*Research Article*

# Independent Processing of Parts and of Their Spatial Organization in Complex Visual Objects

**Martin Arguin**[1] **and Daniel Saumier**[1,2]

[1]*Groupe de Recherche en Neuropsychologie Expérimentale, Département de Psychologie, Université de Montréal, and Centre de Recherche, Institut Universitaire de Gériatrie de Montréal, Montreal, Quebec, Canada, and* [2]*Bloomfield Centre for Research in Aging, Lady Davis Institute for Medical Research, McGill University, Montreal, Quebec, Canada*

ABSTRACT—*A visual search experiment using synthetic three-dimensional objects is reported. The target shared its constituent parts, the spatial organization of its parts, or both with the distractors displayed with it. Sharing of parts and sharing of spatial organization both negatively affected visual search performance, and these effects were strictly additive. These findings support theories of complex visual object perception that assume a parsing of the stimulus into its higher-order constituents (volumetric parts or visible surfaces). The additivity of the effects demonstrates that information on parts and information on spatial organization are processed independently in visual search.*

A fundamental issue in understanding human visual object recognition concerns the representation of shape, which is its main determinant (Biederman & Ju, 1988). The shapes of most everyday objects can be considered complex, in that they subjectively appear to be made of a number of perceptually or functionally distinct components. The present study investigated the representation of complex visual shapes.

A number of theories of visual shape perception explicitly assume that the general problem of shape complexity is broken down into a set of more manageable, smaller problems by parsing the object into its constituent parts (Biederman, 1987; Hoffman & Richards, 1984; Hoffman & Singh, 1997; Kurbat, 1994; Marr, 1982; Marr & Nishihara, 1978; Palmer, 1977; Singh, Seyranian, & Hoffman, 1999). Accordingly, researchers have proposed rules for segmenting complex shapes into component parts and for grouping low-level image features into an integrated representation of these parts (Biederman, 1987; Donnelly, Humphreys, & Riddoch, 1991; Hoffman & Richards, 1984; Hummel & Biederman, 1992). In another theoretical approach, which is broadly compatible with the notion of part-based representations, objects are coded in terms of collections of visible surfaces (Lee & Park, 2002; Leek & Arguin, 2000; Leek, Reppa, & Arguin, 2002; Pentland, 1989). This alternative view agrees with part-based theories in proposing that visual shapes are parsed into higher-order units (i.e., units of a greater scale than local features, such as vertices). However, it differs from part-based theories in that surfaces are not grouped to create explicitly coded parts. In addition, it makes no assumption regarding surfaces that are not visible, such as those on the side of the object that is opposite to the observer (in contrast, e.g., to the widely held part-based theory of geon structural descriptions; Biederman, 1987; Hummel & Biederman, 1992). Although the present experiment was not initially designed to test a surface-based theory of visual shape perception, it provides theoretically relevant observations (see Discussion). Although other theoretical views remain silent or ambiguous with respect to the issue of the parsing of complex shapes into higher-order units, they nevertheless allow for this possibility (see Tarr & Bülthoff, 1998, for a relevant discussion).

Investigators studying visual shape perception do not all agree that shapes are parsed into higher-order units, as proposed by the views we have just outlined. For instance, some authors have suggested that shape may be coded in terms of the global properties of the entire stimulus, such as Fourier descriptors (Cortese & Dyre, 1996). Alternatively, others have suggested that shape representations may rest entirely on local, low-level features (e.g., oriented edges, vertices) and on the relations among them (e.g., Edelman & Weinshall, 1991; Lades et al., 1993; Poggio & Edelman, 1990; Ullman, 1989).

In the present study, we tested the hypothesis that complex visual shapes are represented in terms of distributed collections of parts by

**Address correspondence to Martin Arguin, Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Quebec, Canada H3C 3J7; e-mail: martin.arguin@umontreal.ca.**

Copyright © 2004 American Psychological Society

assessing one straightforward prediction implied by this view—that it should be more difficult to discriminate between complex shapes if they share their component parts than if they do not. Given the construction of the stimuli used in this experiment (described later), the part-sharing effect we assessed is one that is robust across depth rotations of the component parts of the stimuli.

