On the representation of object structure in human vision: evidence from differential priming of shape and location

Shimon Edelman
School of Cognitive and Computing Sciences
University of Sussex at Brighton, Falmer BN1 9QH, UK
shimone@cogs.susx.ac.uk

Fiona Newell
Department of Psychology
University of Durham, South Road, Durham DH1 3LE, UK
Fiona.Newell@durham.ac.uk

November 27, 1998

Abstract

Theories of object representation can be classified as structural, holistic or hybrid, depending on their approach to the mereology and compositionality of shapes. We tested the predictions of some of the current theories in three experiments, by quantifying the effects of various priming cues on response times to 3D objects. In experiment 1, there were two possible locations for the stimulus components: left-right and top-bottom. The prime could be identical to the stimulus, identical in location but with different parts, identical in the complement of differently located parts, or altogether different. Both location and part identity effects were significant. In experiment 2 we added a part-neutral (empty frame) prime condition; the effect of location, but not of part, remained significant. In experiment 3, which included an additional location-neutral prime condition, only the location effect, again, was significant. These findings are not entirely compatible either with the structural description theories of representation (which predict priming by "disembodied" parts or geons) or with the holistic theories (which do not predict priming by "shapeless" location on its own). They may be interpreted in terms of a hybrid theory, according to which conjunctions of shape and location are explicitly represented, and therefore amenable to priming.

1   The Problem of Representation

The nature of the memory trace left by perceived objects in the human visual system is a fascinating problem, whose solution would lead both to a better understanding of vision in the brain, and to the development of better artificial visual systems. From a computational standpoint, the Problem of Representation of object shapes has several distinct aspects. For example, any computational model of object recognition must explain how to represent objects internally in such a manner that the variability of their appearance caused by changing viewing conditions (such as viewpoint or illumination) will not disrupt recognition (Ullman, 1996). In theorizing about human vision, this consideration led to the emergence of two classes of models. On the one hand, there are models that postulate essentially viewpoint-invariant representations, claiming that human performance in recognition does not, by and large, depend on viewpoint (Biederman, 1987; Biederman and Gerhardstein, 1993; Biederman and Gerhardstein, 1995). On the other hand, there are models that

posit viewpoint-dependent representations, motivated by the increasingly extensive psychophysical evidence in favor of viewpoint-dependent performance in a variety of cases (Bülthoff et al., 1995; Newell and Findlay, 1997; Newell, 1998; Jolicoeur and Humphrey, 1998).

In the present psychophysical study, we examine another issue concerning representation: how is object structure — in particular, familiar shapes in new configurations — represented and processed by the human visual system? Our decision to consider the problem of novel objects rather than new views is motivated by two considerations. First, due to the recent advances in the theory of recognition (Ullman, 1996), the computational problem of compensating for viewpoint-related changes seems now tractable. Dealing with new shapes (rather than new views of familiar shapes) is, therefore, the next challenge to be taken on now (Edelman, 1997). Second, comparing the predictions of current models of recognition with observer performance on novel objects should help us distinguish between the various theories, including those that vie for offering the best model of recognition across viewpoints. Thus, the results of this study can be fed back into the main theoretical debate about the Problem of Representation.

2 The representation of novel objects — candidate models

The difficulties facing any attempt to settle major issues concerning cognitive representations empirically have been highlighted repeatedly in the past; see, e.g., (Anderson, 1978; Barsalou, 1990). In view of the caveats mentioned by these authors (e.g., the inseparability of the effects of representation and of processing), we attempt to distinguish among concrete, algorithm- or mechanism-level models, rather than among abstract, computational-level theories. To that end, we proceed to formulate the models that are to be compared, filling in the algorithmic details where these are not available in the original formulation of the model in the literature.

2.1 The Standard Structural Model (SSM)

The first family of models we consider is based on the notion of structural decomposition: object shapes are described in terms of relatively few generic components, joined by spatial relationships chosen from an equally small fixed set. The representation of novel objects is made possible through the standardization of the primitives (components and their relationships): if these are sufficiently varied, a great many shapes can be described, just as tens of thousands of spoken words can be generated using a small number of phonemes as components (Biederman, 1987).

A typical structural theory, Biederman's (1987) Recognition By Components (RBC),[1] postulates a set of 30 or so primitive shapes (geons), claimed to be easily detected in images due to their *nonaccidental* properties. The latter are 3D features that are almost always (that is, barring an accident of viewpoint) preserved by the imaging (projection) process (Lowe and Binford, 1985). A representative example of such a feature is a pair of parallel lines; because a chance image alignment of two segments that are in fact not parallel in 3D is unlikely (Richards and Jepson, 1992), two parallel lines in the image are a good indicator of the presence of a 3D geon such as a cylinder "out there" in the scene.

To be able to deal with novel objects, a model based on structural descriptions must form the representation of the whole in terms of its parts dynamically ("on the fly"), for each shape it encounters. The implementation of the RBC theory described by Hummel and Biederman (1992) is an example of a model that binds the parts to each other dynamically. It is important to note that this implementation includes special relational units dedicated to the binding operation, over

---

[1]This, in fact, is a variant of a structural decomposition theory usually attributed to (Marr and Nishihara, 1978) and popularized as a psychological model by Biederman.
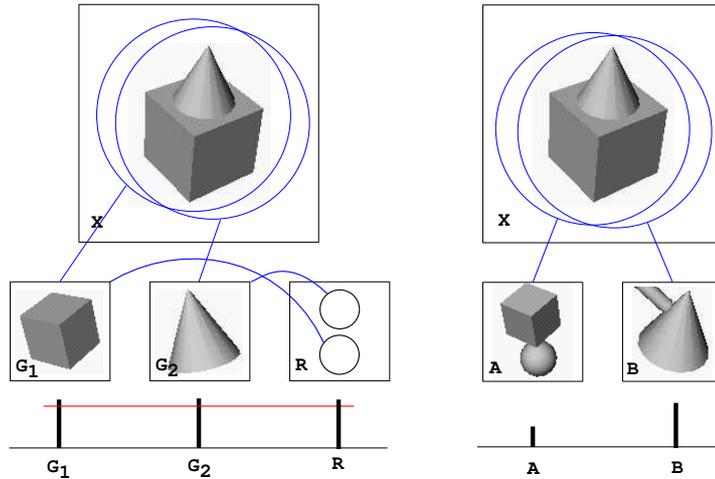
Figure 1: An illustration of the conceptual difference between part-based and holistic approaches to the representation of structure. *Left:* Part-based models such as SSM (section 2.1) would describe this object as a cone on top of a block; here, modules G$_1$ and G$_2$ tuned to the shape of the parts signal their presence (in an all-or-none fashion), from which the structure of the entire object is determined. Separate spatial-relations modules (here, R, charged with the representation of the "on top" relationship) are a crucial component of SSM-like models. See section 2.1. *Right:* holistic models such as Chorus (section 2.2) describe the entire object in terms of its similarity to entire reference shapes (here, A and B). In Chorus, the relative activation levels of graded-response modules convey information about the structure of the object. See section 2.2.

and above the shape units dedicated to each of the geons; see Figure 1, left. An explanation of the role of the relational units can be found in (Hummel and Biederman, 1990), p.619:

> "In [layer] L4, relations are computed separately for each value of each dimension. The relation *below* will be used to illustrate how the L4 cells operate, but the logic generalizes to all relations. [...] Associated with every position in Y (Y$_p$), there is an L3 cell which becomes active when that position is occupied by a geon (L3$_{y=p}$), and there are two L4 cells: one that becomes active when Y$_p$ us below another occupied position (L4$_{below\ at\ y=p}$), and one that becomes active when Y$_p$ is above another occupied position (L4$_{above\ at\ y=p}$)."

A more recent model of this kind, developed to account for the structural alignment performed by subjects in cognitive judgment tasks, is described in (Hummel and Holyoak, 1997). Here too, there are special units devoted to representing the relations: "the predicate unit *loves1* represents the first (agent) role of the predicate "loves" and has bidirectional excitatory connections to all of the semantic units representing that role [...]" (Hummel and Holyoak, 1997, p.435).

