is critical to understanding shape perception in general: The same primitives, specified in the same reference frame, can give rise to radically different behaviors depending on the specification of their configuration (Palmer, 1978; Hummel and Biederman, 1992; Hummel, 1994, 1995).

One proposal that has been influential in theories of human shape perception is that objects are represented as *structural descriptions* (Clowes, 1967; Sutherland, 1969; Marr and Nishihara, 1978; Rock, 1983; Biederman, 1987; Hummel and Biederman, 1992). According to this idea, objects are perceived in terms of their features' or parts' locations relative to one another. For example, a mug might be represented as a curved cylinder (the handle) side-attached to a vertical cylinder (the body) (Biederman, 1987); a human body might be represented as a hierarchical arrangement of cylinders attached at specific angles (Marr and Nishihara, 1978; Marr, 1982). Different structural descriptions theories differ in the details of the parts and reference frames they assume (e.g. the structural descriptions proposed by Marr and Nishihara are specified in a much more object-centered format than those proposed by Hummel and Biederman), but all these theories share the assumption that objects are explicitly represented in terms of their features or parts' relations to one another.

One notable property of such a description—especially as contrasted with a literal 'template' (Neisser, 1967)—is its capacity to code relations categorically. Categorical relations permit object recognition in novel viewpoints and recognition of novel instances of known classes as a natural consequence (Biederman, 1987). For example, the description 'curved cylinder side-attached to a vertical cylinder' applies to many different mugs as they appear in many different viewpoints. The properties of categorical relations make structural descriptions intuitively appealing, and a substantial body of evidence has been interpreted as support for structural descriptions in human shape perception (for reviews, see Pinker, 1984; Quinlan, 1991). However, the vast majority of this evidence speaks to the role of volumetric parts in shape perception (e.g. Biederman, 1987; Biederman and Cooper, 1991), or to the reference frames used for shape perception and recognition (see Quinlan, 1991). Comparatively little is known about how we perceive the spatial relations among an object's features or parts.

There is some evidence for the role of categorical relations in the perception of simple forms, such as pairs of lines. For example, Foster and Ferraro (1989) measured discrimination performance for pairs of displaced lines. The displacement created a gap between the lines in a pair, and pairs differed in the size of this gap. In their Experiment 1, subjects viewed displays containing three such pairs, two of which were identical and one of which was different, and their task was to determine which pair was different. At short exposure durations (100 ms), a categorical effect was evident in subjects' judgments: Discrimination accuracy was sharply peaked for line pairs with no gap vs. line pairs with a small gap relative to line pairs with a small gap vs. line pairs with a slightly larger gap. (At longer exposure durations, performance fell off more smoothly with gap size.) Subsequent experiments revealed categorical effects over other perceptual boundaries (e.g. 'just a gap' vs. 'more than just a gap'; see Foster and Ferraro, 1989).

Findings such as these are suggestive in the context of the claim that shape perception is based (at least partly) on the categorical relations among an object's features or parts. However, in these and related experiments (e.g. De Valois *et al.*, 1990), subjects were explicitly instructed to respond to the relations between the parts of the stimuli. It is not clear what role such relations play in situations where people are not explicitly directed to attend to them. (As elaborated shortly, many recent theories of shape perception posit no role whatsoever for the relations among an object's features, except inasmuch as those relations are implicit in the features' literal coordinates in the image.) In addition, these experiments used stimuli composed of very simple features in very simple relations. As such, it is difficult to generalize from findings such as these to the kinds of spatial relations that figure prominently in structural description theories of shape perception.

This paper presents five experiments investigating the role of categorical relations in the perception of relatively complex, object-like shapes (Fig. 1). The question

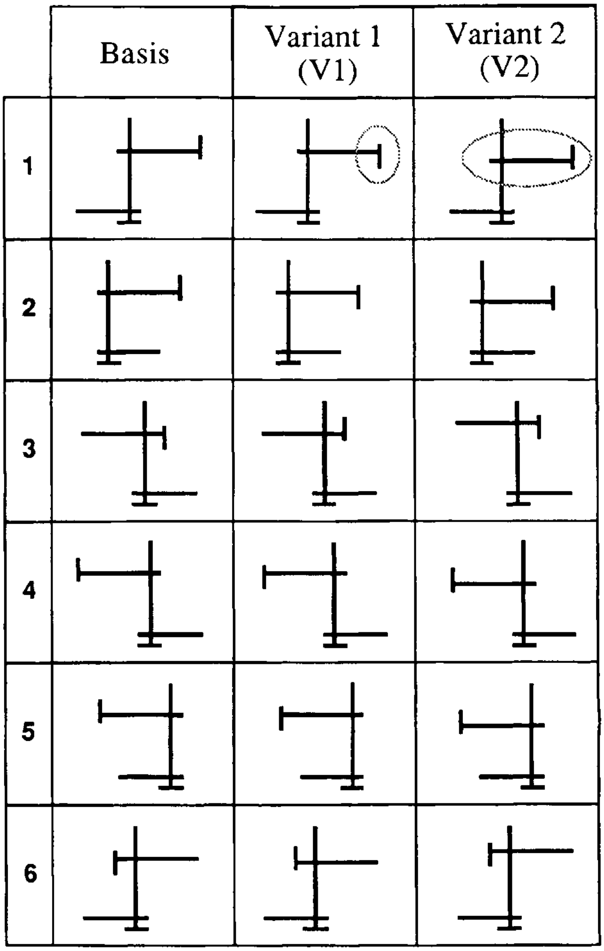

Figure 1. The stimuli (Basis objects and their variants) used in Experiments 1–3.

of whether such relations play a role in shape perception is especially timely in the context of current models of shape perception and object recognition. Although structural description theories were the dominant account of shape perception for several years, 'view-based' models have recently enjoyed growing popularity (see Bülthoff *et al.*, 1995, for a review). These models represent objects in terms of the *coordinates* of their features as they appear in particular 2D views. That is, rather than representing an object's features relative to one another, a view represents them relative to the origin of a coordinate system. These models recognize objects in novel viewpoints by means of vector operations (such as linear combination, view interpolation) on the resulting coordinate vectors (e.g. Poggio and Edelman, 1990; Ullman and Basri, 1991; Vetter *et al.*, 1994). This approach accounts for some effects of viewpoint in human shape perception, at least for some types of objects (see Bülthoff *et al.*, 1995; Tarr, 1995; but see Biederman and Gerhardstein, 1995). If this account of shape perception is correct, then the visual system has no need for categorical relations of the type proposed in structural description theories. Indeed, such relations would be detrimental to performance to the extent that they interfere with the operations that match viewed images to familiar views stored in memory.

2. TESTING SENSITIVITY TO CATEGORICAL RELATIONS IN SHAPE PERCEPTION

If shape perception is sensitive to categorical relations, then two shapes composed of the same features should be easy to discriminate to the extent that they differ in the categorical relations among their features. However, stimuli that differ in the categorical relations among their features necessarily also differ in the coordinates of their features. Therefore, to measure the perceptual effects of categorical relations, it is necessary to control for the perceptual effects of feature coordinates. One way to do so is to compare subjects' sensitivity to the difference between stimuli that differ in both their categorical relations and their coordinates, to their sensitivity to stimuli that differ only in their coordinates. In the experiments reported here, coordinate and categorical differences were always anti-correlated, such that stimuli that differed in their categorical relations were always more similar in terms of their numerical coordinates than stimuli that were similar in their categorical relations.

Coordinates (e.g. as postulated in modern view-based models) differ from the spatial relations postulated in structural description theories in two critical ways. First, by definition, coordinates are specified relative to a single reference point (the origin of the coordinate system), whereas the relations postulated in structural description theories are defined on multiple reference points (e.g. if every part is explicitly related to every other part, then each part serves as a reference point). Second, the coordinates postulated in view-based models vary linearly with the location of the feature in the reference frame: If moving a feature vertically distance d changes its vertical coordinate by amount v, then moving the feature distance $2d$ will change its vertical coordinate by $2v$. (This assumption is important in the context of the vector operations on which view-based models are based; see Ullman and Basri, 1991; Bülthoff *et al.*, 1995.) By contrast, categorical relations do not vary linearly with a feature's location in the image. Rather, they are unaffected (or minimally affected) by changes within categorical boundaries (e.g. from 'above' to 'farther above'), and maximally affected by otherwise equivalent changes across boundaries (e.g. from 'above' to 'below'). Investigating perceptual sensitivity to categorical relations is therefore a matter of investigating (a) whether the configuration of an object's parts is perceived relative to multiple reference points, and (b) whether the part's location relative to any given reference point is specified as a linear or nonlinear function of the part's location in the reference frame as a whole (e.g. the image).

Consider the *Basis* objects in Fig. 1 (column 1), and their V1 and V2 variants (columns 2 and 3, respectively). Each V1 variant was created by moving one part (*part 1*) of the corresponding Basis object six pixels up or down. (The parts that move to create the V1 and V2 variants of Basis object 1 are circled with light gray ellipses in Fig. 1. These ellipses did not appear in the stimuli used in the experiments.) The move was chosen to change the categorical above/below relation between part 1 and the part to which it is attached (*part 2*). Each V2 variant was created by moving both part 1 and part 2 six pixels. No categorical relations change as a result of this move. As such, each basis object has the same pair-wise categorical relations as its V2 variant, but the coordinate difference between a Basis–V2 pair is greater than the coordinate difference between the corresponding Basis–V1 pair (as elaborated shortly). The difference between subjects' ability to discriminate the Basis objects from their categorically different (but metrically similar) V1 variants and their ability to discriminate the Basis objects from their metrically different (but categorically similar) V2 variants thus serves as a conservative index of the role of categorical relations in shape perception.

2.1. Isolating the perceptual effects of categorical relations

To use discrimination performance with Basis–variant pairs as a basis for measuring perceptual sensitivity to coordinates and categorical relations, it is necessary to isolate perceptual effects due to shared coordinates and relations from effects due to similar viewpoints or shared features. To this end, the stimuli were presented at a constant size and viewpoint (except in Experiment 4). To eliminate effects associated with shared features, it is necessary to use stimuli that differ only in the locations—not the identities—of their features. We must therefore know what to regard as a 'feature'. All the stimuli were composed of horizontal and vertical lines (except in Experiment 4, where the figures are composed of volumes; Fig. 3). The analyses presented below are based on the assumption that either the lines or their endpoints serve as the basic features whose coordinates (or relations) are perceptually coded. Importantly, this assumption is consistent with virtually all extant coordinate-based models, which use line endpoints or curvature extrema as primitives (Poggio and Edelman, 1990; Edelman and Weinshall, 1991; Ullman and Basri, 1991; Siebert and Waxman, 1992; Vetter *et al.*, 1994; Edelman *et al.*, 1996). Importantly, the coordinate-based similarity relations among our stimuli are exactly the same regardless of whether the lines themselves (e.g. their midpoints) or the lines' endpoints are perceptually primitive. The analysis also holds if simpler features (e.g. pixels) are assumed to be primitive.

Although it is possible to invent a vocabulary of primitives in which our stimulus manipulations correspond to changes in the identity rather than the location of a feature, we are aware of no models based on such a vocabulary, and it is difficult to imagine any general theory of shape perception that would support such an account. As such, we take the analysis to be very general.

