Shape representation by $\mathcal{S}$econd-order $\mathcal{I}$somorphism, and the $\mathcal{C}$horus model: $\mathcal{SIC}$.

Shimon Edelman*
Center for Biological and Computational Learning
Dept. of Brain and Cognitive Sciences
MIT E25-201
Cambridge MA 02142, USA
edelman@ai.mit.edu

January 1998

Abstract

Proximal mirroring of distal similarities is, at present, the only solution to the Problem of Representation that is both theoretically sound (for reasons discussed in the target article) and practically feasible (as attested by the performance of the Chorus model). Augmenting the latter by a capability for referring selectively to retinotopically defined object fragments should lead to a comprehensive theory of shape processing.

1 An overview of the commentaries

The relationships among the stances taken by the commentators on the various issues having to do with representation and similarity can be visualized with the help of Figure 1. This figure depicts a two-dimensional embedding of a textually defined "commentary space" in which each commentary is represented by a point labeled by its author's initial.

The center of the plot is occupied by commentaries that touch upon relatively few of the 11 issues used to define conceptual similarity in this visualization exercise. Whereas the units along the two dimensions are, of course, arbitrary, the locations and the proximities in the plot can be given an interpretation. For example, the upper right corner contains the minders of computational issues, and, in particular, of top-down influences; the lower right is occupied by the champions of nonlinear dynamics, and the lower left is where the proponents of combined metric and structural representations are. All these issues, along with some of the specific concerns raised by the commentators, are discussed next.

2 Veridicality

The strongest concerns in connection with veridicality are voiced by **Hahn and Chater**, who contend that the notion of an objective shape space in which proximity corresponds to similarity is problematic, because, as pointed out by Goodman and Watanabe, objective similarity is an ill-defined concept. **Eisler** goes even further, stating that he does not use the term "subjective similarity" because there is no such thing as "objective similarity" in the first place.

* Present address: School of Cognitive and Computing Sciences, University of Sussex, Falmer BN1 9QH, UK