Another model of the structural variety is Barsalou's Perceptual Symbol System (PSS), outlined in (Barsalou, 1998).[2] In this model, object structure is represented by spatial templates which possess slots acting as variables. Shape primitives (e.g., object parts) are bound to these variables, resulting in a data structure known in cognitive science as a frame (Minsky, 1975). Because the slots in a frame can be occupied by different shapes, the PSS model is capable of representing an open-ended variety of objects. The representational structure of PSS is postulated to be at least three levels deep. On the first level are the representations of the objects and the locations. On

---

[2]We are grateful to L. Barsalou for clarifying to us some of the details of his theory in a personal communication.
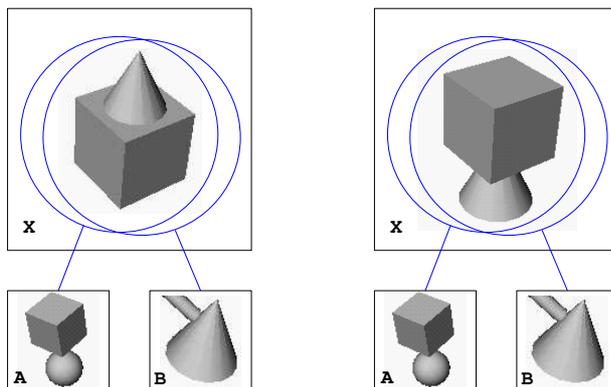
Figure 2: The holistic approach is limited in its ability to make explicit the similarities and the differences between complex objects which human observers would describe as composed of similar parts arranged in different configurations. Consider, for example, the two objects shown here: a cone on top of a cube (*left*) and a cube on top of a cone (*right*). On the one hand, these objects are clearly different and Chorus would indeed easily label them as such. On the other hand, the objects do share some rather conspicuous features, a fact that needs to be represented explicitly in any system that aims at mimicking human competence in visual shape analysis. This example is based on J. Hummel's (1998) argument against holistic models of representation.

the second level are the mappings from the object representations to associative areas (required for controlled activation of object frames, a key feature of PSS), and from the spatial representations to their associative areas. Finally, the third level coordinates the two two-level structures for the objects and space into a complex spatial configuration of objects.

We shall regard models that postulate separate shape and relational units as varieties of the Standard Structural Model (SSM). The activation level of such units over time is, in principle, amenable to manipulation, an observation that can be used to formulate predictions concerning human performance in *priming* experiments. Priming is defined as a modification of performance that (i) stems from exposure to a stimulus, and (ii) persists over time and manifests itself when the subject subsequently encounters similar stimuli (Tulving and Schacter, 1990; Ochsner et al., 1994). In the context of SSMs, priming by two kinds of stimulus characteristics is expected:

1. *shape:* the shape units should respond to their preferred stimuli (geons) irrespective of their location in the image, leading to shape-based priming that is insensitive to the location of the shape;

2. *relative location:* the relational units should give rise to location-based priming in which the relative position of object parts, but not the shapes of the parts, matter.

2.2   The Chorus of Prototypes

Categorization of novel shapes, considered until recently to be the prerogative of structural models, can, in fact, be carried out by a holistic mechanism that does not treat various parts of the same object separately. The main idea here is to represent objects by their similarity[3] to a collection of reference shapes or prototypes (Figure 1, right), which, in turn, are represented by stored chosen views (Ullman and Basri, 1991; Poggio and Edelman, 1990). A model based on this idea (Edelman

---

[3]Although the concept of similarity is often regarded as dangerously vague, it can be very useful if properly treated (Medin et al., 1993). In the model referred to in this section (Edelman and Duvdevani-Bar, 1997), similarity is given a concrete computational definition that does not leave room for vagueness.

and Duvdevani-Bar, 1997; Edelman, 1998) contains a number of reference-shape detection modules, each of which computes the similarity of its preferred shape to the input. The resulting vector of similarities serves as a low-dimensional representation of the input, which is not structural but holistic, because it is based ultimately on the stored views ("snapshots") of the reference objects.

The greatest challenge to holistic models seems to lie in capturing the compositional aspects (Bienenstock and Geman, 1995; Bienenstock et al., 1997) of object representation in human vision. As illustrated in Figure 2, if the structure of parts comprising an object is not made explicit, the model will lack certain features of the human competence in the domain of object perception, such as judging the similarity of composition (as opposed to the similarity of the global shape).

The need to treat object structure explicitly requires relaxing the holistic outlook of Chorus. This can be done without compromising the positive features of this model, such as its computational feasibility (Edelman and Duvdevani-Bar, 1997), by following two general principles: (1) the parts should be defined in an image-based, not object-centered, frame, to alleviate the binding problem (2) the parts should be specific, not generic (geons), to facilitate learning from examples. A model based on the Chorus scheme and on these principles is outlined next.
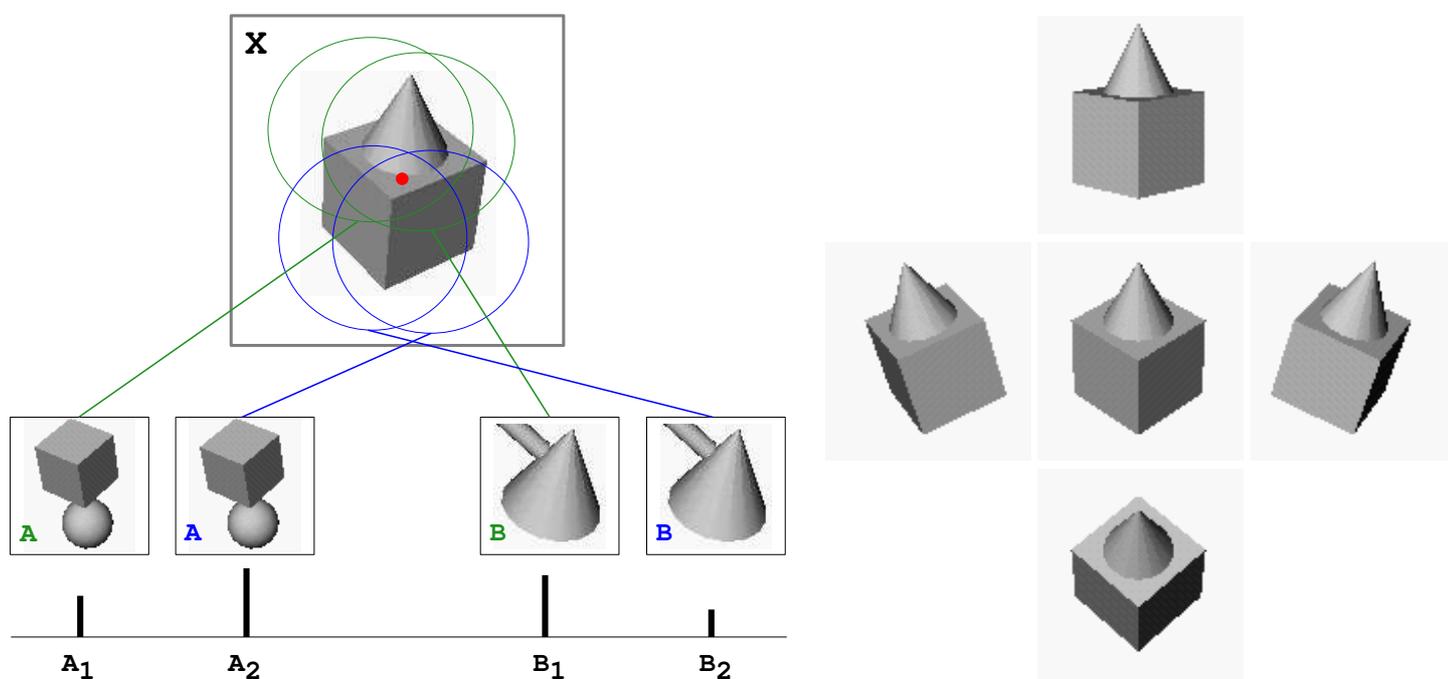
Figure 3: It may be possible to circumvent the problem illustrated in Figure 2 using modules tuned to image fragments, in conjunction with binding by retinotopy. *Left:* In such a scheme, which may be called the Chorus of Fragments (CoF), each object-specific module would come in several varieties, distinguished by the location of the module's receptive field relative to the fixation point (indicated by the thick dot). Here, module A_1 responds optimally when the fixation is above and slightly to the left of a stimulus resembling object A. Likewise, module A_2 prefers the object to be below the fixation point. As in the Chorus of prototypes, a new object X is represented by the pattern of activities across object-specific modules. *Right:* because different aspects (Koenderink and van Doorn, 1979) of solid shapes need to be treated separately in any case, view-specific image-based fragments can be used instead of object-centered parts (which are difficult to detect reliably). The original Chorus scheme can be easily adapted for this purpose; its operation is illustrated in Figure 4.
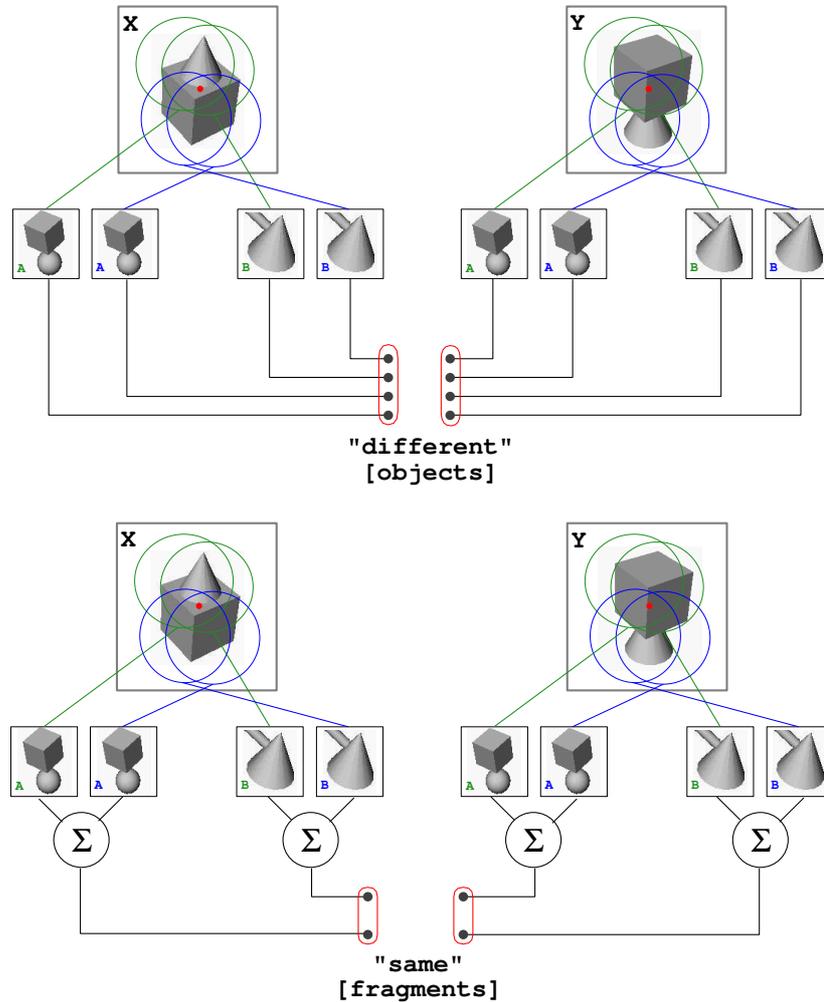
Figure 4: If the receptive fields of Chorus modules are confined to retinally-defined fragments of the entire image, their activities can be made to carry additional information concerning the structure of the stimulus, without recourse either to generic parts, or to any kind of binding mechanisms (beyond co-activation and retinotopy). *Top:* a Chorus system (shown here with duplicated modules differing in their receptive field locations) can easily discriminate between objects X and Y, which consist of the same parts in different configurations. *Bottom:* if the responses of each group of modules tuned to the same shape are pooled, the resulting representation will reflect the fragment-level similarity between objects X and Y (note the relationship between this idea and the feature histogram methods for object recognition (Schiele and Crowley, 1996; Mel, 1997). Top-down connections can then be used to trace back the source of the different contributions to the pooled response, which would amount to the ability to address and describe separately the various retinotopically defined fragments of the stimulus.