2.2. *Quantifying coordinate-based discriminability*

It is now possible to quantify the discriminability of the Basis and variant objects in Fig. 1 in terms of the coordinates of their features. The coordinates of an object's features define a vector $\boldsymbol{c}$ in a DN vector space, where D is the dimensionality of the reference frame in which the coordinates are defined (e.g. 2 for viewer-centered coordinates and 3 for object-centered coordinates), and N is the number of features in each object (Poggio and Edelman, 1990; Poggio and Girosi, 1990). For example, using line endpoints as primitives, the retinotopic coordinates ($D = 2$) of the features of the objects in Fig. 1 ($N = 10$) can be represented as a 20-dimensional vector $[x_1, y_1, x_2, y_2, \ldots, x_{10}, y_{10}]$. The coordinate similarity of two objects, i and j, is given by the similarity of their coordinate vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$, and the discriminability of i and j is proportional to the Euclidean distance between $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ in the DN space. This is true regardless of whether the reference frame is 2D or 3D, Cartesian or polar, or viewer- or object-centered. For convenience and simplicity, we will illustrate the implications of this property using 2D Cartesian coordinates, but it is important to recognize that the analysis is not limited to such spaces.

Recall that each V1 variant was generated from the Basis object in its row by moving one line (part 1) six pixels; each V2 variant was generated by moving two lines (parts 1 and 2) the same direction and distance. In terms of their parts' coordinates, each Basis object is more similar to its V1 variant (which differs in the position of one line) than to its V2 variant (which differs in the position of two). Let the six-pixel shift in a part's location correspond to a distance of 1.0 (in the 2D coordinate space). If coordinates are defined on line midpoints, then the Euclidean distance between each Basis object and its V1 variant (in the DN vector space) is exactly 1.0; the distance between each Basis object and its V2 variant is 1.414 (the square root of 2.0).[1] If coordinates are defined on line endpoints, then the distance between the Basis objects and their V1 variants is 1.414 (because two endpoints differ in their locations, each by 1.0), and the distance between the Basis objects and their V2 variants is 2.0 (because four endpoints differ). Finally, if coordinates are defined on both midpoints and endpoints, then the distance between the Basis objects and their V1 variants is 1.732, and the distance between the Basis objects and their V2 variants is 2.449. Thus, regardless of whether line endpoints, midpoints, or both are taken as features, the Basis–V1 pairs are more similar (less discriminable) than Basis–V2 pairs in terms of their features' coordinates. A coordinate-based measure of match therefore predicts that Basis–V1 pairs will be perceptually more confusable (e.g. in a same-different task) than Basis–V2 pairs.

It is important to emphasize the generality of this prediction. It applies to any coordinate-based measure of match that decreases monotonically with distance in the

DN space of coordinate vectors. For example, it is impossible to change the prediction by differentially weighting the objects' features (e.g. as proposed by Edelman and Poggio, 1991). Part 1 (which moves to make a V1 variant from a Basis object) also moves in the corresponding V2 variant. Therefore, Basis–V1 similarity is greater than or equal to the Basis–V2 similarity for all sets of positive feature weights. Of course, it is possible to make the Basis–V2 coordinate similarity greater than the Basis–V1 coordinate similarity by assigning a negative weight to part 2 (the horizontal that moves in the V2 variants but not in the V1 variants). With this weighting, the Basis–V1 similarity would be reduced (relative to Basis–V2 similarity) by virtue of the fact that part 2 is in the same location in the Basis object and its V1 variant. However, this weighting would also cause the similarity between a Basis object and *itself* to be lower than the similarity between that Basis object and any other object in which part 2 is in a different location. Any plausible coordinate-based measure of similarity predicts that the Basis objects will be more difficult to discriminate from their V1 variants than their V2 variants. As detailed in the General Discussion, all the view-based models of which we are aware make the same prediction. Any trend in the opposite direction suggests that subjects perceive these figures in terms of the categorical relations among their parts.

3. EXPERIMENTS

We ran a total of five experiments to test the perceptual similarity of objects like those in Fig. 1. In each experiment, subjects were trained to name three *target* objects and then tested for their ability to discriminate them from a variety of distractors. In Experiment 1, targets were chosen from the set of Basis objects in Fig. 1, and distractors were chosen from their V1 and V2 variants. The results of Experiment 1 suggest that the Basis objects are perceptually more similar to their V2 variants than to their V1 variants. Indeed, this experiment revealed virtually no evidence for any perceptual sensitivity to the coordinates of the object's features. Experiment 2 tested whether the findings of Experiment 1 would replicate when subjects were given unlimited time to view the objects. Experiment 3 tested whether the findings would replicate when subjects were explicitly trained to distinguish the Basis objects from their variants. Experiment 4 replicated Experiment 1 with 3D objects, and Experiment 5 replicated Experiment 1 with a new set of 2D objects to test an alternative account of the effects observed in Experiments 1–4. All five experiments showed that subjects are much more sensitive to the categorical relations among the objects' parts than to the coordinates of those parts.

3.1. General method

The experiments used similarity judgments (Experiments 1 and 5), identity judgments (Experiments 2 and 4), sequential same-different judgments (all experiments), and naming recognition (all experiments) to assess the perceptual similarity of the objects in Fig. 1 (Experiments 1–3), and similar objects (Figs 2 and 3; Experiments 4 and 5,

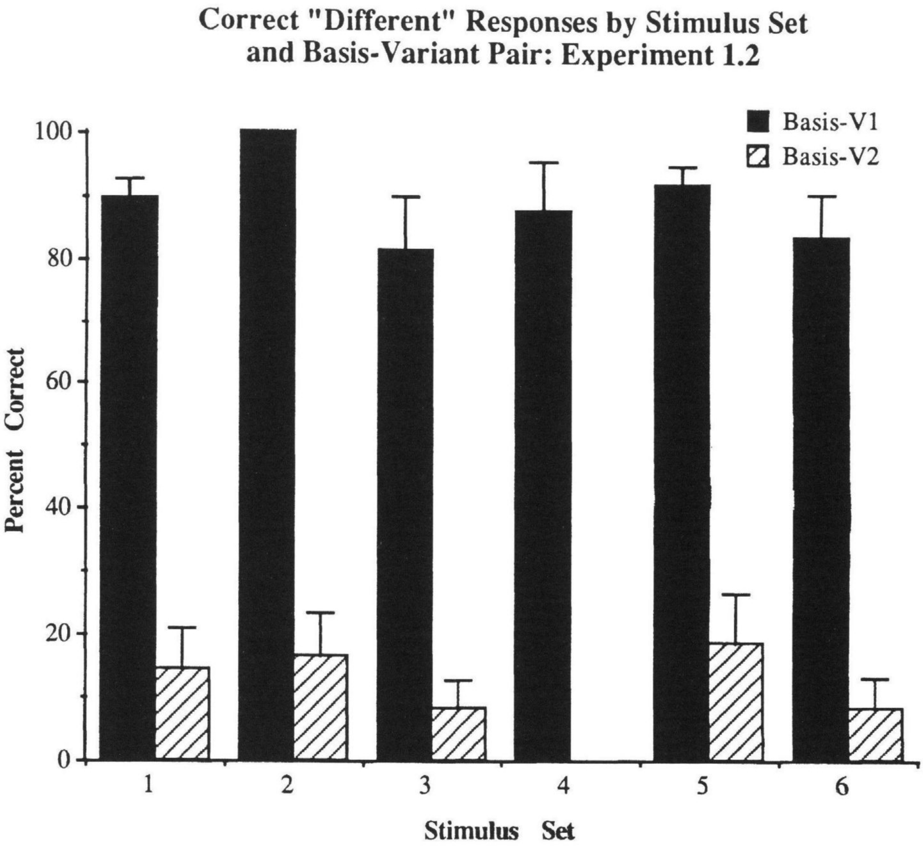

Figure 2. Proportion of correct 'different' responses to Basis–variant pairs in Experiment 1.2.

respectively). Similarity judgments provide a subjective measure of how similar the objects appear, and identity judgments (i.e. 'Which of these stimuli are identical?'), naming recognition, and same-different judgments provide objective measures of the objects' perceptual similarity. The same-different task is particularly appropriate for testing subjects' sensitivity to small differences between objects; for inter-stimulus intervals (ISIs) under about two seconds (Ellis and Allport, 1986), same-different judgments are more sensitive to fine metric differences than is naming recognition (Besner and Coltheart, 1975; Bundesen and Larsen, 1975; Howard and Kerst, 1978; Bundesen *et al.*, 1981; Larsen, 1985; Jolicoeur and Besner, 1987; see Cooper *et al.*, 1992). We used naming recognition to test whether the same effects obtain under recognition as under explicit visual judgment. Subjects always performed all three tasks in any experiment (except Experiment 3), and they always performed them in the same order (similarity or identity, then same-different, then naming recognition). Order was held constant so that subjects would have the most experience with the stimuli during the most difficult task (naming recognition), and so that they would have the opportunity to view all the distractors before performing the same-different and naming recognition tasks.

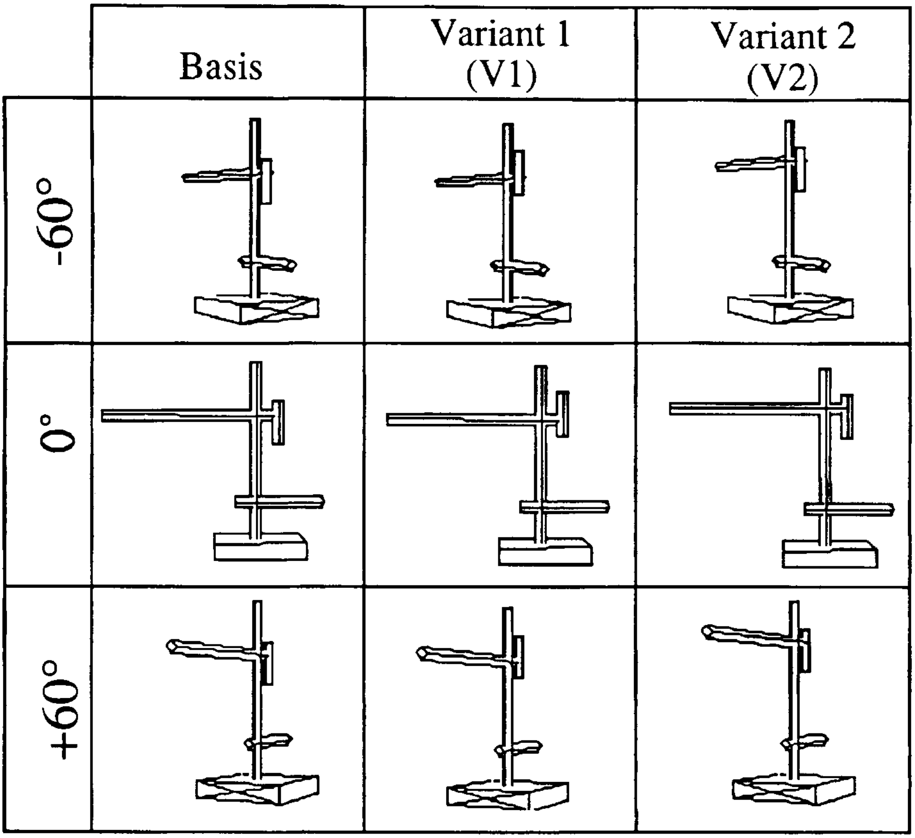

Figure 3. A subset of the 3D stimuli used in Experiment 4.

3.2. Subjects

Undergraduate students at the University of California, Los Angeles, served as subjects in all the experiments. All had normal or corrected-to-normal vision and participated voluntarily for credit in undergraduate psychology courses at UCLA.