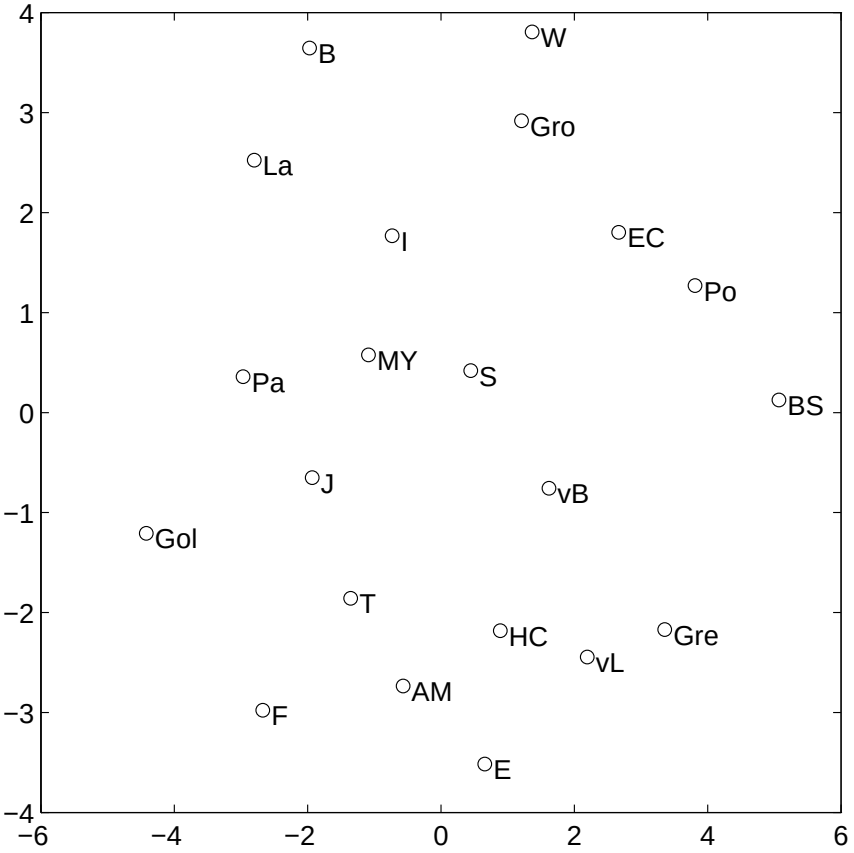

Figure 1: A 2D rendition of an 11-dimensional "commentary space" derived from the 21 commentaries. Each commentary was first described by 11 binary predicates, chosen so as to cover the major issues raised in all 21 of them. The issues were defined by the appearance in the text of the following keywords or key concepts: (1) warped similarity spaces, (2) differences vs. similarities, (3) veridicality, (4) the influence of context on similarity, (5) computational complexity, (6) compositionality and structural similarity, (7) mention of nonlinear dynamics, (8) top-down effects, including adaptive resonance theories, (9) holism, (10) invariances, and (11) neurobiology. If a given key phrase appeared in a particular commentary, the corresponding bit in the feature vector describing that commentary was set to 1; otherwise, it was set to 0. The 21×21 matrix of pairwise Euclidean distances between the commentaries was then formed, and the 21 points were embedded into a 2D space by metric multidimensional scaling (MDS). In the plot, the points are labeled by the initials of the commentators (AM: Andresen and Marsolek, B: Benson, BS: Bonmassar and Schwartz, E: Eisler, EC: Eklundh and Carlsson, F: Foldiak, Gol: Goldstone, Gre: Gregson, Gro: Grossberg, HC: Hahn and Chater, I: Intrator, J: Jüttner, La: Latimer, MY: Markman and Yamauchi, Pa: Palm, Po: Postma, van der Herik and Hudson, S: Sokolov, T: Tovee, vB: van Brakel, vL: van Leeuwen, W: Williamson. The coefficient of congruence (Borg and Lingoes, 1987) between the distances in the MDS-derived 2D configuration and in the original 11D one was 0.97, signifying that much fewer than 11 dimensions were sufficient to describe the contextual similarities among the different commentaries.

A typical argument against the notion of objective similarity can be found in (Murphy and Medin, 1985), who note that the number of attributes shared by plums and lawn-mowers could be infinite: both weigh less than 1000 kilograms (and less than 1001 kilograms), both cannot hear well, both have a smell, etc. Watanabe (1985) formalized this kind of reasoning, by proving that any two objects are as similar to each other as any other two objects, insofar as the degree of similarity is measured by the number of shared predicates (provided that the set of predicates is finite and equally applicable to all objects, and that no two objects are identical with respect to this set).

While being formally impeccable, these arguments leave one with a suspicion of being cheated out of using a perfectly serviceable concept — similarity — by some kind of definitional sleight of hand (what Dennett calls an intuition pump). Somehow, the deep intuitive roots of similarity play a part in this show: without the reader's utter and absolute conviction that plums are *not* similar to lawn-mowers, the impact of Murphy and Medin's example would be considerably weakened. Quite perversely, this conviction emerges unscathed even from the formal argument: plums are not perceived as similar to lawn-mowers no matter what, despite the recruitment of silly features common to both, such as not being able to hear well.

The resolution we are offered for this conundrum consists of bringing into the consideration an *observer*, whose system of "values" (Watanabe, 1985) or "prior spacing of qualities" (Quine, 1969) removes the ambiguity by introducing a bias (Goldstone, 1994). Indeed, in a precursor to the target paper (Edelman, 1995), I cited Watanabe and Quine in support of a particular kind of bias in the perception of similarities — the natural bias imposed by the standard machinery of biological vision (receptive fields with smooth graded profiles, etc.).

A logical continuation of this approach, suggested by **Hahn and Chater**, is to consider the nature (in particular, the veridicality) of the mapping between the representational systems of two observers, instead of the mapping between the world and the observer's similarity space. It is interesting to note that a straightforward rephrasing of the relevant passages of the target paper (substituting "another observer's" for "distal") leaves the computational conclusions concerning veridicality, *mutatis mutandis*, intact. In particular, if the composition of the mappings of the two observers, $M_1 \circ M_2^{-1}$, is smooth, and if no dimensions are lost (projected out) along the way, the two representation spaces will be locally second-order isomorphic.

Establishing the possibility of veridical *communication* between two observers in the manner suggested above shifts the focus of discussion away from the possibility of veridical *perception*. This, however, means that somewhere along the way the real world of shapes gets lost. Do we have to give up the notion of objective similarity altogether just to annul the standard philosophical arguments against it? **Hahn and Chater** answer in the affirmative, drawing an analogy between the discredited correspondence theory of truth and the second-order isomorphic representation of objective similarities. I reject this analogy, and contend that, as far as shape *geometry* is considered, this amounts to throwing out the baby with the bath water.

Intuitively, the geometry of a plum is very different from that of a lawn-mower, because any shape-preserving transformation[1] applied to the former would leave a residual discrepancy that is large relative to the size of the smaller of the objects involved in the comparison — and also large relative to the residual that is left when a plum and a melon are compared. More formally, a survey of the mathematical theory of shape spaces developed in the last decade (and mentioned briefly in the target article) suggests that shape can be formalized naturally along these lines, in such a manner that similarity is unique (defined by proximity along minimal geodesics in the shape space) in all but certain degenerate cases (Kendall, 1984; Carne, 1990; Le, 1991; Le and Kendall, 1993; Bookstein, 1996).