2.3   The Chorus of Fragments (CoF)

The Chorus of Fragments (CoF) model uses prototypical shapes as "parts" that are spatially anchored (i.e., are actually image fragments) rather than floating or holistic. This is necessary to avoid the need for temporal binding of parts — a traditional handicap of the structural approaches (von der Malsburg, 1995; Kirschfeld, 1995).

Instead of temporal binding, CoF uses binding by retinotopy (Edelman, 1994). The line of reasoning leading to this idea is illustrated in Figures 1 through 4. In this approach, structure is represented explicitly, but in an image-based rather than object-centered manner. From the standpoint of functionality required by the structural approach, keeping representations image-based (as they are in the original version of Chorus) is not in itself problematic. In particular, although image-based structure is aspect-specific, so is a full-blown structural description (Biederman and Gerhardstein, 1993), which in any case must be extracted anew for each distinct aspect of the object. Computationally, however, image-based structure is more tractable, especially if the primitives in terms of which structure is represented are encoded by Chorus-like modules. The only modification required for that purpose in the original (holistic) Chorus scheme is control over the location and the size of the retinal receptive field of each module. We conjecture that this can be done in a hard-wired fashion, as depicted in Figure 3, turning the Chorus of prototypes into a Chorus of Fragments.

We can now formulate the predictions of the CoF model with respect to the nature of priming one should expect. Note that in CoF the representations of shape and of retinal location are inextricably interwoven, so that a spatial predicate such as "above" is only represented as the disjunction over the activities of all object-specific modules that "look" at the upper visual field (even then this predicate means "above fixation" and takes one argument, not two). Consequently, priming is expected for location, but not necessarily for translation-invariant spatial relations. Moreover, the priming for shapes is expected to be stronger when the shape appears in the same retinal location in the two trials. These predictions can be contrasted with those of SSM, which predicts priming both for "shape-free" spatial relations and for spatially "floating" geons.

3   Questions for psychophysics

The preceding discussion suggests a number of concrete issues that can be addressed experimentally within the priming paradigm.

3.1   Q0: Are the shape primitives generic or specific?

Remarks. SSM postulates generic hierarchically structured primitives, namely, geons built of lines and curves. CoF postulates primitives that are composed, ultimately, of snapshots of objects familiar to the system. At present, it is not clear to us how to address question Q0 psychophysically; attempts to do so in the past did not yield clear results (Intrator et al., 1995). Methods that complement psychophysics, e.g., electrophysiology, may be more suitable for this purpose. In this connection, one should note the results obtained by the "stimulus reduction" technique introduced by K. Tanaka in a series of single-cell studies of the inferotemporal cortex in the monkey (Tanaka et al., 1991; Tanaka, 1992). It may so happen, however, that the set of intermediate-level features used to code objects is too dependent on the system's prior experience (Kobatake et al., 1998), making generalization of the kind implied in question Q0 risky. Indeed, results of some computational studies indicate that recognition can be based on features that, if examined in isolation, would seem meaningless (Mel, 1997; Amit and Geman, 1997). In view of these considerations, we believe that Q0 is better left aside for the moment.

3.2 Q1: Are the shape primitives represented independently of specific retinotopic locations?

Remarks. According to SSM, the presence of a geon is signaled by the activity of a unit dedicated to that geon, no matter where in the visual field it appears; the binding of a geon to a location is carried out dynamically, by a separate mechanism (a relational unit). In comparison, according to CoF, shape-specific modules have well-defined receptive fields each of which allows rough localization of the stimulus to which the module responds; precise localization is afforded by the ensemble response of a number of modules tuned to the same object.

3.3 Q2: Are the spatial relations represented independently of the shape primitives?

Remarks. In SSM, relations such as "above" are represented explicitly (i.e., by dedicated relational units). In CoF, there is no explicit representation of relations at all; a novel situation — say, "A above B" — would be represented by the conjunction "A in the upper visual field and B in the lower visual field".

3.4 Q3: Are the spatial relations represented independently of the retinotopic locations?

Remarks. Spatial relations can be independent of shape primitives, yet dependent on the location in the visual field, as long as the shape primitives are generic and are not tied to a particular location. This is precisely what SSM predicts. Alternatively, the representation of "A above B" in the upper left quadrant of the visual field may have nothing in common with the representation of "C above D" in the lower right quadrant (as it is the case in CoF).
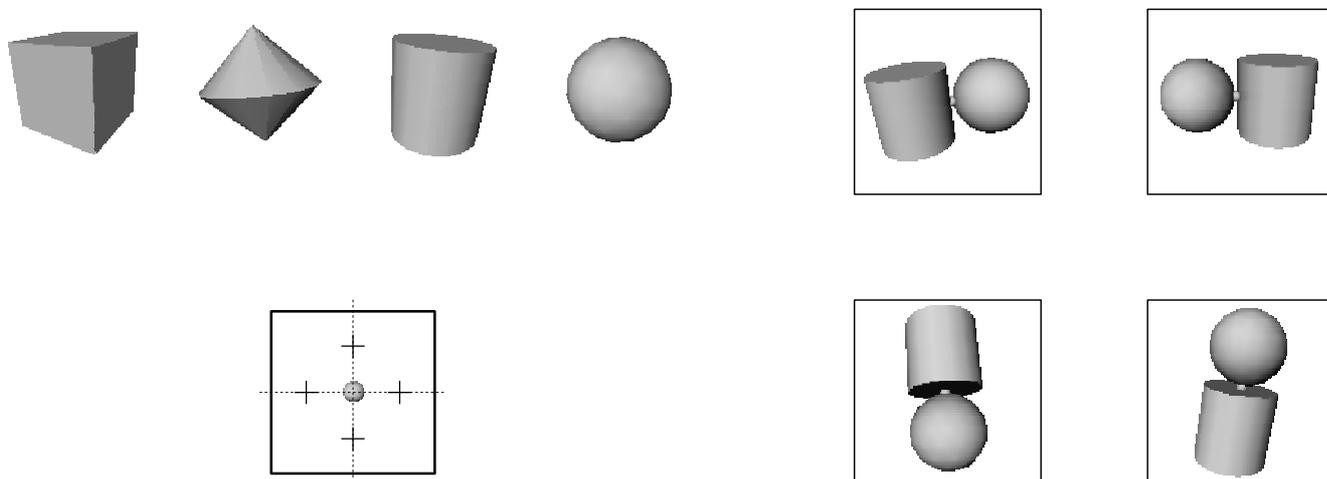
Figure 5: *Left, top:* the four basic parts, Cube, Top, Cylinder, Sphere, used in generating the stimuli for the experiments. *Left, bottom:* each part can be located in one of four possible places relative to the fixation point (designated by the small sphere). For two-part objects, constrained to consist of distinct parts, this arrangement results in $3 \times 4 = 12$ different shapes. *Right:* The four possible objects, composed of two of the parts (cylinder and sphere). Objects such as these served as stimuli for the 4-alternative forced-choice (4AFC) categorization task used in our experiments.