Another crucial issue for theories that assume complex shapes are represented in terms of their constituent parts concerns how the spatial organization of these parts is coded. One possibility is that the organization of parts is coded as a global configuration, that is, as a spatial structure within which each part has its own place (Kosslyn, 1994; Marr, 1982; Marr & Nishihara, 1978). Alternatively, spatial organization may be registered in terms of the spatial relations among parts (e.g., part A is above part B; Biederman, 1987; Hummel & Biederman, 1992). Current empirical knowledge regarding the representation of the spatial organization of complex objects is limited to the fact that part connectedness facilitates the coding of spatial relations (Saiki & Hummel, 1998a, 1998b). The present experiment examined whether the difficulty of discriminating complex shapes is affected by whether items share the spatial organization of their component parts.

Another important question that we considered is whether there is a form of interdependency between the processing of object parts and the processing of their spatial organization, or whether these processes occur independently. We explored this issue by jointly manipulating whether items shared parts and spatial organization. Indeed, if one process (e.g., processing of spatial organization) is contingent upon the other (i.e., coding of parts), for instance, then finding an effect of the factor drawing on this process (i.e., items share vs. do not share their spatial organization) should be contingent on stimuli having the same value on the other factor (i.e., sharing component parts). If the processes are independent, no such contingency should be observed, and the effects of part and organization sharing should be additive. Finally, if there is no functional separation between processing of parts and spatial organization, as some theories assume, then the factors of part sharing and organization sharing should have interactive effects because both should draw on the same processing stage.

We used a visual search task with complex synthetic three-dimensional objects constructed from the juxtaposition of elementary volumetric parts (see Fig. 1). On every trial, the subject indicated whether the target was present or absent. Across trials, the total number of stimuli displayed was varied. Distractors were made of the same (but rearranged) parts as the target or of different parts. In addition, distractors had either the same spatial organization of their component parts as the target or a different spatial organization. The effects of these factors were assessed through their impact on the rate of visual search.

## METHOD

### Subjects

Twelve students from the University of Montreal, Canada, took part in the experiment. All had normal or corrected visual acuity, and they were naive as to the purpose of the experiment.

### Stimuli and Apparatus

Examples of the stimuli used are shown in Figure 1b. Stimuli were constructed and rendered through ray tracing using the Sketch!™ program from Alias Research Inc. (Toronto, Ontario, Canada). They were composed of volumetric parts such as those illustrated in Figure 1a. These elementary parts varied according to the global dimensions of aspect ratio (elongated vs. short), curvature of the major axis (straight vs. curved), tapering along the major axis (tapered vs. not tapered), and shape of cross section (circular vs. square). Biederman (1987) has argued that the latter dimension is relevant for the reliable recovery of shape information across rotations in depth. Empirical support for the psychological validity of the dimensions of aspect ratio, curvature, and tapering with respect to visual shape encoding and memory may be found elsewhere (Arguin, Bub, & Dudek, 1996; Arguin & Saumier, 2000; Saumier & Arguin, 2003; Stankiewicz, 2002).

