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

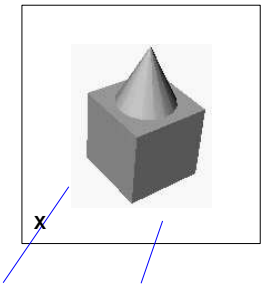
November 27, 1998

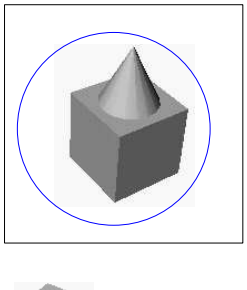
Abstract

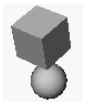
The ries f bject representati n can be classified as structural, h listic r hybrid, depending n their appr ach t the mere l gy and c mp siti nality f shapes. We tested the predicti ns f s me f the current the ries in three experiments, by quantifying the effects f vari us priming cues n resp nse times t 3D bjects. In experiment 1, there were tw p ssible 2.31e1sTd(Durham,)Tj,1Td(cu1T

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


x





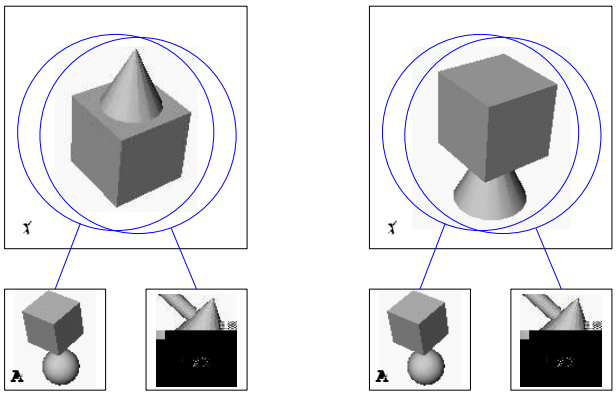


Figure 2: The holistic approach is limited in its ability to make explicit the similarities and the differences between complex objects which human observers would describe as composed of similar parts arranged in different configurations. Consider, for example, the two objects shown here: a cone on 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

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.

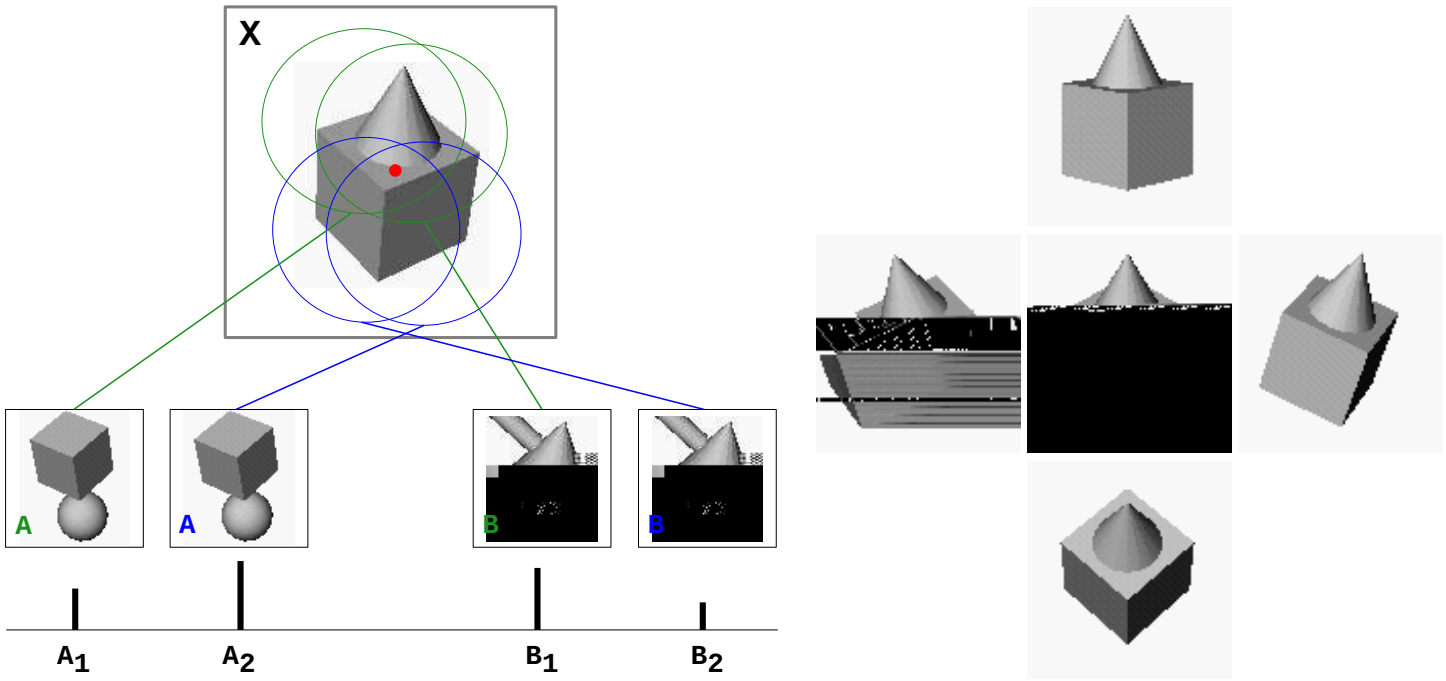


Figure 3: It may be possible to circumvent the problem illustrated in Figure 2 using modules tuned to image fragments, in conjunction with binding by retinotopy. *Left:* In such a scheme, which may be called the Chorus of Fragments (CoF), each object-specific module would come in several varieties, distinguished by the location of the module's receptive field relative to the fixation point (indicated by the thick dot). Here, module $\mathtt{A}_1$ responds optimally when the fixation is above and slightly to the left of a stimulus resembling object `A`. Likewise, module $\mathtt{A}_2$ 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

2.3 The Cho us of F agments (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

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.

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

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 Expe iment 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

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 $\mathtt{A}_2$ and $\mathtt{B}_2$ 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 $\mathtt{A}_1$ and $\mathtt{A}_2$ will lead to a facilitation in the processing of the shape denoted by $\mathtt{A}$ — 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 wo k: 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.

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 compute vision pe spective

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). Lik

could clarify whether

Bülthoff, H. H., Edelman, S., and Tarr, M. J. (1995). How are three-dimensional

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

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.