4 The experiments

We addressed questions Q1 through Q3 in a study centered around the priming paradigm, leaving question Q0 for future research. As we argued above, repetition priming (Tulving and Schacter, 1990; Ochsner et al., 1994) provides a convenient route for studying the nature of memory representations of objects. With everyday objects, repetition priming has been shown to depend both on semantic relatedness of the prime and the target, and on their visual similarity (Bartram, 1974). For novel objects, such as those used in the present experiments (see below), object similarity, which is of direct interest to the study of visual representation, is likely to preponderate.
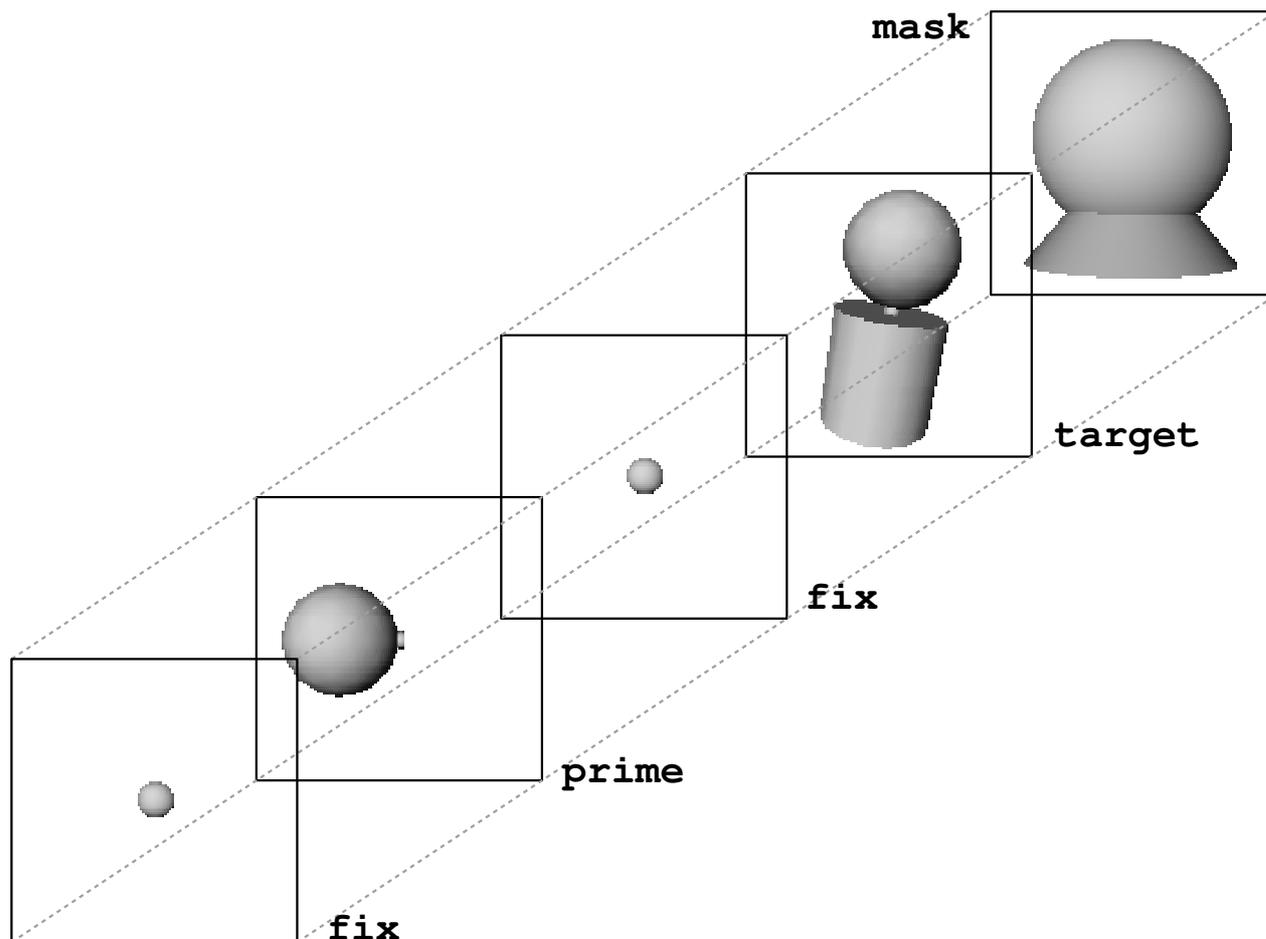
Figure 6: The order of events in an experimental trial. Each trial begins with a brief presentation of the fixation aid (a small sphere, shown for 400 *ms*). Subsequently, the prime and the fixation are shown (for 100 and 300 *ms*, respectively), followed by the target object, which is displayed until the subject responds. After the response is made, the mask appears for 700 *ms*. Each of the objects displayed on the screen undergoes a precession around two axes in depth (i.e., it is seen as "wobbling"); the amplitude of the precession is set to $20°$, so as to impart to the subject a perception of depth through the structure from motion mechanism, while keeping each of the constituent parts roughly in the same retinotopic location.

4.1 Method

The stimuli were two-part objects composed of four basic constituents, as described in Figure 5, left. Those components are, in fact, qualitatively distinct (in Biederman's sense), and are, therefore, referred to as "geons." In contradistinction to fragments (see section 2.3), which are defined both by their apparent shape and by their retinotopic location, geons are "disembodied" (that is, not associated with a particular location).[4] Three distinct families of four shapes each were used in each of the three experiments described below.

The experiments were designed to compare the relative strength of priming for two kinds of structural similarity between the prime and the target stimuli. One of these had to do with location similarity (the variable we call Loc). Thus, in a trial in which Loc=*same* the prime and the target had possibly different parts, located in the same spots with respect to fixation. The other kind of similarity had to do with shape (the variable Geo). In a trial in which Geo=*same*, the prime and the target had the same constituent shapes, possibly located differently with respect to fixation. The full spectrum of prime/target relationships in each of the experiments is illustrated in Figures 7 through 9, top panels.

The event sequence of each trial, which consisted of priming and target stimuli shown in succession, is depicted and explained in Figure 6. The objects were rendered in real time, using the Lambertian shading model, by a graphics workstation (Silicon Graphics Inc., O2). They were displayed in a 256×256 window, within which they subtended a visual angle of approximately $3°$. The objects were always displayed as "wobbling" — undergoing precession around the vertical axis, with an amplitude of $20°$ and period of 500 ms. This mode of display enhanced the 3D appearance of the stimuli, without changing qualitatively their image-plane orientation (e.g., a vertically oriented object remained approximately vertical throughout its precession cycle, etc.).

The subjects were first taught to carry out a 4-alternative forced-choice (4AFC) classification of a family of shapes (Figure 5, right). The response was made by pressing one of the four designated buttons (1, 2, 3 or "enter") on the numeric keypad of the computer keyboard. The response time was recorded using a sub-millisecond-precision routine that accessed a hardware timer available in the SGI workstation (the code for this operation is available from the first author). When their performance reached 90% correct in the trailing 30 trials, the subjects progressed to the test phase, and were shown a series of prime/target pairs from the same family of objects. The number of trials in the test phase varied between 296 in experiment 1 and 92 in experiment 3. The order of the trials was randomized for each subject.

We considered the response time (RT) as the dependent variable, and used analysis of variance (SAS procedure GLM) to examine its dependence on the Loc and Geo variable.[5] Trials in which an incorrect response had been made, or in which the RTs were shorter than 250 ms or longer than 2000 ms were discarded from further analysis. The proportions of such trials were 7.8%, 12.5% and 8.9% in experiments 1 through 3, respectively.

4.2 Experiment 1

In experiment 1, the geons (Geo) in the prime and the target could be the *same* or *different* (see Figure 7, top). Likewise, the location (Loc) of the geons could be either the *same* or *orthogonal*. Note that the four object classes had been defined in such a manner that the alignment of the parts in two of them was orthogonal to that in the other two classes. Hence, in the Loc=*orthogonal*

---

[4]Using the terminology of (Treisman, 1992), geons are, therefore, *types* of shapes, while fragments are the *tokens* that instantiate these types.