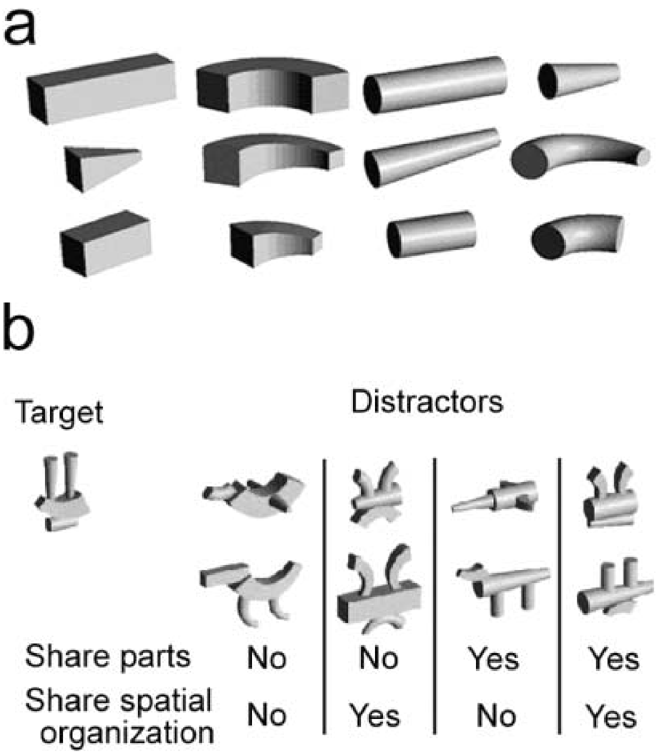

**Fig. 1.** Examples of the volumetric parts used to construct the stimuli (a) and of the objects used as stimuli (b).

Each object was made of a large component, which had a constant three-dimensional orientation across objects, plus three smaller components attached to it. Two of these minor components were replications of the same basic shape and were placed symmetrically about the main component. All objects had a constant matte gray surface and were rendered with the same source of illumination. Objects varied according to their constituent parts and their spatial organization. Three objects served as targets; these differed from one another in both their constituent parts and their spatial organization. One form of spatial organization is illustrated by the target in Figure 1b and the distractors in the second and fourth columns of that figure; objects with this organization were informally referred to as "plugs." The other two kinds of objects were informally referred to as "birds" and "animals"; examples of these spatial organizations are displayed in the first and third columns of distractors in Figure 1b. For each condition and for each possible target, two objects served as distractors. The largest horizontal and vertical extents of the stimuli were 2.2° and 2.1°, respectively, as seen from a viewing distance of 90 cm.

When the target and distractors shared their parts, these shared parts generally had a different orientation in depth so that their retinal projections were markedly different (see Fig. 1b). Therefore, an effect of part sharing between the target and distractors would indicate that the part representations mediating search performance are resistant to depth rotations.

The experiment was run on a Macintosh PowerPC 7100 computer, and stimuli were displayed on a high-resolution, 19-in. Apple monitor.

### Procedure and Design

Subjects were instructed to indicate, on every trial, whether a particular target object was present (50% of trials) in a display of a variable number of items. On target-present trials, the two relevant distractors were replicated an equal number of times. We maintained constant display size across target-present and target-absent trials by presenting an unequal number of replications of the two distractors on target-absent trials. For these trials, one distractor was replicated one more time than the other, with the two distractors occurring an equal number of times within each block.

Trials began with a 500-ms fixation point presented at the center of the computer screen. This was followed by a 500-ms blank screen and then by the stimulus display, which remained visible until the subject responded. Each target and distractor was randomly presented at 1 of 12 locations equally spaced on an imaginary circle that was 9.5° in diameter and centered on the fixation point.

The factors of part and organization sharing were blocked, and each target was tested in different blocks as well, resulting in a total of 12 blocks of trials of 160 trials each. An equal number of trials for each combination of the different levels of the factors of target presence and number of items was presented in a random order within each block. Each subject completed all 12 blocks of trials in three 45-min sessions (only one target served in each session) that were scheduled on different days over a 3-week period. Block order was random across subjects.

The independent variables were part and organization sharing between the target and distractors (shared vs. not shared for each factor), target presence (present or absent), and the number of items displayed (3, 5, 7, or 9). These were all within-subjects factors.

## RESULTS

Response times (RTs) for correct trials in each condition are shown in Figure 2. Table 1 presents the results of the linear regression analyses of RTs as a function of the number of items. The correlation between correct RTs and error rates was +.70 ($p < .01$), thus indicating no speed-accuracy trade-off. In addition, the overall error rate was only 2.1%. Therefore, no further analyses of error rates were conducted.

The RT data indicate that the factors of part sharing and spatial organization sharing had strictly additive effects. This can be seen most easily in Table 2, which presents the separate and joint effects of part and spatial-organization sharing on target-present and target-absent slopes, as well as on the estimated processing time per item. Effect magnitudes were calculated in the following way from the slopes of correct RTs as a function of the number of items displayed:

> effect of part sharing = slope when only parts were shared minus slope when neither parts nor spatial organization was shared