3.3. Stimuli

The 18 objects in Fig. 1 served as stimuli in Experiments 1–3. In Experiments 1 and 2, each subject was trained to recognize three Basis objects by name; their V1 and V2 variants served as distractors during the post training tasks. In Experiment 3, each subject was trained to recognize one Basis object and its two variants. In all experiments, stimuli were presented in black on a white background. Each object was drawn with lines two pixels wide and was approximately 2 cm high and 2 cm wide. Subjects sat in a darkened room approximately 90 cm from the display (the objects subtended approximately 1.3 deg of visual angle). Subjects were not restrained with a chin rest or any similar apparatus.

3.4. Apparatus

Stimuli were displayed on a high-resolution color monitor controlled by a Macintosh IIfx computer using the MacProbe computer program (Aristometrics, CA). Responses were gathered via response-box and voice-key using a computer interface box.

3.5. Training

The same basic training procedure was used in all experiments. Each subject was trained to recognize three objects by name. Objects were grouped for counterbalancing into sets of three. In Experiments 1 and 2, set A consisted of Basis objects 1, 2, and 3; set B consisted of Basis objects 3, 4, and 5; and set C consisted of Basis objects 5, 6, and 1. In Experiment 3, each subject's training set consisted of one Basis object and its variants. Here, there was one set for each Basis object. We will refer to the objects in a subject's training set as his or her *target* objects.

Training proceeded in four phases. The first introduced the subject to his or her target objects by displaying them on the screen along with their names ('kip', 'kef', and 'kor', after Tarr and Pinker, 1989). Subjects were allowed to view this display as long as they wished and were instructed that they would later be tested for their ability to recognize and name the objects. In the second phase, the target objects appeared on the screen one at a time and the subject selected the object's name using the mouse. Each trial began with one randomly selected object in the center of the screen and the names 'kip', 'kef', and 'kor' above it. Subjects were allowed to take as long as they needed to identify the object. Following the response, the computer displayed 'Correct' or 'Incorrect', the correct response was '[name]' at the bottom of the screen. The 'correct' message was displayed for one second and the 'incorrect' message was displayed for three seconds paired with a beep lasting 150 ms. This phase of training ended after the subject made 15 correct responses in a row.

The third phase was the same as the second except that Basis objects not in the training set also appeared on the screen. One object was displayed in the center of the screen with the three names and the word 'none' above it. Subjects selected the name of the object if it was in their training set and selected 'none' if it was not. This phase ended after the subject made 15 correct responses in a row.

To discourage subjects from adopting a strategy in which they simply classified the objects on the basis of diagnostic features, the fourth phase of training required them to draw the target objects from memory. Each trial began by displaying a 46×46 mm^2 square on the screen along with instructions telling the subject which object to draw (e.g. 'Draw kip'). Using the mouse, the subject drew the object inside the square. The computer only permitted the subject to draw straight lines, but a subject could draw as many lines as he or she wished. When the subject ended the trial (by double-clicking the mouse button), the corresponding target object was displayed above the drawing. The subject was allowed to compare the drawing to the object, and indicated they were finished by pressing a response-box button. Except for giving the subjects the opportunity to compare their drawings to the actual objects, we did not evaluate the accuracy of their drawings. Subjects drew each object three times in random order.

4. EXPERIMENT 1

Experiment 1 used similarity judgments, speeded same-different judgments, and naming recognition to test whether subjects perceive the V1 or V2 variants (Fig. 1) as more similar to the Basis objects.

4.1. Method

4.1.1. Subjects. Twelve subjects participated in this experiment.

4.1.2. Stimuli. The objects in Fig. 1 served as stimuli.

4.2. Experiment 1.1: Similarity judgment

4.2.1. Design and procedure. On each trial, one target was displayed in the center of the screen with its V1 and V2 variants 1.8 cm apart in a row 6.5 cm above it. Subjects used the mouse to indicate which variant looked most like the target. They were allowed to take as much time as they wished to make each judgment. After making the judgment, the subjects were tested for their ability to recognize the target. The variants were removed from the screen and the names 'kip', 'kef' and 'kor' appeared in a row above the target. The subject's task was to select the target's name with the mouse. The assignment of names to positions in the row was randomized. Trials were run in two blocks of nine. Each target appeared three times (in a random order) in each block.

4.2.2. Results and discussion. Similarity judgments followed by incorrect naming responses (4.2%) were omitted from the analysis. Subjects chose the V2 variant as more similar to the target 75.5% of the time and the V1 variant 24.5% of the time. A two-tailed, matched-pairs t-test revealed that this difference is statistically reliable ($t(11) = 2.40$, $p < 0.05$). Subjects' judgments about Basis–variant similarity were in the direction predicted by the objects' categorical relations.

4.3. Experiment 1.2

Experiment 1.2 used speeded same-different judgments to test subjects' ability to discriminate the Basis objects from their variants. On each trial, a subject viewed two objects in rapid succession and indicated whether they were identical or differed in any way. If subjects perceive the objects in terms of their parts' pairwise categorical relations, then they should more often say 'same' in response to a target paired with its V2 variant than to a target paired with its V1 variant, even though the latter are objectively more similar in terms of their parts' coordinates.

4.3.1. Design and procedure. Subjects were told they would see two objects briefly and that their task was to judge whether they were exactly the same or differed in any way; they were encouraged to respond as quickly as they could without making errors. Following the instructions, subjects viewed a reminder display depicting their target objects paired with their names. They were allowed to view this display as long as they wished.

Each trial began with a fixation cross (2 cm × 2 cm [1.3 deg]) displayed for 1995 ms, followed by a blank screen (495 ms), the first object (240–750 ms as detailed below, and a blank screen for 90 ms), a pattern mask (240 ms; blank 45 ms), the second object (195 ms; blank 90 ms), and a second mask (240 ms). The mask was composed of large number of random straight lines. Subjects responded by pressing buttons marked 'SAME' and 'DIFF' on a response box. They sat with their fingers just above the buttons throughout the experiment.

Trials were run in three blocks of 24. The first was a practice block, in which the objects were displayed for 750 ms on the first trial and the display time was reduced by 30 ms every trial until it reached 240 ms. Display times were 240 ms in both post-practice blocks. The first object to appear on a given trial was placed in one of four locations (randomly chosen) on the screen (3.5 cm above, below, left, or right of the fixation cross); the second object appeared in a location randomly chosen from the three in which the first did not appear. The mask covered all four locations. Half the trials in a block were 'same' trials and half were 'different'. Every 'different' trial contained one target; half paired it with its V1 variant, and half with its V2 variant. On half of the 'different' trials, the target appeared first, and on half it appeared second. Of the 12 'same' trials, six paired a target with itself (two trials for each target), and one paired each variant with itself, so an object's identity as a target or variant was uncorrelated with the correct response ('same' or 'different'). Subjects were not given accuracy feedback during the task. They were required to rest for at least 20 s after each block. The data reported below are from the second and third blocks only.

4.3.2. Results and discussion. Table 1 shows how often subjects said 'same' as a function of which object pair appeared on a trial (e.g. 'Basis–V1' shows the percentage of 'same' responses on V1–Basis and Basis–V1 trials). Subjects were substantially more likely to incorrectly say 'same' to a Basis–V2 pair than to a Basis–V1 pair (87.96% vs. 11.58%, respectively). A two-tailed matched-pairs t-test revealed that this

Table 1.
Means and standard errors of percent 'same' responses by object pair, Experiment 1.2

Object pair	Percent 'same' responses
Basis–Basis	91.67 (3.10)
V1–V1	89.67 (3.98)
V2–V2	92.59 (2.50)
Basis–V1	11.58 (3.02)
Basis–V2	87.96 (3.80)

difference is statistically reliable ($t(11) = 12.64$, $p < 0.01$). Subjects were nearly as likely to say 'same' in response to a Basis object paired with its V2 variant as to a Basis object paired with itself (Table 1, rows 5 and 1, respectively), as predicted by the hypothesis that subjects perceive these shapes in terms of the categorical relations between their parts. As shown in Fig. 2, this trend obtained for all object pairs. Throughout the remainder of this paper, we will present the results for individual object pairs only when they differ from one another.

4.4. Experiment 1.3

In Experiment 1.2, subjects were much more likely to confuse Basis objects with their categorically similar V2 variants than with their metrically similar V1 variants. Experiment 1.3 used speeded naming recognition to test whether similar effects obtain when subjects had to compare a stimulus to a representation in memory. On each trial, one object appeared briefly on the screen. The subject's task was to say its name (e.g. 'kip') if it was in their training set, or to say 'none' if it was not. The dependent measure of interest is the frequency with which subjects incorrectly say the name of a target object in response to its V1 and V2 variants. If subjects represent these objects in terms of the categorical relations among their parts, then they should incorrectly give the name of a target in response to its V2 variant more often than they give its name in response to its V1 variant.

4.4.1. Design and procedure. Prior to the naming task, subjects performed a reminder task identical to the third training phase (object identification with distractors chosen from the set of untrained Basis objects). Next, they were told that objects would appear on the screen one at a time, and their task was to say the object's name if it had appeared in their training set, or to say 'none' if it had not. The instructions emphasized that unless the object was *identical* to one of the objects in their training set, the correct response was 'none'. The instructions were followed by a display showing the target objects along with their names. Subjects were allowed to examine the display for as long as they wished.

Each trial began with a fixation cross (2 cm; displayed for 1995 ms in the center of the screen), followed by blank screen (495 ms), an object in the center of the screen (150–750 ms; see below), a blank screen (90 ms), and a pattern mask (240 ms). The experimenter remained in the room during the entire session and recorded the subject's responses via the keyboard. The experimenter could not see the display.

Trials were run in five blocks of 12. Within a block, half the trials presented distractor objects (chosen from the targets' V1 and V2 variants) and the other half presented targets (Basis objects on which the subject had been trained). Each target appeared twice per block (for six trials), and each distractor appeared once per block (for six trials). Presentation order was randomized. The first block was a practice block, during which the display times started at 750 ms and decreased by 45 ms every trial until they reached 150 ms. Objects were displayed for 150 ms in blocks two through five. The data reported below are from blocks two through five only.

Table 2.
Means and standard errors of percent responses by object, Experiment 1.3

	Object		
Response	Basis	V1	V2
Basis name	92.01 (3.14)	6.25 (1.81)	86.11 (3.45)
'None'	3.125 (0.75)	88.20 (4.52)	11.11 (3.30)
Other	4.86 (2.71)	5.56 (3.88)	2.78 (1.57)

4.4.2. Results and discussion. Table 2 summarizes subjects' responses as a function of which object appeared on a given trial. For example, the row *Basis Name* shows how often a subject said the name of a target Basis object (e.g. 'kip') in response to that Basis object (column *Basis*), its V1 variant (column *V1*), and its V2 variant (column *V2*). '*None*' shows how often subjects said 'none' in response to a target (column *Basis*), its V1 variant (column *V1*), and its V2 variant (column *V2*). *Other* shows other errors (e.g. saying 'kip' in response to an image of 'kef', etc.). Of primary interest is subjects' tendency to say the name of a target in response to its V1 and V2 variants (row 1, columns 2 and 3). Subjects mistook V2 variants for targets substantially more often than they mistook V1 variants for targets (86.11% vs. 6.25%, respectively; $t(11) = 24.07$, $p < 0.01$); they very rarely mistook a V1 variant for a target. Like the pattern observed in Experiment 1.2, this pattern is consistent with the hypothesis that subjects perceive these objects (and encode in them memory) in terms of the categorical relations among their parts.