[5]Instead of RT, one can look at the correct classification rate. For that, however, the stimulus presentation time would have to be very short, to drive the performance below ceiling. We felt that this choice would have made our results more difficult to generalize to a normal object processing setting.
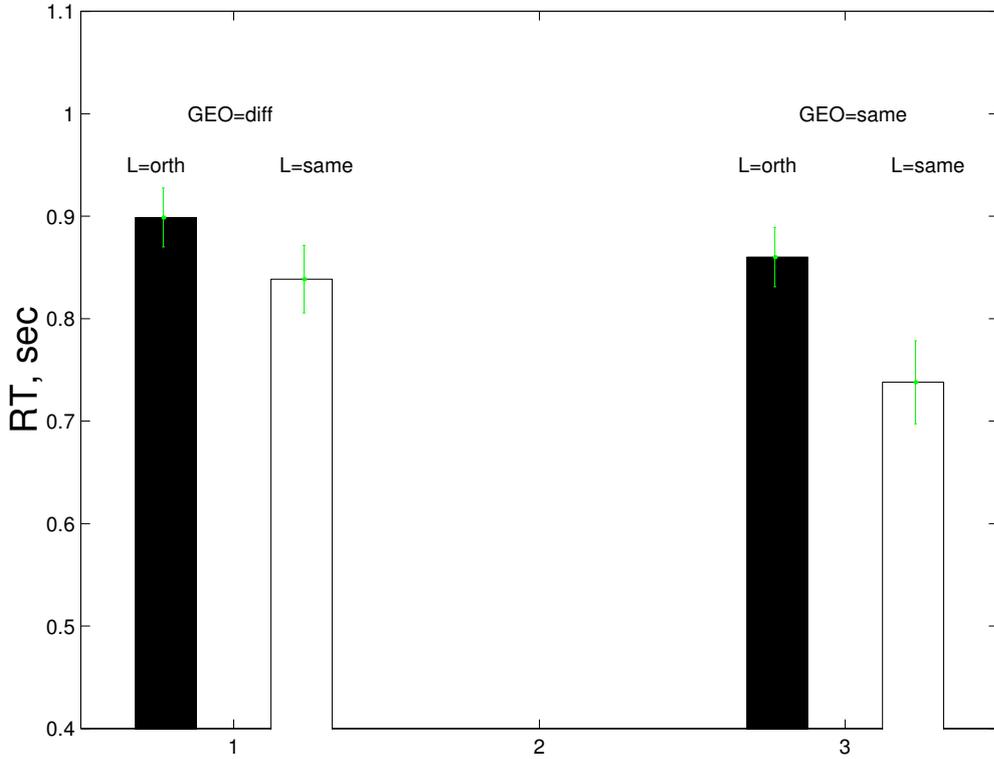
Figure 7: Experiment 1. *Top:* the four priming conditions — identical fragments (i.e., same parts in same places); identical complement of geons (but not locations); identical image locations (but not geons); both geons and locations different. *Bottom:* the response times. The error bars, showing ± 1 standard error of the mean, were computed after the inter-subject variability has been taken out, by transforming each observation according to the formula $y' = y - \bar{y}_s + \bar{y}_G$, where $\bar{y}_s$ is the subject mean, and $\bar{y}_G$ the grand mean (Loftus and Mason, 1994).

priming condition, the location of the prime necessarily overlapped that of another target category. This constraint was removed in the later experiments. To avoid confounding the effects of LOC and GEO, we excluded priming stimuli in which the parts were the same as in the target but their relative location was inverted (i.e., top-bottom instead of bottom-top, or left-right instead of right-left).

4.2.1 Results

Four subjects participated in this experiment. The mean RT was 833 *ms*; the breakdown of RT by priming condition is plotted in Figure 7, bottom. Changing location from *orthogonal* to *same* (corresponding to the two levels of the LOC variable) resulted in priming (that is, reduction of RT) of 91 *ms*. Likewise, changing part shapes from *different* to *same* (corresponding to the two levels of the GEO variable) resulted in a priming of 70 *ms*.

An analysis of variance was conducted for the variables LOC and GEO, and for RECENCY (a *post hoc* variable, defined to be equal to 1 in trials in which the stimulus in the immediately preceding trial was identical to the present stimulus, and 2 otherwise). In addition, the influence of SUBJECT, declared as a random effect, was examined. The main effect of SUBJECT was significant ($F[3, 165] = 29.91$, $p < 0.0001$), but its interactions with the other variables were not.

The analysis of variance revealed significant main effects of LOC ($F[1, 165] = 7.28$, $p < 0.008$) and GEO ($F[1, 165] = 4.25$, $p < 0.041$). The main effect of RECENCY was significant ($F[1, 165] = 4.91$, $p < 0.029$), and so was its interaction with LOC ($F[1, 165] = 5.27$, $p < 0.023$); there was also a hint of interaction of RECENCY with GEO ($F[1, 165] = 1.89$, $p = 0.17$). A separate analysis by levels of RECENCY revealed that the effects of LOC and GEO were mostly confined to trials in which RECENCY was equal to 1.

4.2.2 Discussion

The absence of interaction between SUBJECT and the variables of interest, LOC and GEO, means that the effects of the latter were the same across subjects (despite the large differences in the mean RT between various subjects). Thus, the SUBJECT differences can be safely omitted from further discussion.

The pattern of RTs in this experiment (see Figure 7 and Table 1) conforms to the expectations. The mean RT was the fastest when the prime was identical to the target, and the slowest when the prime was different both in its complement of parts and in the location of the parts. The success of the experimental manipulation of LOC and GEO manifested itself in that the RTs for the other two combinations of these variables was intermediate. Thus, both the effects of LOC and of GEO were significant, although the former was somewhat stronger (as judged by the priming time and by the ANOVA sum-of-squares criteria).

The identity of the stimulus in the immediately preceding trial (coded by the RECENCY variable) also had a strong effect on RT, as expected from the literature (Luce, 1986). The confinement of the LOC and GEO effects to trials with RECENCY=1 can be explained tentatively by noting that the four categories of stimuli in the present experiment were quite similar to each other, and, moreover, that there were only four distinct priming conditions. In this situation, the differential effect of the prime/target similarity on the activity of the representational mechanism probably needed the additional boost imparted by an identical preceding target.

The finding of pronounced LOC and GEO effects in experiment 1 confirmed the feasibility of exploring the nature of structure representation by differential priming of shape and location. The range of conditions tested did not, however, allow us to draw conclusions concerning the particular mechanism involved in the priming phenomenon. Both on the SSM and on the Chorus accounts, it is unlikely that each of our four categories of stimuli activated a separate mechanism in the subject's
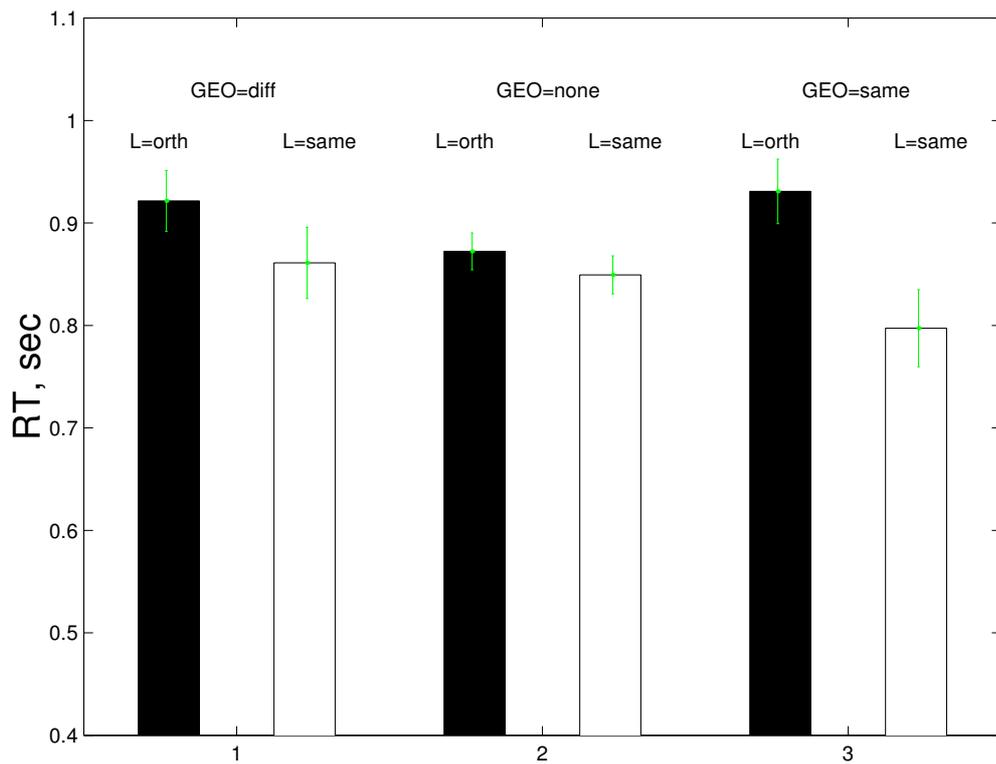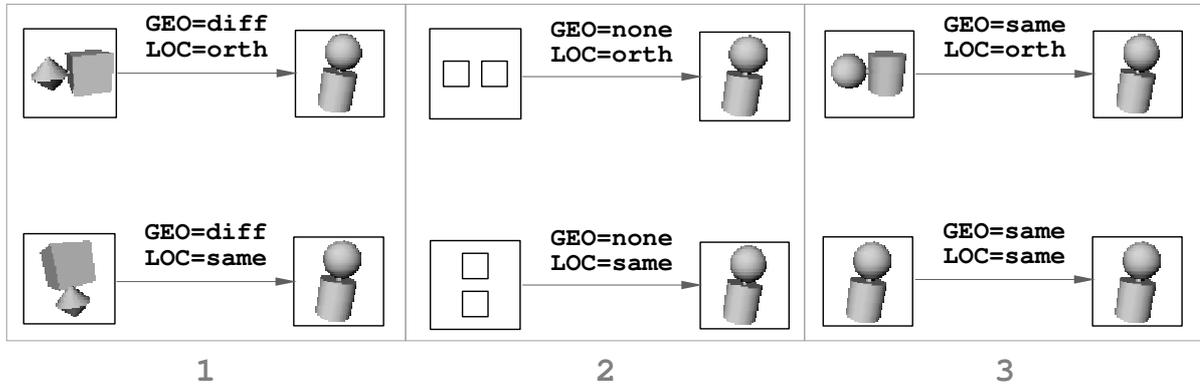
Figure 8: Experiment 2. *Top:* the six priming conditions; these were the same as in experiment 1, plus two conditions in which an empty frame was substituted for visible parts. *Bottom:* the response times.

visual system. With the same mechanisms being activated (in varying degrees) by all the stimuli, the constraint inherent in the structure of the priming objects in experiment 1 (namely, the identity of Loc=*orth* priming condition for one category of targets to the Loc of another category) could have led to an interference between Loc and Geo effects. This constraint was removed, in two steps, in the next two experiments.