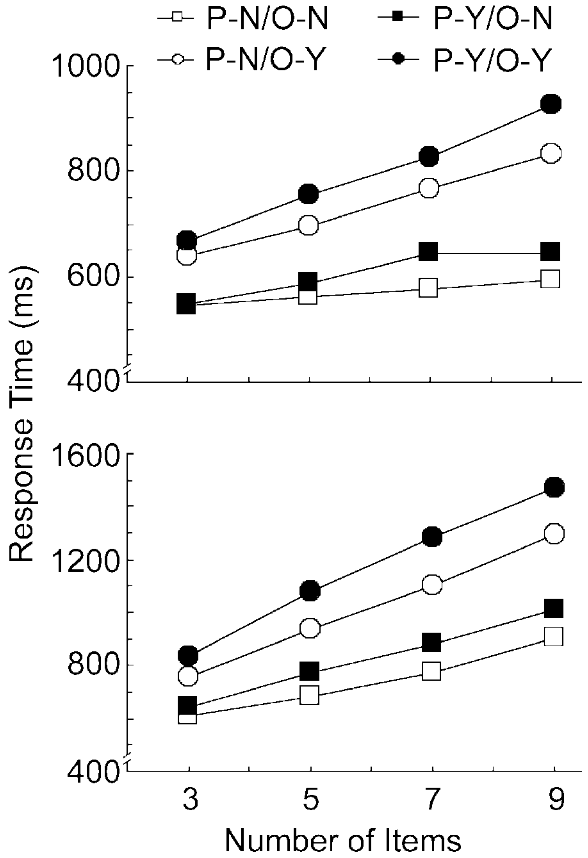

**Fig. 2.** Response time on correct trials as a function of the number of items displayed on target-present (top panel) and target-absent (bottom panel) trials. Results are shown separately for the four combinations of part and spatial-organization sharing: P-N = target and distractors did not share parts; P-Y = target and distractors did share parts; O-N = target and distractors did not share spatial organization; O-Y = target and distractors did share spatial organization.

> effect of spatial organization sharing = slope when only spatial organization was shared minus slope when neither parts nor spatial organization was shared

> joint effect of part and organization sharing = slope when both parts and spatial organization were shared minus slope when neither parts nor spatial organization was shared

With the assumption that visual search was serial and self-terminating, the processing time per item was estimated as $[(2*p)+n]/2$, where $p$ and $n$ are the slopes on target-present and target-absent trials, respectively. A comparison of the two right-most columns in Table 2 shows that the actual impact of the target sharing both its parts and spatial organization with the distractors was precisely predicted by the addition of the separate effects of these factors.

An analysis of variance conducted on correct RTs confirmed this analysis. All main effects were highly significant. RTs were longer on target-absent than on target-present trials, $F(1, 11) = 83.2, p < .001, \eta^2 = .88$; they were also longer when the target and distractor shared parts, $F(1, 11) = 46.3, p < .001, \eta^2 = .81$, or spatial organization, $F(1, 11) = 66.6, p < .001, \eta^2 = .86$. The main effect of number of items, $F(3, 33) = 86.0, p < .001, \eta^2 = .89$, reflects the linear increase

**TABLE 1**

*Outcome of the Linear Regression Analyses of Response Times on Correct Trials as a Function of the Number of Items Displayed*

| Condition | | | | |
|---|---|---|---|---|
| Part sharing | Spatial-organization sharing | Intercept | Slope | $R^2$ |
| | Target-present trials | | | |
| No | No | 522 | 8.0 | 1.0 |
| Yes | No | 501 | 17.5 | .91 |
| No | Yes | 537 | 32.6 | 1.0 |
| Yes | Yes | 540 | 42.4 | 1.0 |
| | Target-absent trials | | | |
| No | No | 448 | 49.0 | .98 |
| Yes | No | 466 | 60.1 | 1.0 |
| No | Yes | 489 | 89.1 | 1.0 |
| Yes | Yes | 536 | 105.3 | 1.0 |

**Note.** Intercepts are measured in milliseconds and slopes in milliseconds per item.

of correct RTs with increasing number of stimuli displayed (see Table 1 and Fig. 2). Apart from the important exception of the Part × Spatial Organization interaction, which was not significant, $F(1, 11) = 3.7$, every other two-way interaction was significant. The two-way interactions involving the factor of target presence revealed that the other factors had larger effects on target-absent than on target-present trials: number of items, $F(3, 33) = 46.0$, $p < .001$, $\eta^2 = .81$; part sharing, $F(1, 11) = 15.0$, $p < .005$, $\eta^2 = .58$; spatial-organization sharing, $F(1, 11) = 44.2$, $p < .001$, $\eta^2 = .80$ (see Fig. 2 and Table 1). The Part Sharing × Number of Items interaction, $F(3, 33) = 7.7$, $p < .001$, $\eta^2 = .41$, and Spatial-Organization Sharing × Number of Items interaction, $F(3, 33) = 39.4$, $p < .001$, $\eta^2 = .78$, indicate that the sharing of parts or spatial organization between the target and distractors led to increased slopes of RTs as a function of number of items (see Tables 1 and 2). The only other significant interaction was the Target Presence × Spatial-Organization Sharing × Number of Items interaction, $F(3, 33) = 10.8$, $p < .001$, $\eta^2 = .50$. This interaction reflects the fact that the effect of spatial-organization sharing on the slope of RTs as a function of number of items was magnified on target-absent trials relative to target-present trials (see Table 2).