4.5. General discussion, Experiment 1

In terms of their parts' coordinates (in any type of reference frame), the Basis objects are more similar to their V1 variants than to their V2 variants. If subjects represent these objects in terms of their parts' coordinates, then they should both judge the V1–Basis similarity higher than the V2–Basis similarity, and they should mistake V1 variants for Basis objects more often than they mistake V2 variants for Basis objects. Neither of these effects obtained. In Experiment 1.1, subjects judged the target (Basis) objects more similar to their V2 variants than to their V1 variants. In Experiment 1.2, they incorrectly called a target and its V2 variant 'same' much more often than they called a target and its V1 variant 'same'. And in Experiment 1.3, confusions in naming recognition were again much more frequent for V2 variants than for V1 variants. These findings are consistent with the hypothesis that subjects perceive the categorical above/below relation that distinguishes each Basis object from its V1 variant. Interestingly, subjects' performance discriminating the Basis objects from their categorically-similar (but metrically very different) V2 variants was often well below chance. This finding suggests that the role of coordinates in shape perception is at best weak compared to the role of categorical relations.

5. EXPERIMENT 2

One limitation of Experiment 1 is that our failure to find evidence for coordinate-based representations may simply reflect the limits of the subjects' visual acuity. If subjects represent the objects in Fig. 1 in terms of their parts' coordinates but those coordinates have only coarse spatial resolution, then they may have had difficulty perceiving the small differences between the Basis objects and their V2 variants. To address this possibility, Experiment 2.1 used an identity judgment task. One target (Basis) object and three probe objects were simultaneously displayed on the screen and the subject's task was to identify which probe was *identical* to the target. One probe was identical to the target, and the others were its V1 and V2 variants. If the results of Experiment 1 reflect subjects' inability to perceive the differences between the Basis objects and their V2 variants, then they should be near chance choosing the Basis probe over the V2 probe in this task. A second limitation of Experiment 1 is that the rapid displays may have led us to underestimate subjects' sensitivity to the differences between the Basis objects and their V2 variants: Given time to examine the objects, subjects might easily reject the V2 variants as different from the Basis objects. To address this possibility, Experiments 2.2 and 2.3 replicated Experiments 1.2 and 1.3 using subject-controlled display times.

5.1. Method

5.1.1. Subjects. Twelve subjects participated in this experiment.

5.1.2. Stimuli. The stimuli were the same as those in Experiment 1.

5.2. Experiment 2.1

On each trial, the subject viewed a target (trained Basis) object and three probe objects, and the task was to say which probe was identical to the target. If the results of Experiment 1 reflect the limits of visual acuity, then subjects should be near chance distinguishing the Basis probes from the V2 probes. This experiment was designed to address an additional question as well. In Experiment 1.1 subjects tended to choose the V2 variants as most similar to the Basis objects. These judgments may have established a bias that carried forward to Experiments 1.2 and 1.3. In Experiment 2.1, we explicitly instructed subjects to compare the targets to the probes by 'mentally overlaying them' to determine how much contour they have in common. This operation is analogous to the alignments some coordinate-based models use to match images to object memory (e.g. Lowe, 1987; Ullman, 1989; Olshaussen *et al.*, 1993). Our instructing subjects to compare the objects in this manner was intended to bias them to use a coordinate-based measure of match in the identity judgment task.

5.2.1. Design and procedure. Before beginning Experiment 2.1, subjects were trained to name three Basis (target) objects as described previously. They were told that they would see a target and three probes, and that their task was to decide which probe was identical to the target. Subjects were instructed to '...mentally place the [target] over each [probe], and try to determine which [probe] is identical to the [target]'. The computer illustrated the 'mental overlaying' process by aligning line drawings of common objects. The judgment trials began after the subject indicated that he or she understood the instructions. On each trial, three probes were displayed above a single target. The target was always one of the trained Basis objects, and the probes were the Basis object and its two variants. Subjects were allowed to view the display for as long as they wished. They were not given feedback. Once the subject selected a probe, the probes were replaced with the names 'kip', 'kef', and 'kor', and the subject selected the target's name with the mouse. Trials were run in two blocks of nine.

5.2.2. Results and discussion. Trials followed by incorrect names (0.7%) were omitted from the analysis. On the remainder, subjects selected the correct probe 87.84% of the time. They incorrectly chose the V2 probe 10.98% of the time, and incorrectly chose the V1 probe 1.19% of the time. These response rates are all reliably different from chance (33.33%) (Target: $t(11) = 10.00$, $p < 0.01$; V2: $t(11) = 4.83$, $p < 0.01$; V1: $t(11) = 27.01$, $p < 0.01$). We also tested the Basis and V2 probes against a 50% criterion of chance to determine whether subjects had a better than 50/50 chance of choosing the Basis probe over the V2 probe. Using this criterion, the Basis and V2 probe choice rates are still reliably different from chance (Target: $t(11) = 3.44$, $p < 0.01$; V2: $t(11) = 8.43$, $p < 0.01$). Allowed to view the Basis and V2 objects simultaneously, subjects can easily discriminate them, suggesting that the results of Experiment 1 do not simply reflect the limits of human visual acuity. Despite our instructing subjects to make these judgments by comparing the objects' common contour, they were still over nine times as likely to incorrectly choose the V2 probe than the V1 probe (10.98% vs. 1.19%, respectively; $t(11) = 2.45$, $p < 0.05$). Apparently, the 'mental overlaying' instruction had little effect on subjects' judgments.

5.3. Experiment 2.2

Experiment 2.2 was designed to test whether the results of Experiment 1.2 (speeded same-different judgment) would replicate when subjects were given unlimited time to view the objects.

5.3.1. Design and procedure. This experiment was identical to Experiment 1.2 except for the exposure durations. Each trial began with a fixation cross (1995 ms) followed by a blank screen (495 ms). The first object was then displayed in one of four locations, randomly chosen (3.5 cm above, left, right, or below fixation). Subjects viewed this display for as long as they wished and terminated it by pressing a response box button. A mask then appeared (240 ms), followed by a blank screen (45 ms). The second object was then displayed in one of the three positions in which the first did not appear (randomly chosen). Subjects terminated the display by pushing

Table 3.
Means and standard errors of percent 'same' responses by object pair, Experiment 2.2

Object pair	Percent 'same' responses
Basis–Basis	95.37 (1.79)
V1–V1	98.15 (1.85)
V2–V2	94.45 (1.67)
Basis–V1	4.63 (2.25)
Basis–V2	51.85 (11.97)

'same' or 'different' on the response box. Following the response, the second target was replaced by a mask (240 ms).

Trials were run in three blocks of 24, counterbalanced as in Experiment 1.2. The first block was practice, the first ten trials of which contained additional instructions. Specifically, the first display contained the statement 'Press either button when you are finished examining the object'. The second contained 'Press the appropriate response button (i.e. SAME or DIFF)'. Subjects were required to rest for at least 20 s after each block. The data reported below are from the second and third blocks only.

5.3.2. Results and discussion. Subjects' responses are summarized in Table 3. The data of interest are again the incorrect 'same' responses to Basis–V1 and Basis–V2 pairs. Even when they were given unlimited time to view the objects, subjects incorrectly classified the Basis objects and their V2 variants as 'same' 51.85% of the time. This error rate is reliably greater than the 4.63% error rate with the V1 variants ($t(11) = 3.96$, $p < 0.01$), replicating the basic pattern observed in Experiment 1.2. Subjects are apparently insensitive to the coordinates of the objects' features, even given unlimited time to view them.

5.4. Experiment 2.3

Experiment 2.3 was designed to test whether the effects observed in Experiment 1.3 (greater V2 than V1 confusions in naming recognition) would replicate when subjects were allowed to view each object for as long as they wished.

5.4.1. Design and procedure. The method was identical to that of Experiment 1.3 except that subjects controlled the display times. Each object remained on the screen until the subject responded (saying 'kip', 'kef', 'kor', or 'none' into a microphone attached to a voice key), after which the object was replaced by a mask. Trials were run in five blocks of twelve, the first of which was practice. Trials were counterbalanced as in Experiment 1.3. The results presented below are from blocks two through five.

5.4.2. Results and discussion. Table 4 summarizes subjects' responses to the target (Basis) objects and their variants. The data of primary interest are the confusion rates for the V1 and V2 variants (i.e. saying a target's name in response to one of its variants;

Table 4.
Means and standard errors of percent responses by object, Experiment 2.3

	Object		
Response	Basis	V1	V2
Basis name	93.06 (3.71)	11.11 (4.02)	56.25 (13.88)
'None'	3.12 (2.77)	76.39 (8.81)	34.72 (12.93)
Other	3.82 (2.78)	12.50 (8.97)	9.03 (8.30)

row 1, columns 2 and 3). As in Experiment 1.3, subjects were reliably more likely to mistake a V2 variant for a target than to mistake a V1 variant for a target (56.25% vs. 11.11%; $t(11) = 3.04$, $p < 0.05$). Not surprisingly, subjects were less likely to mistake V2 variants for targets in this experiment than in Experiment 1.3 (86.11%). Giving them unlimited time to view the V2 variants improved their performance, but it did not bring their ability to make this discrimination to the same level as their ability to discriminate the targets from the V1 variants. Subjects' tendency to mistake V2 variants for targets in Experiment 1.3 did not wholly reflect the rapid presentations used in that experiment.

5.5. General discussion, Experiment 2

Experiment 2 was designed to explore whether the results of Experiment 1 simply reflect the limits of visual acuity or effects of the rapid displays. In Experiment 2.1, subjects successfully distinguished the Basis objects from their V2 variants under simultaneous presentation, indicating that acuity alone cannot explain subjects' poor performance with the V2 variants in Experiments 1.2 and 1.3. The results of Experiments 2.2 and 2.3 suggest that the results of Experiment 1 do not simply reflect the brief displays. These experiments used subject-controlled display times and replicated the basic findings of Experiments 1.2 and 1.3.

A more serious limitation of Experiments 1 and 2 is that the results may simply reflect our subjects' response criteria. It is possible that subjects represent the target objects in terms of their parts' coordinates but simply did not realize that small changes in a part's coordinates (without accompanying changes in its position relative to another part) constituted a change in the object. This is especially plausible given that the subjects received no feedback during the same-different and naming-recognition tasks. Moreover, subjects were trained only to discriminate the various Basis objects, which differ from one another in the categorical relations among their parts. As such, the training procedure may have biased subjects to attend selectively to such properties. Experiment 3 was designed to address both these possibilities.

6. EXPERIMENT 3

In Experiment 3, subjects were trained to name one Basis object and its V1 and V2 variants. Explicitly training subjects to use different names for the Basis, V1, and V2 objects should make it clear that even subtle differences are sufficient to define two objects as 'different'. If Experiment 3 replicates the effects observed in Experiments 1 and 2, then we can have some confidence that those effects do not wholly reflect the subjects' expectations or biases. Because completing the training indicates that a subject can distinguish the objects given enough time, the same-different and naming tasks used speeded presentations like those of Experiment 1.

6.1. Method

6.1.1 Subjects. Twelve subjects participated in this experiment.

6.1.2. Stimuli. Each subject was trained to name one Basis object and its V1 and V2 variants. Two subjects were randomly assigned to each of the six object sets (corresponding to the six rows in Fig. 1).

6.2. Experiment 3.1

6.2.1. Design and procedure. Experiment 3.1 used a speeded same-different judgment task. The experimental procedure was identical to that of Experiment 1.2 except for the counterbalancing of trials within blocks. Here, the twelve 'different' trials consisted of two presentations of all possible combinations of trained objects in each order. The twelve 'same' trials consisted of four presentations of each trained object paired with itself.