4.3   Experiment 2

To reduce the possible interference between the effects of Loc and Geo, in experiment 2 we added a part-neutral prime condition. Specifically, in some of the trials empty box-like frames were used as the priming stimuli, to offer the subject the proper location/relational cues, but no shape information. We note that past attempts to prime abstract frames of reference met with mixed success. For example, (Koriat and Norman, 1984), who studied mental rotation, found that pre-cuing the attitude of the target by displaying an appropriately oriented empty frame ahead of time did little to reduce the response-time cost of misorientation. In comparison, (Treisman, 1992) found that an object can facilitate the response to its reappearance in a different location, if the two locations are linked by a continuous drift of an empty frame.

4.3.1   Results

Five subjects participated in this experiment. The mean RT was 872 *ms*; the breakdown of RT by priming condition is plotted in Figure 8. Changing location from *orthogonal* to *same* (corresponding to the two levels of the Loc variable) resulted in RT gain (priming) of 72 *ms*. In comparison, changing part shapes from *different* and from *none* to *same* (corresponding to the three levels of the Geo variable) resulted in smaller RT differences of 27 *ms* and -3 *ms*, respectively.

As before, an analysis of variance was conducted for the variables Loc, Geo, and Recency, as well as for Subject. The main effect of Subject was significant ($F[4, 498] = 29.91$, $p < 0.0001$), but its interactions with the other variables were not.

The analysis of variance revealed a significant main effect of Loc ($F[1, 498] = 9.51$, $p < 0.0022$), but not of Geo ($F < 1$). The main effect of Recency was significant ($F[1, 498] = 18.26$, $p < 0.0001$), and so was its interaction with Loc ($F[1, 498] = 6.76$, $p < 0.0097$). As in experiment 1, a follow-up analysis revealed the source of this interaction to be in the difference between the effect of Loc for the two levels of Recency. As before, this effect was confined to trials in which Recency was equal to 1.

4.3.2   Discussion

Perhaps the most surprising outcome of experiment 2 is the dwindling of the Geo effect. Let us consider the results observed for the newly introduced Geo=*none* condition. An examination of the RT data (Figure 8 and Table 1) reveals that even for Loc=*same* trials alone, the change of Geo from *different* to *none* reduced RT only by 12 *ms*. The further reduction of RT as Geo changed from *none* to *same* — 52 *ms* — was apparently not enough to make the overall effect of Geo significant. This, however, may have happened because of the opposite effect of Geo for Loc=*orthogonal*, where the same change caused an *increase* of RT by 59 *ms*. We attempted to disentangle these effects in the next experiment.

4.4   Experiment 3

This experiment included an additional location-neutral prime condition, resulting in nine conditions altogether (see Figure 9, top. This allowed us to scrutinize the effects of Loc and Geo independently,

14
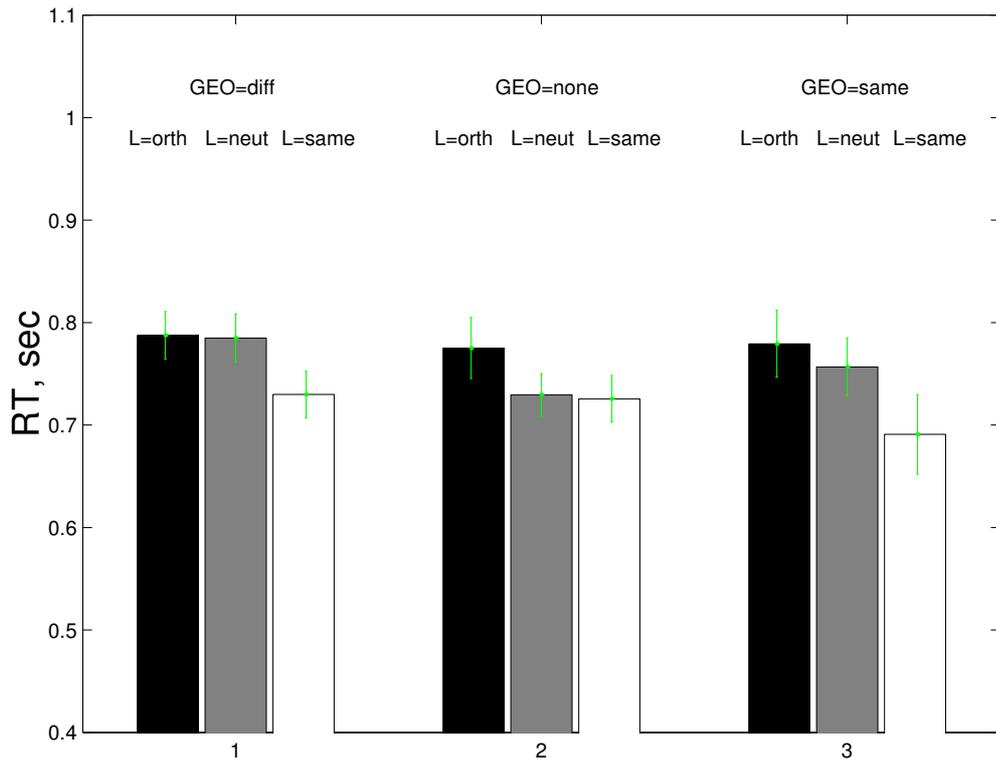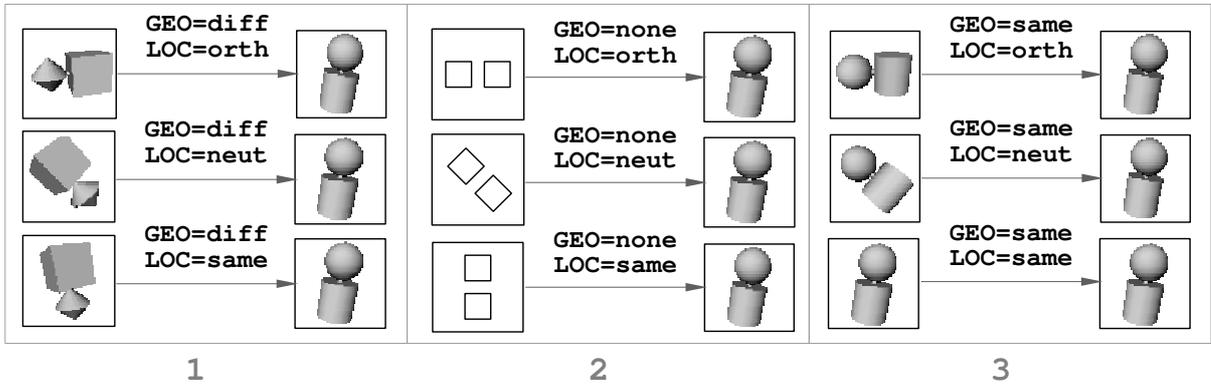
Figure 9: Experiment 3. *Top:* the nine priming conditions; these are the same as in experiment 2, plus three conditions in which the prime could be rotated by 45° relative to the target. *Bottom:* the response times.

by observing change in RT as Loc changed from *neutral* to *same*, and Geo — from *none* to *same*.

4.4.1 Results

Three subjects participated in this experiment. The mean RT was 751 *ms*; the breakdown of RT by priming condition is plotted in Figure 9, bottom. Changing location from *orthogonal* to *same* and from *neutral* to *same* (corresponding to the three levels of the Loc variable) resulted in RT gains (priming) of 66 *ms* and 42 *ms*, respectively. In comparison, changing part shapes from *different* and from *none* to *same* (corresponding to the three levels of the Geo variable) resulted in smaller RT differences of 25 *ms* and 1 *ms*, respectively.

As in the previous experiments, an analysis of variance was conducted for the variables Loc, Geo, and Recency, as well as for Subject. The main effect of Subject was significant ($F[4, 496] = 36.31$, $p < 0.0001$), but its interactions with the other variables were not.

The analysis of variance revealed a significant main effect of Loc ($F[2, 496] = 3.36$, $p < 0.0356$), but not of Geo ($F < 1$). The main effect of Recency was significant ($F[1, 496] = 15.93$, $p < 0.0001$), but its interactions with the other variables were not.