**TABLE 2**

*Separate and Combined Effects of Part and Spatial-Organization Sharing*

| | Factor | | | |
|---|---|---|---|---|
| Measure of effect | Sharing parts | Sharing spatial organization | Actual combined effect | Predicted combined effect |
| Target-present slope (ms/item) | 9.5 | 24.7 | 34.4 | 34.2 |
| Target-absent slope (ms/item) | 11.1 | 40.1 | 56.3 | 51.2 |
| Estimated ms/item | 15.1 | 44.7 | 62.5 | 59.8 |

## DISCUSSION

The sharing of parts and the sharing of spatial organization between the target and distractors both affected visual search performance. Most important, these effects were strictly additive. Part and spatial-organization sharing both resulted in increased RTs, as well as in increased slopes of RTs as a function of the number of items displayed. These findings have important implications with respect to the processes involved in the perception of complex visual objects in the context of speeded visual tasks such as visual search.

The cost of part sharing supports the hypothesis that complex shapes are parsed into their component parts. An alternative explanation of the part-sharing effect in terms of the global shape properties that define the component parts cannot hold because all conditions were matched in this respect. Whether or not they shared their constituent parts, all target-distractor sets were made of objects that instantiated every possible value of the global dimensions used to construct the parts, and, therefore, no target possessed a unique shape property not found in its distractors. Similarly, an account of the part-sharing effect based on global representations of entire objects or on local feature information (e.g., vertices and edges) is highly unlikely because the manipulations appear to be orthogonal to what these theories assume to be psychologically important shape properties.

Another theory congruent with the part-sharing effect proposes that complex objects are represented as collections of visible surfaces, instead of as collections of parts (Lee & Park, 2002; Leek & Arguin, 2000; Leek et al., 2002; Pentland, 1989). The relevance of this theory can be seen in Figure 1b, which shows that the visible surfaces of distractors largely overlapped with those of the target when these items shared parts. When parts were not shared, however, the target had no visible surfaces in common with the distractors. The present findings do not adjudicate between part-based and surface-based theories of shape perception, and future studies should attempt to address this issue.

Whether the correct account of the part-sharing effect involves part-based or surface-based representations, these representations should be considered relatively resistant to rotations in depth. The parts or surfaces the target and distractors shared were typically shown from markedly different viewpoints. The occurrence of a part-sharing effect under these circumstances indicates that the visual system spontaneously, and perhaps automatically, generalizes part representations across viewpoints (see also Stankiewicz, 2002). This finding may appear incongruent with previous observations from same/different matching tasks indicating that the visual representation of even simple shapes, like those in Figure 1, is completely viewpoint-specific (i.e., depth-rotation cost on same-shape trials; Tarr, Williams, Hayward, & Gauthier, 1998; see also Hayward & Williams, 2000, for a relevant discussion). This apparent contradiction is resolved by noting that a cost in matching a pair of differently oriented shapes does not mean that it should be impossible to make the match. In fact, the typical finding is that shapes from different views are matched accurately. The part-sharing effect documented here indicates that the visual system can establish the correspondence between differently oriented instances of the same shape. It is likely, however, that such a correspondence is performed less effectively than the correspondence of matching shapes that have the same orientation.

Another important finding of the present experiment is that the sharing of spatial organization between the target and distractors was associated with a large performance cost, even when the objects were made of different parts. This clearly demonstrates that the spatial structure of the constituent higher-order units of a complex shape is a determining aspect of its representation.

Finally, the strict additivity of the effects of part and spatial-organization sharing is a crucial observation. This finding indicates that independent processes mediate the perception of these two types of information and that there is no cross talk between them. Obviously, this means that neither process is contingent upon the other. The additivity of the part- and organization-sharing effects also appears incompatible with shape representations that are exclusively based on either global properties of the entire object or local contour information (Cortese & Dyre, 1996; Edelman & Weinshall, 1991; Lades et al., 1993; Poggio & Edelman, 1990; Ullman, 1989). Indeed, these two views fail to dissociate between the representation of the higher-order constituents of a shape and the shape's spatial organization. Consequently, such theories would assume that the factors of part and organization sharing tap common processes and thus that their effects should interact with one another, which is clearly not the case.