6.2.2. Results and discussion. The results are summarized in Table 5. Subjects incorrectly called the Basis objects and V2 variants 'same' 61.80% of the time, and incorrectly called the Basis objects and their V1 variants 'same' 9.72% of the time ($t(11) = 7.17$, $p < 0.01$). Although Basis–V2 confusions were less frequent in this experiment than in Experiment 1.2 (87.96%), performance was not reliably better than chance ($t(11) = 1.66$, $0.2 > p > 0.1$). This finding suggests that subjects' poor Basis–V2 discrimination performance in Experiments 1.2 and 2.2 is attributable neither to confusion about what constitutes a 'different' object nor to response biases established during training. Rather, it likely reflects subjects' inability to distinguish the Basis and V2 objects, which differ only (although substantially) in terms of their

Table 5.
Means and standard errors of percent 'same' responses by object pair, Experiment 3.1

Object pair	Percent 'same' responses
Basis–Basis	90.28 (3.05)
V1–V1	94.44 (1.87)
V2–V2	87.50 (4.41)
Basis–V1	9.72 (3.05)
Basis–V2	61.80 (7.71)

parts' coordinates. Basis–V1 discrimination performance did not improve relative to performance in Experiment 1.2 (11.16%).

6.3. Experiment 3.2

6.3.1. Design and procedure. Experiment 3.2 used the same basic procedure as Experiment 1.3 (naming recognition with speeded displays). Except for the difference in the subjects' training sets, the only modification was that the distractors in this experiment were chosen from the five non-target Basis objects. Subjects were instructed to identify objects that were not in their training set as 'none'. Trials were run in five blocks of twelve. Trials within a block were counterbalanced as in Experiment 1.3 except for the 'none' trials. Here, each non-target Basis object was used as a distractor once, except for one (randomly chosen) that was used twice. Presentation order was randomized within blocks. The dependent variable of primary interest is subjects' tendency to confuse the V1, V2, and Basis objects on which they had been trained (i.e. giving the name of one trained object in response to the image of a different trained object), rather than their tendency to confuse trained objects with untrained objects.

6.3.2. Results and discussion. Subjects correctly named the Basis objects 81.25% of the time (standard error = 5.44%), the V1 variants 98.96% of the time (1.04%), and the V2 variants 78.13% of the time (6.73%). They confused the trained Basis objects with their V2 variants (giving the name of one in response to the other) 16.15% of the time, and with their V1 variants 1.56% of the time ($t(11) = 3.92$, $p < 0.01$). Subjects confused V1 variants with V2 variants 2.08% of the time. Performance distinguishing the trained Basis objects from their V2 variants was substantially better than performance on the same discrimination in Experiment 1.3 (86.11% errors), but subjects were still more than five times as likely to mistake a trained Basis object for its V2 variant than to mistake it for its V1 variant. Importantly, this pattern obtained even though subjects were explicitly trained to distinguish these objects.

6.4. General discussion, Experiment 3

The training phase of Experiment 3 required subjects to learn to distinguish a Basis object from its V1 and V2 variants, but subjects still had substantial difficulty distinguishing Basis objects from their categorically similar V2 variants. Like the results of the previous experiments, these findings suggest that the pair-wise categorical relations between an object's parts are perceptually much more salient than the parts' coordinates. They also suggest that the effects observed in Experiments 1 and 2 cannot wholly reflect response criteria, strategies, or biases developed during training. Moreover, there was no similarity judgment task in this experiment, but subjects still made many more Basis–V2 confusions than Basis–V1 confusions in both the same-different and naming tasks. This indicates that performance on the same-different and naming recognition tasks of Experiments 1 and 2 cannot be explained in terms of biases developed during the similarity- and identity-judgment tasks.

7. EXPERIMENT 4

Experiments 1–3 were designed to explore how subjects represent the *configuration* of an object's parts, but it is possible that some of our stimulus manipulations affected the *identities* of those parts. For example, subjects may perceive the shift in the position of the short vertical line from the Basis objects to their V1 variants as a change in the identity of a two-line feature, e.g. from a 'downward-pointing asymmetrical T vertex' to an 'upward-pointing asymmetrical T vertex'. Such an interpretation of the stimuli could explain why subjects made more V2–Basis than V1–Basis confusions in Experiments 1–3. This interpretation is especially plausible given that the objects used in these experiments are composed of simple lines. Although no extant theory predicts that 'downward-pointing asymmetrical Ts' and 'upward-pointing asymmetrical Ts' are different 'features' for shape perception, numerous theories assume that features are defined by specific configurations of line segments (or edges) (e.g. McClelland and Rumelhart, 1981; Biederman, 1987; Hummel and Biederman, 1992; Dickenson *et al.*, 1993; among others). As such, it is not difficult to imagine that our subjects treated these simple configurations of lines as different features. Experiments 4 and 5 were designed in part to address this issue.

Experiment 4 was essentially the same as Experiment 1 except that the stimuli were line drawings of 3D objects composed of convex, volumetric parts (Fig. 3). Although numerous theories of shape perception posit primitive features defined by collections of lines, none posit features defined by collections of volumes. Among other limitations, such a theory would have to posit an enormous number of primitive features for recognition (Biederman, 1987). There are also data suggesting that we perceptually segment objects into separate, convex parts (e.g. at pairs of matched concavities in the image) rather than integrating multiple convex parts into single 'features' (Hoffman and Richards, 1985; Biederman, 1987; Biederman and Cooper, 1991; Palmer and Rock, 1994). As such, one way to increase our confidence that our results speak to the representation of the *configuration* of an object's parts—rather than the identities of those parts—is to see whether they replicate with images of 3D objects composed of convex volumes that meet at matched concavities. 3D objects are also more natural than 2D objects, so it is important to know whether the observed effects generalize to them. The stimuli used in Experiment 4 were designed according to exactly the same criteria as the 2D stimuli of Experiments 1–3, except that the lines were replaced by volumes and the objects were displayed at multiple orientations in depth (Fig. 3).

7.1. Method

7.1.1. Subjects. 18 subjects participated in this experiment.

7.1.2. Stimuli. The stimuli were line drawings of 3D objects (a subset are depicted in Fig. 3) adapted from the 2D objects in Fig. 1 by replacing each line segment with an elongated rod (except for the base, which was replaced by a slab). They were created with Professional Swivel 3D (Paracomp, San Francisco, CA). There were six

Basis objects, each with one V1 and one V2 variant. Each object was displayed at three orientations in depth separated by 60 deg about the vertical axis.

7.1.3. Training. Training was like that of Experiments 1–3 except that the objects were presented at multiple orientations in depth. In phase 1, the three Basis objects in a subject's training set (target objects) were displayed only in the 0 deg orientation. On each trial in phase 2, one target was displayed at a randomly chosen orientation (0, +60, or −60 deg) and the subject's task was to identify it as 'kip', 'kef', or 'kor'. Subjects were instructed to ignore an object's orientation. In phase 3, both target and non-target Basis objects were displayed at randomly chosen orientations in the center of the screen and subjects were instructed to identify each as 'kip', 'kef', 'kor' or 'none'. They were instructed to ignore orientation and to respond 'none' only when the 3D shape of an object did not exactly match any of the objects in the training set. In the final phase, subjects drew one stick-figure version of each target object at the 0-deg orientation.

7.2. Experiment 4.1

7.2.1. Design and procedure. Experiment 4.1 was exactly like Experiment 2.1 (identity judgment), except that it used the 3D objects and subjects were not explicitly instructed to use an alignment-based criterion of match. (That instruction apparently had no effect in Experiment 2.1.) On each trial, four objects appeared on the screen (one target and three probes) at the same orientation. The task was to select the probe that was identical to the target. After selecting a probe, the subject identified the target by name. Trials were run in two blocks of nine. Each trained Basis object appeared as a target three times in each block, once in each orientation, in a random order.

7.2.2. Results and discussion. Trials followed by incorrect names were omitted from the analysis (1.23%). Subjects correctly selected the trained Basis probe 79.10% of the time (standard error = 5.56%). They incorrectly selected the V2 probe 14.48% of the time (3.94%), and the V1 probe 6.42% of the time (2.15%) ($t(17) = 2.63$, $p < 0.05$). Selection rates for all three probes were reliably different from chance (33%) (V2: $t(17) = 4.78$, $p < 0.01$; V1: $t(17) = 12.36$, $p < 0.01$; Basis: $t(17) = 8.29$, $p < 0.01$). These error rates are greater than those with the 2D objects (in Experiment 2.1, error rates were 9.66% and 0.93% for the V2 and V1 probes, respectively), suggesting that these objects are more generally difficult to discriminate than their 2D counterparts.

7.3. Experiment 4.2

7.3.1. Design and procedure. Experiment 4.2 used a speeded same-different judgment task identical to that of Experiment 1.2 except for the display times. Each trial began with a fixation cross (1995 ms), followed by a blank screen (495 ms), the first

Table 6.
Means and standard errors of percent 'same' responses by object pair, Experiment 4.2

Object pair	Percent 'same' responses
Basis–Basis	82.41 (5.13)
V1–V1	84.57 (4.59)
V2–V2	78.40 (5.35)
Basis–V1	40.12 (5.83)
Basis–V2	81.17 (4.86)

object (495 ms; blank 90 ms), a mask (140 ms; blank 45 ms), the second object (495 ms; blank 90 ms), and a second mask (240 ms). The two objects appeared at the same orientation on any given trial. Trials were run in three blocks of 24, where the first block served as practice. Blocks were counterbalanced in the same was as in Experiment 1.2. The data below are from the second and third blocks only.

7.3.2. Results and discussion. The results are summarized in Table 6. Subjects were reliably more likely to incorrectly respond 'same' to a target paired with its V2 variant (81.17%) than to a target paired with its V1 variant (40.12%) ($t(17) = 3.18$, $p < 0.05$). Although this difference is less marked than the same difference with the 2D objects (in Experiment 1.2, confusion rates were 88.96% and 11.58% for V2 and V1 variants, respectively), it is nonetheless substantial (the interaction, V1 vs. V2 × Experiment 1 vs. Experiment 4 is reliable; $F(1, 56) = 12.138$, $p < 0.01$). The basic effect observed in Experiments 1.2, 2.2, and 3.1 obtained with the 3D objects. The effect is attenuated relative to the effect in the experiments primarily because subjects were more likely to make Basis–V1 confusions here than in Experiments 1–3 (Experiment 1 vs. Experiment 4, $F(1, 56) = 12.138$, $p < 0.01$). The reason for this increase in V1–Basis confusions is unclear, but it is consistent with the idea that these objects are simply more difficult to discriminate than the 2D objects. It is tempting to speculate that the increase in Basis–V1 confusions reflects subjects' adopting a more coordinate-based criterion of match with these objects than with the 2D objects used in the previous experiments. However, this speculation is not supported by the fact that subjects made reliably more Basis–V2 than Basis–V1 confusions even with the 3D objects.

7.4. Experiment 4.3

7.4.1. Design and procedure. Experiment 4.3 was identical to Experiment 1.3 (speeded naming recognition) except that the objects appeared in a variety of orientations in depth. Subjects were instructed to identify each object as 'kip', 'kef', 'kor' or 'none', ignoring orientation. Trials were run in two blocks of 36. They were counterbalanced as in Experiment 1.3 except that they were also counterbalanced with respect to orientation. Presentation order was randomized within blocks. The data reported below are from both blocks.