4.4.2 Discussion

The results of experiment 3 indicate that Geo does not have as strong a facilitatory effect on RT as Loc. To see that, let us leave aside the difficult-to-interpret conditions in which Loc=*orthogonal*, or Geo=*different*. A scrutiny of the data (see Table 1) then reveals that (1) for Geo=*same*, the change of Loc from *neutral* to *same* resulted in RT becoming faster by 66 *ms*, while (2) for Loc=*same*, the change of Geo from *none* to *same* reduced RT only by 35 *ms*.

One should keep in mind, of course, that the changes in Geo and Loc underlying the effects just reported are formally incommensurable: it is meaningless to draw a comparison between (1) the 45° rotation, giving rise to the Loc change, and (2) the appearance of two geons instead of an empty frame, giving rise to the Geo change. Still, the *outcome* of the change in Geo, which did not reach statistical significance in the overall ANOVA, is about half as strong as that of the change in Loc.

Geo	Loc	Exp. 1	Exp. 2	Exp. 3
diff	orth	899	922	788
diff	diag	—	—	785
diff	same	839	861	730
none	orth	—	872	775
none	diag	—	—	729
none	same	—	849	726
same	orth	860	931	779
same	diag	—	—	757
same	same	738	797	691

Table 1: Mean RTs (*ms*) by condition, in the three experiments. The RTs were estimated by the LSMEANS option of the General Linear Models (GLM) procedure we used for the analysis of variance (SAS, 1989).

5 General discussion

Some provisional conclusions that can be drawn from the results of the three experiments described above are:

1. Similarity in either shape (GEO) or location (LOC) between the prime and the target can facilitate (speed up) the response to the target in a 4AFC setting.

2. The contribution of shape (GEO) to this facilitation is quantitatively weaker than that of location (LOC), and tends to be not statistically significant in a setting where the two effects can be separated experimentally.

These findings are not entirely compatible with the structural description models of representation. For example, a central prediction of SSM is priming by "disembodied" parts or geons, corresponding to our GEO effect, which experiments 2 and 3 showed to be weak and not statistically significant. Nor are our results compatible with the holistic models, such as Chorus. Specifically, Chorus cannot account for the LOC effect — priming by "shapeless" location — which we found in all three experiments.

This combination of results can be interpreted in terms of a hybrid model such as Chorus of Fragments as follows. According to CoF, conjunctions of shape and location are explicitly represented, making each potentially amenable to priming, perhaps to different degrees. Consider again the schematic depiction of CoF in Figure 4, left. Priming the two modules labeled as A_2 and B_2 will facilitate subsequent processing of stimuli in the lower visual field; this could be the source of a LOC effect, of the kind we found in the psychophysical experiments. Likewise, priming the two modules labeled as A_1 and A_2 will lead to a facilitation in the processing of the shape denoted by A — a GEO-like effect. The relative strength of these two effects, which depends on the contribution of the various modules to the decision-making stage, can be made to fit the observed pattern within the general computational framework specified by the CoF model. We shall propose some experimental ways to strengthen our conclusions concerning the three classes of models, after discussing related data from several disciplines.

5.1 Related work: psychophysics

Results stemming from priming studies were the major source of support for the structural models of recognition of which SSM is an example. In particular, (Biederman and Cooper, 1991a) reported complete translational (and rotational) invariance of priming, as predicted by SSM. The results of another study, which examined the pattern of priming across several conditions in which the objects' contours were partially deleted, suggested explicit involvement of geon-like intermediate representations postulated by SSM (Biederman and Cooper, 1991b).

Other studies that used priming yielded evidence of incomplete invariance with respect to rotation in depth (Srinivas, 1993; Lawson et al., 1994; Gauthier and Tarr, 1997; Williams and Tarr, 1998). The strong influence of view-to-view similarity on priming is consistent with an extensive body of data obtained within other experimental paradigms, as reviewed by (Jolicoeur and Humphrey, 1998). We note that recognition that generally falls short of being invariant under rotation is a hallmark of the view-interpolation scheme of representation (Poggio and Edelman, 1990; Bülthoff et al., 1995), from which both the Chorus and the CoF models are derived. Interestingly, a lack of invariance has been reported even for translation, especially for the stimulus moving from one quadrant of the visual field to another (Bar and Biederman, 1998).[6] Such an outcome is a direct consequence of the kind of split treatment of the visual field postulated by the CoF model.

---

[6]A much earlier report of a similar effect of translation can be found in (Wallach and Austin-Adams, 1954). We thank S. Kaufmann for bringing this reference to our attention.

Psychophysical studies of object representation more often than not involved quantifying the effects of manipulating entire intact objects rather than object parts. The contour-deletion study of (Biederman and Cooper, 1991b) mentioned above is among the few exceptions in this respect. Another exception is the work by (Cave and Kosslyn, 1993), who compared the effects of various kinds of decomposition of line drawings of everyday objects on the naming time. They report that the way an object is divided into parts has very little influence on its identification, and note that this evidence speaks against part decomposition being necessarily prior to recognition.

The issues of part structure and translation invariance were addressed jointly in a recent study by (Dill and Edelman, 1997), who tested *same-different* discrimination of animal-like shapes generated and controlled by computer graphics. In a sense, that study complemented the present one, by examining the effects of larger-extent translation than what we achieved here by manipulating the Loc variable. The two stimuli in each trial were displayed at the same or at different locations of the visual field (with differences varying between about 2° and 4°). Both for intact and for scrambled animal shapes, Dill and Edelman found complete translation invariance — but only if the shapes were distinguishable on the basis of local cues. The invariance was lost when the stimuli were made to be distinguishable only on the basis of structural (relational) properties involving more than one fragment — an outcome that is compatible with the Chorus of Fragments (CoF) model.

The idea that the representation of a structure may be tied to a particular location in the visual field where it is first observed is compatible both with the CoF model, and with the notion of *object file* — a hypothetical record which is created by the visual system for every encountered object and which persists as long as the object is observed (Kahneman et al., 1992). Results obtained by Treisman and her associates, summarized in (Treisman, 1992), indicate that "location" (as it appears, e.g., in CoF) should perhaps be interpreted relative to the focus of attention, rather than retinotopically, a distinction that certainly deserves further research.

5.2 Related work: physiology

Although priming is a behaviorally defined phenomenon, it has a physiological counterpart — the sustained change in the activity of a unit, caused by the animal's exposure to a stimulus that is effective for that unit (Wiggs and Martin, 1998). A subsequent exposure to an identical (or a sufficiently similar) stimulus further modifies the response. Electrophysiological studies in the monkey revealed visual priming both of the excitation and of the suppression variety which operates in an object-specific manner (Miller et al., 1993; Miller and Desimone, 1994). Thus, one may hope that integration of psychophysical and physiological findings concerning object representation would eventually become possible.

One way to pursue this goal in human subjects is already available, in the form of the functional magnetic resonance imaging (fMRI) technology. The results of two recent studies are especially relevant to the present investigation. The first of these (Grill-Spector et al., 1998b) investigated the impact of progressive scrambling of stimulus images on the activation of the visual cortex. It was found that early visual areas were activated by images of natural objects cut up into 32×32 squares and rearranged randomly to the same extent as by the intact images of these objects. In comparison, the Lateral Occipital (LO) complex (Malach et al., 1995), being roughly the human homologue of areas V4 and TEO, exhibited a much reduced activation when the subjects were shown images scrambled into 16×16, but not into 4×4 squares (it is interesting to note in this connection that (Bar and Biederman, 1998) conjectured the source of the quadrant-specific priming they observed to lie in area V4). This result suggests that the "grain" of the representation in area LO is intermediate, that is, smaller than entire objects, yet larger than local features, as postulated by the CoF model.

The second fMRI study (Grill-Spector et al., 1998a) exploited the rapid shape-adaptation effect (a variety of priming), in which repeated presentations of identical images gradually result in a

reduced activation. Subjects were shown repeatedly either identical images of an object (face or car) or the same object but under various translations, illuminations or viewpoints. In all subjects, voxels in area LO were activated maximally by images of different exemplars compared to scrambled images. Presentation of identical images produced 53% of the maximal signal. In comparison, images of the same object but at different translations yielded 78% and changing illuminations or viewpoint — 89% of the maximal signal. These results indicate that object processing mechanisms in human vision treat various image transformations differentially, as suggested also by some recent computational models (Vetter et al., 1995; Riesenhuber and Poggio, 1998). In particular, this means that translation need not be fully compensated for — a phenomenon that could give rise to the effect of LOC in the present study.

The neural basis for these fMRI data may be provided by the columnar structure revealed in the inferotemporal (IT) cortex of the monkey by electrophysiological means (Tanaka, 1992). Specifically, the spatial structure of columns of cells tuned to similar shapes may exist in the brain on a sufficiently large scale to be detectable by fMRI despite the relatively coarse resolution of this technique (Edelman et al., 1998). This finding in itself constitutes support for models of the Chorus variety, and, in particular, for CoF. A closer look at the receptive fields of the object-tuned units in IT shows that they are frequently located eccentrically (Kobatake and Tanaka, 1994; Ito et al., 1995). These units may, therefore, carry location and not only shape information, as postulated by the CoF model.