The additivity of the part- and organization-sharing effects appears incompatible with a functional architecture whereby the parts and spatial organization of an object are processed sequentially. Indeed, this factor additivity implies that neither process depends on the output of the other, and a functional architecture made of two parallel streams, one concerned with parts and the other with spatial organization, would appear more probable. We should point out, however, that the present observations are not conclusive in this respect and that further investigations will be required.

In conclusion, additive effects of part sharing and of spatial-organization sharing between the target and distractors were observed in a visual search task involving complex synthetic visual objects. These observations indicate that, in visual search, the perception of complex shapes involves one process concerned with the properties of their higher-order constituents (parts or surfaces) and another process concerned with the spatial organization of these units.

*Acknowledgments*—The technical help of Christine Lefebvre in carrying out the present research is acknowledged. We thank E. Charles Leek for stimulating discussions regarding the present findings. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Martin Arguin. Martin Arguin is chercheur-boursier of the Fonds de la Recherche en Santé du Québec.

## REFERENCES

Arguin, M., Bub, D., & Dudek, G. (1996). Shape integration for visual object recognition and its implication in category-specific visual agnosia. *Visual Cognition*, *3*, 221–275.

Arguin, M., & Leek, E.C. (2003). Orientation-invariance in visual object priming depends on prime-target asynchrony. *Perception & Psychophysics*, *65*, 469–477.

Arguin, M., & Saumier, D. (2000). Conjunction and linear non-separability effects in visual shape encoding. *Vision Research*, *40*, 3099–3115.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*, 65–96.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 1–38.

Cortese, J.M., & Dyre, B.P. (1996). Perceptual similarity of shapes generated from Fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 133–143.

Donnelly, N., Humphreys, G.W., & Riddoch, M.J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 561–570.

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

Hayward, W.G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science*, *11*, 7–12.

Hoffman, D.D., & Richards, W.A. (1984). Parts of recognition. *Cognition*, *18*, 65–96.

Hoffman, D.D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *63*, 29–78.

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

Kosslyn, S.M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Kurbat, M.A. (1994). Structural description theories: Is RBC/JIM a general-purpose theory of human entry-level object recognition? *Perception*, *23*, 1339–1368.

Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Marlsburg, C., Wurtz, R.P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, *42*, 300–311.

Lee, Y.-L., & Park, R.-H. (2002). A surface-based approach to 3-D object recognition using a mean field annealing neural network. *Pattern Recognition*, *35*, 299–316.

Leek, E.C., & Arguin, M. (2000). Surface-based shape representation in object recognition. *Abstracts of the Psychonomic Society*, *5*, 87.

Leek, E.C., Reppa, I., & Arguin, M. (2002). Surface-based shape representations mediate object recognition: Further evidence from part-whole priming. *Abstracts of the Psychonomic Society*, *7*, 87.

Marr, D. (1982). *Vision*. New York: Freeman.

Marr, D., & Nishihara, H.K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society of London*, *207*, 187–217.

Palmer, S.E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*, 441–474.

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

Pentland, A. (1989). Shape information from shading: A theory about human perception. *Spatial Vision*, *4*, 165–182.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3-dimensional objects. *Nature*, *343*, 263–266.

Saiki, J., & Hummel, J.E. (1998a). Connectedness and part-relation integration in shape category learning. *Memory & Cognition*, *26*, 1138–1156.

Saiki, J., & Hummel, J.E. (1998b). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 227–251.

Saumier, D., & Arguin, M. (2003). Distinct mechanisms account for the linear non-separability and conjunction effects in visual shape encoding. *Quarterly Journal of Experimental Psychology*, *56A*, 1373–1388.

Singh, M., Seyranian, G.D., & Hoffman, D.D. (1999). Parsing silhouettes: The short-cut rule. *Perception & Psychophysics*, *61*, 636–660.

Stankiewicz, B.J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 913–932.

Tarr, M.J., & Bülthoff, H.H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*, 1–20.

Tarr, M.J., Williams, P., Hayward, W.G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, *1*, 275–277.

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

(RECEIVED 7/10/03; REVISION ACCEPTED 9/16/03)