Table 7.
Means and standard errors of percent responses by object, Experiment 4.3

		Object	
Response	Basis	V1	V2
Basis name	91.05 (2.98)	30.25 (8.42)	87.66 (3.47)
'None'	8.02 (2.78)	68.52 (8.39)	10.49 (3.42)
Other	0.93 (0.50)	1.23 (0.85)	1.85 (1.00)

7.4.2. Results and discussion. Table 7 summarizes subjects' naming responses as a function of which object appeared on a given trial. As before, the measure of interest is the likelihood of saying the name of a target in response to its V1 and V2 variants (row 1, columns 2 and 3). Subjects were again more likely to give the name of a target (Basis) object in response to its V2 variant than its V1 variant (87.66% vs. 30.25%, respectively; $t(35) = 5.54$, $p < 0.01$). However, they were also more likely to mistake a V1 variant for a target than they were in Experiment 1.3. This interaction (Experiment 1 vs. Experiment 4 x V1 vs. V2) approaches reliability ($F(1, 56) = 3.71$, $0.05 < p < 0.06$).

7.5. General discussion, Experiment 4

Experiment 4 replicated Experiments 2.1, 1.2, and 1.3 with 3D objects. In Experiment 4.1, subjects successfully discriminated the trained 3D Basis objects from their V1 and V2 variants under simultaneous presentation. With the 3D objects, as with the 2D objects, subjects' confusions in sequential same-different judgments and naming recognition were predicted by the objects' similarity in terms of the categorical relations between their parts, rather than their parts' coordinates. This finding is less amenable than the findings of Experiments 1–3 to explanation in terms of the V1 variants' differing from the Basis objects in the identity of a primitive feature. To explain these data in such terms, one would have to assume that volumetric parts in specific configurations can serve as primitive features for shape perception. There are both theoretical and empirical reasons to regard this assumption as suspect, but it is impossible to falsify the feature-based interpretation of our results in any strong sense until we know with certainty what constitutes a 'feature' for shape perception. The best we can do is accumulate evidence that the observed effects reflect the way or ways subjects represent the configuration of an object's parts, rather than the specific vocabulary of parts. Experiment 5 was designed to provide such evidence by testing another account of the effects observed in Experiments 1–4.

8. EXPERIMENT 5

In Experiments 1–4, it is possible that subjects perceived the difference between the Basis and V1 objects as a difference of the form, 'the part composed of the long horizontal connected to the short vertical has rotated 180 deg about a horizontal axis'[2]

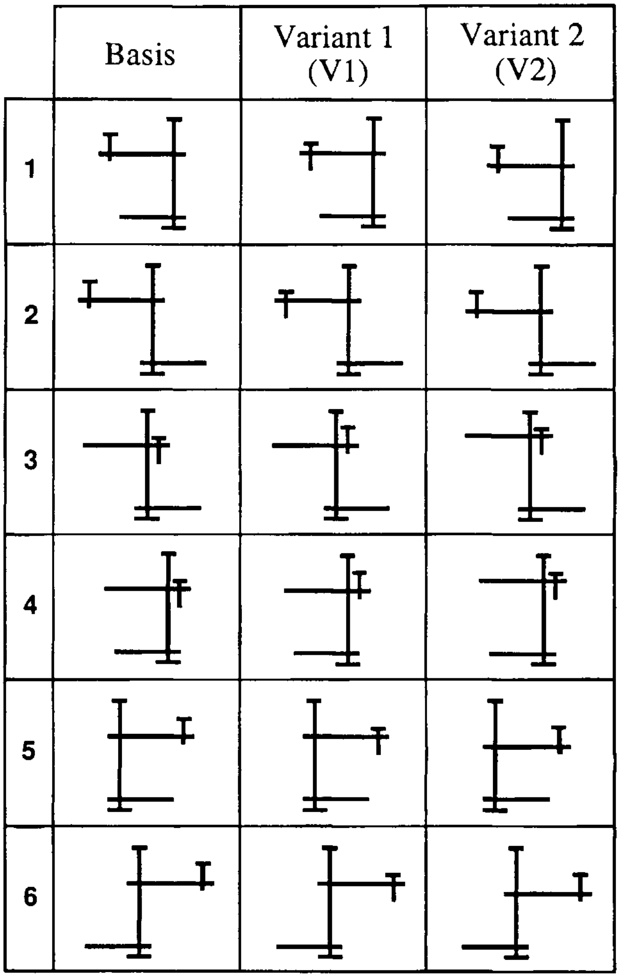

Figure 4. The stimuli (Basis objects and their variants) used in Experiment 5.

(compare the Basis and V1 objects in Fig. 1). This interpretation applies to both the 2D and the 3D stimuli. If this interpretation is correct, then our experiments would be comparing a change in the *location* of a two-part part in the V2 variants (the part formed by the horizontal and vertical lines/parts) with a change in the *orientation* of that two-part part in the V1 variants. Experiment 5 was designed to address this possibility.

The stimuli (Fig. 4) were designed to make it impossible to perceive the 'location change' in the V1 variants as a rotation of a two-part part. They were also designed to provide an additional test of the hypothesis that subjects perceive that change as a basic feature change. In these stimuli, the short vertical line that undergoes the change in the V1 variants has been moved inward along the horizontal to which it is attached (compare the V1 variants in Figs 1 and 4). The result is that the short vertical now forms an X junction with the horizontal (they formed a T junction in the original stimuli). This is important because it is possible that a T junction is not as strong a parsing cue as an X junction. If so, then subjects should be less inclined to perceive these vertical–horizontal pairs as a single two-part part. (If subjects group over both X and T junctions, then every object is just one 'holistic feature', and there is no basis—according to any theory—for predicting any difference between subjects' responses to the V1 and V2 variants.) The most important modification is the addition of a short horizontal line to the top of the short vertical. This new line serves two functions. First, it makes the rotation interpretation of the V1 variants implausible. Even if subjects group the short vertical with the long horizontal, then by virtue of the new horizontal, the resulting multi-part part in the V1 variants (here, it contains three sub-parts rather than two) is not the same as a rotation of the corresponding multi-part part of the Basis object (compare the V1 variants in Figs 1 and 4). If the pattern of effects observed in Experiments 1–4 persists with these new stimuli, then we can have confidence that they do not reflect subjects' perceiving the change in the vertical's position as a rotation.

The second function performed by the new short horizontal is that it moves with the short vertical in both the V1 and V2 variants. As such, according to a coordinate-based account, it should provide an additional source of information to distinguish both types of targets from the distractors: In the new V1 variants, two parts move (rather than one), and in the new V2 variants, three move (rather than two). (To keep the number of parts that move in the V2 variants below half the total number of parts, we also added a short horizontal to the top of the tall vertical of each object. In the new V2 variants, three of seven parts move, and in the V2 variants in Fig. 2, two of five parts move.) A coordinate-based measure of match predicts that the Basis and variant objects in Fig. 4 will be more discriminable than those in Fig. 1 (since a larger fraction of their parts move); this greater discriminability should manifest itself as lower error rates in this experiment than in Experiment 1 (for both the V1 and V2 variants). By contrast, if subjects perceive these objects in terms of their parts' categorical relations, then the opposite pattern is expected: The new horizontals introduce additional relations that do *not* change from a Basis object to either of its variants (making the basis objects more similar to their variants in Fig. 4 than in Fig. 1), so the basis–variant pairs in Fig. 4 should be more difficult to discriminate than those in Fig. 1.

8.1. Method

8.1.1. Subjects. Twelve subjects participated in this experiment.

8.1.2. Stimuli. The objects in Fig. 4 served as stimuli.

8.2. Experiment 5.1

Experiment 5.1 was identical to Experiment 1.1 (similarity judgment) except for the stimuli.

Table 8.
Means and standard errors of percent 'same' responses by object pair, Experiment 5.2

Object pair	Percent 'same' responses
Basis–Basis	87.04 (2.85)
V1–V1	86.11 (4.35)
V2–V2	96.30 (1.58)
Basis–V1	63.96 (9.31)
Basis–V2	86.57 (4.86)

8.2.1. Results and discussion. Similarity judgments followed by incorrect names (7.73%) were omitted from the analysis. Subjects chose the V2 variants as more similar to the Basis objects (targets) 72.62% of the time and the V1 variants 27.38% of the time ($t(11) = 2.34$, $p < 0.05$). This pattern replicates the basic pattern observed in Experiments 1.1, 2.1, and 4.1.

8.3. Experiment 5.2

Experiment 5.2 was identical to Experiment 1.2 (speeded sequential same-different judgment) except for the stimuli.

8.3.1. Results and discussion. Table 8 summarizes subjects' responses to the various object pairs. Subjects were again more likely to say 'same' in response to Basis–V2 pairs than Basis–V1 pairs (86.57% vs. 63.96%, respectively; $t(11) = 2.55$, $p < 0.05$), replicating the basic pattern observed in Experiments 1.2, 2.2, 3.1, and 4.2. However, the difference between the Basis–V1 and Basis–V2 'same' response rates was less marked than the corresponding difference in Experiment 1 (the interaction, V1 vs. V2 × Experiment 1 vs. Experiment 5 is statistically reliable; $F(1, 44) = 20.82$, $p < 0.01$): Subjects were more likely to respond 'same' to a Basis–V1 pairs in this experiment than in Experiment 1.2 (63.96% vs. 11.58%, respectively; $t(22) = 5.25$, $p < 0.01$). This increase in Basis–V1 confusions is predicted by the account that subjects represent the objects in terms of their parts' locations relative to one another; it is inconsistent with the idea that they represent them in terms of their coordinates. However, counter to the account that subjects represent these objects in terms of their parts interrelations, the Basis–V2 confusion rate was about the same in this experiment (86.57%) as in Experiment 1.2 (87.96%) ($t(22) = 0.824$, $0.9 > p > 0.8$), although this may reflect a floor effect.

8.4. Experiment 5.3

Experiment 5.3 was identical to Experiment 1.3 (speeded naming recognition) except for the stimuli.

8.4.1. Results and discussion. Table 9 summarizes subjects' responses to the target objects and their variants. Subjects again mistook V2 variants for targets reliably more

Table 9.
Means and standard errors of percent responses by object, Experiment 5.3

	Object		
Response	Basis	V1	V2
Basis name	88.19 (5.42)	45.14 (11.53)	77.08 (7.19)
'None'	10.07 (5.22)	53.47 (11.98)	18.75 (6.42)
Other	1.74 (0.95)	1.39 (0.94)	4.17 (2.41)

often than they mistook V1 variants for targets (77.08% vs. 45.14%, respectively; $t(11) = 2.33$, $p < 0.05$). As predicted by the hypothesis that subjects perceive these objects in terms of their parts' interrelations, subjects mistook the V1 variants for targets more frequently in this experiment than they did in Experiment 1.3 (45.14% vs. 6.25%, respectively; $t(22) = 3.33$, $p < 0.01$). But counter to this account, subjects did not mistake V2 variants for targets more often in this experiment (77.08%) than in Experiment 1.3 (86.11%). Indeed, there was a slight trend in the opposite direction, as predicted by a coordinate-based account, but this trend does not approach statistically reliability ($t(22) = 0.27$, $0.3 > p > 0.2$). Again, the basic effect observed in Experiments 1.3, 2.3, 3.3, and 4.3 replicated with the new stimuli.

8.5. General discussion, Experiment 5

Experiment 5 replicated the basic pattern of effects observed in Experiments 1–4, suggesting that those effects do not reflect subjects' perceiving our manipulation as a rotation of a two-part part. Rather, it is more likely perceived as a change in the location of the vertical line, as intended. Moreover, as predicted by the hypothesis that subjects perceive these objects in terms of their parts' relations to one another, Basis–V1 confusions were more frequent in this experiment than in Experiment 1. Neither of these effects is consistent with the idea that we represent an object's features or parts in terms of their coordinates in a spatial reference frame.