5.3 A computer vision perspective

In computer vision, there have been some recent attempts to combine the simplicity of representing objects by multiple views (as it is done in the Chorus model) with the robustness of structural descriptions (as in SSM). Like CoF, these approaches involve the estimation of 2D, image-based feature layout (as opposed to 3D, object-centered structure). In one example, evidence concerning object identity is iteratively refined by considering mutual constraints based on relative locations of simple template-like features in an image (Amit and Geman, 1997). Likewise, encoding the rough structure of objects in image coordinates can support object recognition in the presence of occlusion and clutter (Nelson and Selinger, 1998). The latter system can also perform categorization of novel instances of familiar classes. Importantly, it represents object structure explicitly, making it, in principle, capable of reasoning about object parts — a serious challenge for holistic methods such as Chorus, but not, we believe, for hybrid approaches such as CoF.

5.4 Summary

The results of the psychophysical experiments reported here, along with a wide range of data gleaned from recently published psychophysical, physiological and computational works, suggest that object structure may be represented in human vision by a hybrid system, based on two principles: (1) statistically defined image fragments serving as the basic features that comprise shape prototypes, and (2) the rough topographical layout of such fragments serving as the representation of object/scene structure. Clearly, further work (including computational simulations) needs to be done to substantiate this hypothesis.

Among the psychophysical issues that need to be resolved, our highest priority is assigned to clarifying the nature of the effective similarity of spatial structure. In a representation of object structure, the locations of parts can be defined relative to each other (as in SSM), or relative to a coordinate system centered at the fixation point (as in CoF). We plan to distinguish between these possibilities by manipulating object structure in two different ways. First, we shall shift the relative locations of object parts in the prime (keeping the target shape intact). Data from this experiment

could clarify whether spatial relations are represented in a quantized, all-or-none fashion (as in SSM), or in a graded fashion (as in CoF). Second, we shall change the location of the prime relative to the target (which will be kept fixed). The prediction of SSM here is that priming should not depend on the relative displacement of the prime and the target. In comparison, CoF predicts that progressive displacement should disrupt the priming and eventually reduce it to nil.

Acknowledgments

References

Amit, Y. and Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation*, 9:1545–1588.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 85:249–277.

Bar, M. and Biederman, I. (1998). Subliminal visual priming. *Psychological Science*, -:–. in press.

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In Srull, T. K. and Wyer, R. S., editors, *Advances in social cognition, Volume III: Content and process specificity in the effects of prior experiences*, pages 61–88. Lawrence Erlbaum Associates, Hillsdale, NJ.

Barsalou, L. W. (1998). Perceptual symbol systems. *Behavioral and Brain Sciences*. in press.

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, 6:325–356.

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

Biederman, I. and Cooper, E. E. (1991a). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, 20:585–593.

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

Biederman, I. and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for 3D viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19:1162–1182.

Biederman, I. and Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff. *Journal of Experimental Psychology: Human Perception and Performance*, 21:1506–1514.

Bienenstock, E. and Geman, S. (1995). Compositionality in neural systems. In Arbib, M. A., editor, *The handbook of brain theory and neural networks*, pages 223–226. MIT Press.

Bienenstock, E., Geman, S., and Potter, D. (1997). Compositionality, MDL priors, and object recognition. In Mozer, M. C., Jordan, M. I., and Petsche, T., editors, *Neural Information Processing Systems*, volume 9. MIT Press.

Bülthoff, H. H., Edelman, S., and Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, 5:247–260.

Cave, C. B. and Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. *Perception*, 22:229–248.

Dill, M. and Edelman, S. (1997). Translation invariance in object recognition, and its relation to other visual transformations. A. I. Memo No. 1610, MIT.

Edelman, S. (1994). Biological constraints and the representation of structure in vision and language. *Psycoloquy*, 5(57). FTP host: ftp.princeton.edu; FTP directory: /pub/harnad/Psycoloquy/1994.volume.5/; file name: psyc.94.5.57.language-network.3.edelman.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Science*, 1:296–304.

Edelman, S. (1998). Representation is representation of similarity. *Behavioral and Brain Sciences*, 21:449–498.

Edelman, S. and Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. *Phil. Trans. R. Soc. Lond. (B)*, 352(1358):1191–1202.

Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. (1998). Towards direct visualization of the internal shape representation space by fMRI. *Psychobiology*, -:–. to appear.

Gauthier, I. and Tarr, M. J. (1997). Orientation priming of novel shapes in the context of viewpoint-dependent recognition. *Perception*, 26:51–73.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., and Malach, R. (1998a). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Proc. Israeli Neuroscience Symposium.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., and Malach, R. (1998b). A sequence of early object processing stages revealed by fMRI in human occipital lobe. *Human Brain Mapping*, 6:316–328.

Hummel, J. E. (1998). Where view-based theories of human object recognition break down: the role of structure in human shape perception. In Dietrich, E. and Markman, A., editors, *Cognitive Dynamics: conceptual change in humans and machines*, pages –. MIT Press. in press.

Hummel, J. E. and Biederman, I. (1990). Dynamic binding: a basis for the representation of shape by neural networks. In *Proc. 12th Annual Conference of the Cognitive Science Society*, pages 614–621, Hillsdale, NJ. Erlbaum.

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

Hummel, J. E. and Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104:–.

Intrator, N., Edelman, S., and Bülthoff, H. H. (1995). An integrated approach to the study of object features in visual recognition. *Network*, 6:603–618.

Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.*, 73:218–226.

Jolicoeur, P. and Humphrey, G. K. (1998). Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In Walsh, V. and Kulikowski, J., editors, *Perceptual constancies*, chapter 10. Cambridge University Press, Cambridge, UK. in press.

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, 24:175–219.

Kirschfeld, K. (1995). Neuronal oscillations and synchronized activity in the central nervous system: functional aspects. *Psycoloquy*, 6(36). available electronically as ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1995.volume.6/psyc.95.6.36.brain-rhythms.11.kirschfeld.

Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.*, 71:856–867.

Kobatake, E., Wang, G., and Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.*, 80:324–330.

Koenderink, J. J. and van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32:211–217.

Koriat, A. and Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10:421–434.

Lawson, R., Humphreys, G., and Watson, D. G. (1994). Object recognition under sequential viewing conditions: evidence for viewpoint-specific recognition procedures. *Perception*, 23:595–614.

Loftus, G. and Mason, M. (1994). Using confidence intervals in within subjects designs. *Psychonomic Bulletin and Review*, 1:476–490.

Lowe, D. G. and Binford, T. O. (1985). The Recovery of Three-Dimensional Structure from Image Curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(3):320–326.

Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization.* Oxford University Press, Oxford.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, J., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., and Tootell, R. B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Science*, 92:8135–8139.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B*, 200:269–294.

Medin, D. L., Goldstone, R. L., and Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100:254–278.

Mel, B. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition. *Neural Computation*, 9:777–804.

Miller, E. K. and Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, 263:520–522.

Miller, E. K., Li, L., and Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neuroscience*, 13:1460–1478.

Minsky, M. (1975). A framework for representing knowledge. In Winston, P. H., editor, *The psychology of computer vision*. McGraw-Hill, New York.

Nelson, R. C. and Selinger, A. (1998). Large-scale tests of a keyed, appearance-based 3-D object recognition system. *Vision Research*, 38:2469–2488.

Newell, F. N. (1998). Stimulus context and view dependence in object recognition. *Perception*, 27:47–68.

Newell, F. N. and Findlay, J. M. (1997). The effect of depth rotation on object identification. *Perception*, 26:1231–1257.

Ochsner, K. N., Chiu, C.-Y. P., and Schacter, D. L. (1994). Varieties of priming. *Current Opinion in Neurobiology*, 4:189–194.

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

Richards, W. and Jepson, A. (1992). What makes a good feature? A.I. Memo No. 1356, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Riesenhuber, M. and Poggio, T. (1998). Just one view: Invariances in inferotemporal cell tuning. In M. I. Jordan, M. J. K. and Solla, S. A., editors, *Advances in Neural Information Processing*, volume 10, pages –. MIT Press. in press.

SAS (1989). *User's Guide, Version 6*. SAS Institute Inc., Cary, NC.

Schiele, B. and Crowley, J. L. (1996). Object recognition using multidimensional receptive field histograms. In Buxton, B. and Cipolla, R., editors, *Proc. ECCV'96*, volume 1 of *Lecture Notes in Computer Science*, pages 610–619, Berlin. Springer.

Srinivas, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19:582–602.

Tanaka, K. (1992). Inferotemporal cortex and higher visual functions. *Current Opinion in Neurobiology*, 2:502–505.

Tanaka, K., Saito, H., Fukada, Y., and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.*, 66:170–189.

Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, 47:862–875.

Tulving, E. and Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247:301–306.

Ullman, S. (1996). *High level vision*. MIT Press, Cambridge, MA.

Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992–1005.

Vetter, T., Hurlbert, A., and Poggio, T. (1995). View-based models of 3d object recognition: Invariance to imaging transformations. *Cerebral Cortex*, 5:261–269.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5:520–526.

Wallach, H. and Austin-Adams, P. (1954). Recognition and the localization of visual traces. *American Journal of Psychology*, 67:338–340.

Wiggs, C. L. and Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.*, 8:227–233.

Williams, P. and Tarr, M. J. (1998). Orientation-specific possibility priming for novel three-dimensional objects. *Perception and Psychophysics*, -:–. in press.