9. GENERAL DISCUSSION

The question of how we perceive the configuration of an object's features or parts is important, but virtually nothing is known about its answer. The five experiments reported here used a variety of perceptual tasks to test subjects' perceptual sensitivity to the categorical relations and coordinates of the parts of simple 2D and 3D objects. The results consistently supported the hypothesis that people perceive the configuration of an object's features or parts in terms of their pair-wise categorical relations. Across a wide variety of tasks and stimuli, subjects were consistently more likely to confuse Basis objects with their V2 variants, which share categorical relations with the Basis objects, than with their V1 variants, which differ from the Basis objects in the categorical relations among their parts. This effect obtained even though the V1 variants are more similar to the Basis objects in terms of their parts' coordinates.

These data pose a challenge to current 'view-based' models of shape perception, which represent objects in terms of their features' numerical coordinates in a spatial reference frame (Ullman, 1989; Poggio and Edelman, 1990; Ullman and Basri, 1991; Bülthoff and Edelman, 1992; Olshaussen *et al.*, 1993; Vetter *et al.*, 1994; Bülthoff *et al.*, 1995; Tarr, 1995; Edelman *et al.*, 1996). Our subjects' inability to distinguish the Basis objects from their V2 variants with above-chance accuracy suggests that, to the extent that coordinates play any explicit role in shape perception, they are likely to be specified only very coarsely (Hummel and Stankiewicz, 1996). The six-pixel change that distinguishes the Basis objects from their V2 variants is over 10% of those objects' total height (57 pixels)—a figure well within the numerical precision required by the normalization procedures performed by extant view-based models.

It is important to emphasize the generality of this result. The stimuli were created such that one of the two parts that moved in any V2 variant was always the part that moved in the corresponding V1 variant. Therefore, as noted in the Introduction, the predictions of a coordinate-based model with regard to our stimuli cannot be reversed (i.e. to conform to our findings) simply by differentially weighting the various features in our stimuli (e.g. as described by Edelman and Poggio, 1991). Any weighting that would make the Basis–V1 discriminability greater than the Basis–V2 discriminability would also make the Basis–Basis discriminability greater than the Basis–V2 discriminability.

It is also impossible to account for our findings simply by assuming that objects are perceived in terms of a hierarchy of coordinate systems. This point bears elaborating. A natural hypothesis to generate in response to our findings is the following: Although the data are inconsistent with perception based on any single linear[3] coordinate system, it might be possible to account for them in terms of a hierarchy of linear coordinate systems. For example, each part might be represented in terms its coordinates relative to a reference point on some other part (e.g. the part or parts to which the former is attached). Like a structural description, this type of representation would specify the locations of an object's features relative to multiple reference points; but like a view-based representation, each reference point would serve as the origin of a simple linear coordinate system. This proposal cannot account for our findings because any set of hierarchical linear coordinates is equivalent to some set of 'flat' (i.e. single-reference-point) linear coordinates, so the similarity relations that derive from any hierarchy of linear coordinate systems are equivalent to the similarity relations that derive from a single linear coordinate system. The findings reported here are inconsistent with any model of shape perception based on any set of linear coordinates.

An intuitive explanation of our data—and the hypothesis that motivated the experiments—is that our subjects perceived our stimuli in terms of the categorical relations among their parts (as suggested by structural description theories of shape perception; Biederman, 1987; Hummel and Biederman, 1992; Dickenson *et al.*, 1993). More specifically, the data suggest that (a) the configuration of an object's parts is perceived with respect to multiple reference points, and (b) the perceived value of a relation varies in a non-linear fashion with the relative location of the points in the image.

What are the reference points and what is the nonlinearity? In the next section, we propose a simple, preliminary model that attempts to address these questions. The model defines relations on the midpoints of connected lines (as we assumed earlier with coordinate-based analyses), and codes the relation between two such points as a logistic function of the difference between their coordinates in the image. (In the general model, relations may be defined on several points, including endpoints, midpoints and others. For simplicity, the simulations reported below are based on midpoints only.) We chose the logistic function for three reasons. First, it is attractive from a computational perspective: A logistic coding of relative location has all the advantages associated with a categorical representation (e.g. robustness to stimulus noise and variations in viewpoint, as noted in the Introduction) without many of the disadvantages (e.g. insensitivity to potentially important metric properties; see Stankiewicz and Hummel, 1996). Second, it is consistent with categorical effects that have been observed in vision (e.g. Foster and Ferraro, 1989), and other domains, such as speech perception. And third, it is simple. The resulting model provides an excellent qualitative account of our findings with only a single free parameter.

9.1. A preliminary model of the perception of spatial relations

Consider two points, i and j, and let the *scaled position* of i relative to j along axis a (where $a = X$ and $a = Y$ are the horizontal and vertical image axes) be

$$P_a(i, j) = (a_i - a_j)/(l_{a_i} + l_{a_j}), \tag{1}$$

where a_i is the coordinate of i on a, a_j is the coordinate of j on a, and l_{a_i} and l_{a_j} are the lengths, along a, of the parts to which i and j belong. The scaled position of i relative to j along a is simply the difference of their coordinates on a scaled by the lengths (on a) of the parts to which they belong. Scaling P in this fashion makes it scale-invariant (i.e. it will not change with the absolute size of the object's image). Shape perception (at least for recognition) is scale-invariant (Biederman and Cooper, 1992). As discussed shortly, the nonlinearity introduced by this scaling also provides a good account of the experimental findings reported here all by itself (although it does not provide a general account of categorical effects in perception).

Consider Fig. 5(a). Let a be the vertical axis (Y), let i be point A, and let j be point B. $l_{Y\mathrm{B}}$ is the length, along Y, of the line to which B belongs (15; $l_{Y\mathrm{A}} = 1$). By Eqn (1), $P_Y(\mathrm{A}, \mathrm{B})$ is $(30 - 25)/(1 + 15)$, or 0.31. For fixed values of $l_{Y\mathrm{A}}$ and $l_{Y\mathrm{B}}$ (i.e. for parts of a constant size), $P_Y(\mathrm{A}, \mathrm{B})$ varies linearly with Y_A (the location of A on Y). Let us define $R_a(i, j)$, the relation of i to j along a, as the logistic function

$$R_a(i, j) = \frac{1}{1 + \mathrm{e}^{-\kappa P_a(i, j)}}, \tag{2}$$

where κ is a scaling constant. Figure 5(b) illustrates R as a function of P. R is 0.5 when P is 0, that is, when i and j have the same coordinate on axis a ($a_i = a_j$); R is less than 0.5 when a_i is less than a_j (e.g. for $a = Y$, when i is below j); and R is greater than 0.5 when a_i is greater than a_j (for $a = Y$, when i is above j). Most importantly, the derivative of R is greatest where $P = 0$; that is, for a constant change in P, the change in R is greatest near the categorical boundary of P, and a

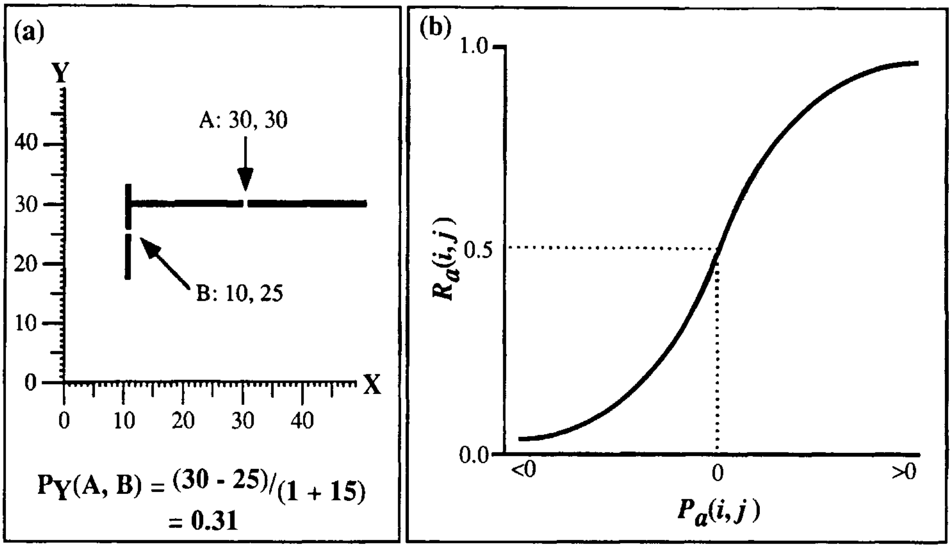

Figure 5. (a) Illustration of $P_a(i, j)$, the scaled position of point i relative to point j along axis a. Here, $a = Y$ (the vertical image axis), i = point A, and j = point B. The horizontal and vertical lines correspond to the parts to which A and B belong, respectively. A and B are depicted as white gaps in the lines. $P_Y(\mathrm{A}, \mathrm{B})$ is greater than zero because point A is above point B. (b) $R_a(i, j)$ as a function of $P_a(i, j)$.

Table 10.
Simulation results with the logistic model of relations: discriminability of the basis objects and variants (Experiments 1–3 and Experiment 5)

	Stimulus set	
Object pair	Experiments 1–3	Experiment 5
Basis–V1	0.2451	0.1634
Basis–V2	0.0311	0.0207

given change in P produces the greatest change in R when P crosses the categorical boundary (e.g. when i goes from below j to above it). In this respect, R is categorical. However, R differs from a strictly categorical relation in that it does not discard all metric information. κ determines the steepness of R. When $\kappa = \infty$, R is a step function that evaluates to zero for all $P < 0$ and to 1 for all $P > 0$ (here, R is completely categorical); for all other $\kappa > 0$, R is a sigmoid whose steepness is proportional to κ.

Let an object be represented as a vector $\boldsymbol{r}$ of relations $R_a(i, j)$ between the midpoints on all pairs of connected[4] parts. Let $D(m, n)$, the discriminability of objects m and n, be proportional to the Euclidean distance between $\boldsymbol{r}_m$ and $\boldsymbol{r}_n$, the vectors describing m and n, respectively. Specifically,

$$D(m, n) = \|\boldsymbol{r}_m - \boldsymbol{r}_n\| / N_r, \tag{3}$$

where N_r is the dimensionality of $\boldsymbol{r}_m$ and $\boldsymbol{r}_n$. Scaling D by N_r causes D to vary with the proportion (rather than the absolute number) of relations that differ between m and n. With $\kappa = 20$, this model produces the Basis–V1, and Basis–V2 discriminability values shown in Table 10. Standard errors are not reported because the values are the same across all sets of stimuli (i.e. the variance in the data is zero).

A few things are worthy of note in Table 10. First, the V2 variants are substantially more similar to (less discriminable from) the Basis objects than are the V1 variants. The model's estimate of the Basis–V1 discriminability is almost eight times its estimate of the Basis–V2 discriminability (with the stimuli used in both Experiments 1–3 and Experiment 5). The model predicts that Basis–V2 confusions are much more likely than Basis–V1 confusions (the basic pattern observed in all the experiments). Second, the Basis–V2 discriminability is non-zero: Albeit difficult, it should be possible to distinguish the Basis objects from their V2 variants (as observed in Experiment 3). And third, the Basis–variant discriminability is lower for the objects used in Experiment 5 than for the objects used in Experiments 1–3, predicting more confusions in Experiment 5 than in Experiment 1. This prediction was only partially supported by the data. Basis–V1 confusions were more frequent in Experiment 5 than in Experiment 1, but Basis–V2 confusions were not. However, as noted previously, the failure to observe more Basis–V2 confusions in Experiment 5 than in Experiment 1 may reflect a floor effect. In summary, this simple model—based on a non-linear numerical representation of the pair-wise relations between object parts—does a remarkably good job accounting for virtually all the effects reported here.

The single free parameter in the model is κ, the steepness parameter. Although we have not run the necessary simulations, it is interesting to speculate that κ (or something functionally equivalent; see Stankiewicz and Hummel, 1996) may provide at least a partial account of Foster and Ferraro's (1989) finding that judgments of relative position become progressively less categorical at stimulus exposure durations over 100 ms. Imagine that, immediately after the presentation of a stimulus (i.e. early in perceptual processing), κ takes a relatively high value and that, with additional processing, it gradually decays toward a lower value. If so, then perception would initially be relatively categorical, and with time, become progressively more linear. The categorical properties of a stimulus are more robust to noise (either in the stimulus or in the nervous system) than are its precise metric properties (see e.g. Biederman, 1987). This kind of categorical-to-metric processing could therefore permit the visual system to make a rapid initial guess about stimulus shape (based on its categorical properties), followed by progressively more refined estimates of shape with additional sampling time (Stankiewicz and Hummel, 1996). At this point, these considerations are merely speculation, but they make sense from a computational perspective (i.e. in terms of the mathematics of stimulus sampling in the face of noise), and they are consistent with the findings Foster and Ferraro.

It is important to note a model based strictly on $P_a(i, j)$, the scaled position of point i relative to point j (Eqn (1)), can also provide a qualitative account of our findings (although it provides no basis for accounting for other categorical effects in perception, such as those reported by Foster and Ferraro). In a manner analogous

Table 11.
Simulation results with the non-logistic model of relations: discriminability of the basis objects and variants (Experiments 1–3 and Experiment 5)

	Stimulus set	
Object pair	Experiments 1–3	Experiment 5
Basis–V1	0.1154	0.0769
Basis–V2	0.0200	0.0134

to the logistic model, let objects m and n be perceptually coded as vectors $\mathbf{p}_m$ and $\mathbf{p}_n$ of the scaled relative locations $P_a(i, j)$ of all pairs of touching parts; also like the previous model, let us define, $D(m, n)$, the discriminability of objects m and n, according to Eqn (3) (with $\mathbf{p}$ substituted for $\mathbf{r}$). The only difference between this model and the previous logistic model is that the logistic model uses the logistic function (Eqn (2)), whereas this one does not. The resulting model produces the Basis–V1, and Basis–V2 discriminability values shown in Table 11; again, there are no standard errors because there is no variance in the data.

Like the logistic model, this model predicts that Basis–V2 confusions will be more frequent than Basis–V1 confusions, and that both types of errors will be more frequent with the stimuli used in Experiment 5 than with those used in Experiments 1–3. Indeed, with respect to the stimuli used in the experiments reported here, the only difference between the models is that the logistic model predicts a greater difference in performance between Basis–V1 and Basis–V2 pairs than does the non-logistic model. The logistic model predicts that Basis–V1 pairs will be about eight times as discriminable as Basis–V2 pairs, whereas the non-logistic model predicts that Basis–V1 pairs will be about six times as discriminable. Importantly, both models' capacity to account for our findings rests on the fact that, in both models, the representation of a feature's location is a nonlinear function of its location in the image. (In the case of the non-logistic model, the nonlinearity is in the scaling by the parts' extent in the image; Eqn (2).)

9.2. Summary and conclusions

Naturally, the data reported here do not warrant proposing either of these models model as a general theory of the perception of spatial relations. Far too little is known about this issue to propose such a theory at this time. But the data for which the models provide a preliminary account underscore the importance of relations in human shape perception. The effects reported here are not subtle; they are completely intuitive on casual inspection of our stimuli. However, they are exactly the opposite of what is expected on the assumption that we perceive objects in terms of the coordinates of their features or parts. As such, they cast doubt on the generality of current view-based theories of human shape perception and object recognition. Rather, they suggest that the visual system may exploit the properties of categorical relations—robustness to noise, variations in viewpoint, and the vagaries of an object's precise shape—for the purposes of shape perception and object recognition.

Acknowledgements

This work was supported by a research grant from the UCLA Academic Senate to JEH, and by a Grant-in-Aid of Research to BJS from the National Academy of Sciences, through Sigma Xi, The Scientific Research Society. We are grateful to Adam Reeves, David Foster, Keith Holyoak, Eric Cooper, Nancy Kanwisher, and Lisa Travis, and two anonymous reviewers for helpful comments on earlier drafts of this manuscript, and to Irv Biederman, Nancy Kanwisher, and Jun Saiki for discussions about this work.

NOTES

1. The Euclidean distance between two vectors is the square root of the sum of squared differences between corresponding elements. In the case of the Basis objects and their V1 variants, all coordinates are identical except for one, which differs by 1. Thus, the sum of squared vector differences is 1.0, the square root of which is 1.0. In the case of the Basis objects and their V2 variants, all coordinates are identical except for two, each of which differs by 1. Here, the sum of squared vector differences is 2.0, so the Euclidean distance is $\sqrt{2.0}$.

2. We are grateful to Jierre Jolicoeur for calling this alternative interpretation of the effect to our attention.

3. A linear coordinate system is a coordinate system in which the numerical value of a coordinate scales linearly with the location of the corresponding point in the reference frame. The coordinate systems used by view-based models are all linear in this sense.

4. Saiki and Hummel (1996) showed that the visual system is more sensitive to the relations between connected parts than to the relations between non-connected parts.

REFERENCES

Besner, D. and Coltheart, M. (1975). Mental size scaling examined. *Memory and Cognition* **4**, 525–531.

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

Biederman, I. and Cooper, E. E. (1991). Priming contour deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology* **23**, 393–419.

Biederman, I. and Cooper, E. E. (1992). Size invariance in visual object priming. *J. Exper. Psychol.: Human Percep. Perform.* **18**, 121–133.

Biederman, I. and Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: A critical analysis. *J. Exper. Psychol.: Human Percep. Perform.* **21**, 1506–1514.

Bülthoff, H. H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Nat. Acad. Science* **89**, 60–64.

Bülthoff, H. H., Edelman, S. Y. and Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex* **3**, 247–260.

Bundesen, C. and Larsen, A. (1975). Visual transformation of size. *J. Exper. Psychol.: Human Percep. Perform.* **1**, 214–220.

Bundesen, C., Larsen, A. and Farell, J. E. (1981). Mental transformations of size and orientation. In: *Attention and Performance IX*. J. Long and A. Baddeley (Eds). Erlbaum, Hillsdale, pp. 279–294.

Clowes, M. B. (1967). Perception, picture processing and computers. In: *Machine Intelligence*. N. L. Collins and D. Michie (Eds). Oliver and Boyd, Edinburgh, pp. 181–197.

Cooper, E. E., Biederman, I. and Hummel, J. E. (1992). Metric invariance in object recognition: A review and further evidence. *Can. J. Psychol.* **46**, 191–214.

De Valois, K. K., Lakshminaryanan, V., Nygaard, L., Schlussel, S. and Sladky, J. (1990). Discrimination of relative spatial position. *Vision Res.* **30**, 1649–1660.

Dickinson, S. J., Pentland, A. P. and Rosenfeld, A. (1992). 3D shape recovery using distributed aspect matching. *IEEE Trans. Pattern Anal. Machine Intelligence* **14**, 174–198.

Edelman, S., Cutzu, F. and Duvdevani-Bar, S. (1996). Similarity to reference shapes as a basis for shape representation. In: *Proc. 18th Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, pp. 260–265.

Edelman, S. and Poggio, T. (1991). Bringing the grandmother back into the picture: A memory-based view of object recognition. *MIT A.I. Memo* No. 1181, April.

Edelman, S. and Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics* **64**, 209–219.

Ellis, R. and Allport, D. A. (1986). Multiple levels of representation for visual objects: A behavioural study. In: *Artificial Intelligence and its Applications*. A. G. Cohen and J. R. Thomas (Eds). Wiley, New York, pp. 245–257.

Foster, D. H. and Ferraro, M. (1989). Visual gap and offset discrimination and its relation to categorical identification in brief line-element displays. *J. Exper. Psychol.: Human Percep. Perform.* **15**, 771–784.

Hoffman, D. D. and Richards, W. A. (1985). Parts of recognition. *Cognition* **18**, 65–96.

Howard, J. H. and Kerst, S. M. (1978). Directional effects of size change on the comparison of visual shapes. *Am. J. Psychol.* **91**, 491–499.

Hummel, J. E. and Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review* **99**, 480–517.

Hummel, J. E. (1994). Reference frames and relations in computational models of object recognition. *Current Directions in Psychological Science* **3**, 111–116.

Hummel, J. E. (1995). Object recognition. In: *The Handbook of Brain Theory and Neural Networks*. M. Arbib (Ed.). MIT Press, Cambribge, pp. 658–660.

Hummel, J. E. and Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In: *Attention and Performance XVI: Information Integration in Perception and Communication*. T. Inui and J. McClelland (Eds). MIT Press, Cambridge, pp. 93–121.

Jolicoeur, P. and Besner, D. (1987). Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. *J. Exper. Psychol.: Human Percep. Perform.* **13**, 478–487.

Larsen, A. (1985). Pattern matching: Effects of size ratio, angular differences in orientation, and familiarity. *Perception and Psychophysics* **38**, 63–68.

Lowe, D. G. (1987). The viewpoint consistency constraint. *Int. J. Computer Vision* **1**, 57–72.

McClelland, J. L. and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* **88**, 375–407.

Marr, D. (1982). *Vision*. Freeman, San Francisco.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of three dimensional shapes. *Proc. Royal Soc. London, Series B* **200**, 269–294.

Neisser, U. (1967). *Cognitive Psychology*. Appleton-Century-Crofts, New York.

Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neuroscience* **13**, 4700–4719.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In: *Cognition and Categorization*. E. Rosch and B. B. Lloyd (Eds). Lawrence Erlbaum, Hillsdale, pp. 259–303.

Palmer, S. E. and Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review* **1**, 29–55.

Pinker, S. (1984). Visual cognition: An introduction. In: *Visual Cognition*. S. Pinker (Ed.). Elsevier, Amsterdam, pp. 1–62.

Poggio, T. and Edelman, S. (1990). A neural network that learns to recognize three-dimensional objects. *Nature* **317**, 314–319.

Poggio, T. and Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science* **247**, 978–982.

Quinlan, P. T. (1991). Differing approaches to two-dimensional shape recognition. *Psychological Bulletin* **109**, 224–241.

Rock, I. (1983). *The Logic of Perception*. MIT Press, Cambridge, MA.

Saiki, J. and Hummel, J. E. (1996). Connectedness and the integration of parts with relations in shape perception. Manuscript submitted for publication.

Siebert, M. and Waxman, A. M. (1992). Learning and recognizing 3D objects from multiple views in a neural system. In: *Neural Networks for Perception, Vol. 1, Human and Machine Perception*. H. Wechsler (Ed.). Academic Press, pp. 427–444.

Stankiewicz, B. J. and Hummel, J. E. (1996). MetriCat: A representation for basic and bubordinate-level classification. In: *Proc. 18th Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, pp. 254–259.

Sutherland, N. S. (1968). Outlines of a theory of visual pattern recognition in animals and man. *Proc. Royal Soc. London, Series B* **171**, 95–103.

Tarr, M. J. and Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology* **21**, 233–283.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review* **2**, 55–82.

Ullman, S. (1989). Aligning pictoral descriptions: An approach to object recognition. *Cognition* **32**, 193–254.

Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Trans. Pattern Anal. Machine Intelligence* **13**, 992–1006.

Vetter, T., Poggio, T. and Bülthoff, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology* **4**, 18–23.