Unfortunately, all the commentators who had had problems with my notion of veridicality ignored the proposal mentioned above, despite its appearance in the target paper. An exception is **van Brakel**'s commentary, where the idea of a common parameterization basis for distal similarity is mentioned, only to be dismissed as "highly disputable." In support of this dismissal, the reader is given two examples. The first

[1] Shape-preserving transformations are the rigid motions and uniform scaling; stretching and bending, which could bring a plasticine plum into congruence with a toy lawn-mower, are disallowed.

of these deals with color, and is, therefore, irrelevant in the context of shape description and representation (except as a psychological rather than psychophysical theory; see **Sokolov**'s commentary). The second example is essentially a paraphrase of Quine's Gavagai-observing situation (Quine, 1960), translated into the Cheyenne language of two centuries ago: the challenge is to reify a highly ambiguous term, *vovetas* that may refer to a black vulture, or to a swarm of dragonflies, or, for all a non-speaker of Cheyenne knows, to the left hind leg of a rabbit. Van Brakel admits that Chorus would be able to acquire the *vovetas* concept, but implies that in doing so, Chorus would not be reflecting anything objective or veridical about the world. My reply is that this does not preclude Chorus from acquiring a genuinely veridical representation in a more natural situation: one that has to do with *natural kinds*. I dare say that van Brakel's tacit assumption that *vovetas* is a natural kind would have been resisted by Quine. Lumping together black vultures and tornadoes may sound exotically appealing, but is about as useful for *prediction* — the main reason for having categories in the first place (Shepard, 1987) — as the classes of animals in the famous excerpt from an ancient encyclopaedia cited by Jorge Luis Borges.[2]

3 Compositionality and the representation of structure

Intrator, **Foldiak**, **Goldstone**, **Markman and Yamauchi**, and **Postma, van der Herik and Hudson** all point out the lack of explicit representation of structure (or, more generally, of various dimensions of similarity) in the Chorus scheme. Of these commentators, Foldiak is the only one who rejects representation by similarities to prototypes altogether. The arguments raised by Foldiak are based on the assumption that this representation scheme is *necessarily* holistic, and, in particular, that dimensions of shape cannot be separated from those of texture or color in the processing of complex objects. This assumption is, however, unwarranted: the Chorus scheme described in the target article can be adapted to attend selectively to different dimensions of variation of the stimuli, in several ways. First, the input space of the prototype modules can be "skewed" and some of its dimensions stressed, as proposed by Foldiak himself, as well as by Postma, van der Herik and Hudson (this is, of course, a standard technique in pattern recognition). Second, the imposition of class labels on a set of stimuli can steer the system towards the formation of a low-dimensional space in which some of the directions of variation are downplayed and others accentuated. In this manner, the system can be made to treat different views of the same object or its different parametrically related versions equivalently, while maintaining discriminability along other dimensions (Intrator and Edelman, 1997). Third, selective association between prototype modules can make some dimensions more important in certain situations. The action of such an association mechanism can be illustrated on Foldiak's example: "there is no way to know whether . . . a 'giraffe' [represented by similarity to a camel and a leopard] is an ungulate with spots or a predator with a hump." Indeed, if I see, for the first time, a thing that resembles a spotted camel or a deformed leopard, I *cannot* tell whether it is going to try to hunt me down or start grazing. One of these acts, however, would immediately suggest the strengthening of an association between the representation of the novel animal and that of its proper class.[3]

Any of these approaches effectively creates a stimulus bias in the similarity space (Shepard, 1964; Nosofsky, 1991), whose action resembles that of assigning a larger weight to some dimensions (i.e., to similarities to some of the prototypes), at the expense of others. However, such adjustment, which may be task-specific

[2]Borges quotes, in the essay "The Analytical Language of John Wilkins" (E. R. Monegal and A. Reid, Borges: A Reader (New York: Dutton, 1981, pp. 141-143) a list, "attributed by Dr. Franz Kuhn to a certain Chinese encyclopaedia entitled 'Celestial Emporium of Benevolent Knowledge.' On those remote pages it is written that animals are divided into: (a) those that belong to the Emperor, (h) embalmed ones, (c) those that are trained, (d) suckling pigs, (c) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance."

[3]This is but an echo of the famous discussion of induction, found in (Hume, 1748), 23ff, which starts thus: "Let an object be presented to a man of ever so strong natural reason and abilities; if that object be entirely new to him, he will not be able, by the most accurate examination of its sensible qualities, to discover any of its causes or effects."

(Schyns et al., 1998), only makes sense if the underlying representation reflects as many as possible stimulus dimensions, because different subsets of these dimensions will be relevant in different situations. Such a *sparse* code, advocated by Barlow (1959) and by others (including **Foldiak**), can be achieved in two ways: by a combination of abstract features (such as "red," an example suggested by Foldiak), or by a combination of multidimensional concrete prototypes (such as "similar to a cherry," as in the Chorus scheme). There is no reason why the former kind of feature should be *a priori* preferable; in fact, abstract features are a very poor basis for categorization and generalization (what do we learn about the nature of an object by being told only that it is red?). In comparison, holistic features such as similarities to prototypes are both useful for generalization and easy to acquire, by a process which Quine calls learning by ostension (as in "this is a cherry," pointing to a cherry). Indeed, infants at the peak of the concept acquisition period around the age 2 exhibit precisely this tendency to attribute labels (words) to shapes of entire objects, rather to their color, or to the shapes of their parts (Markman, 1989; Smith et al., 1997), and so do perceptual novices in general (Tanaka and Gauthier, 1997). Only upon receiving a different label for an already encountered object do they associate it with the object's color, material, or local features.

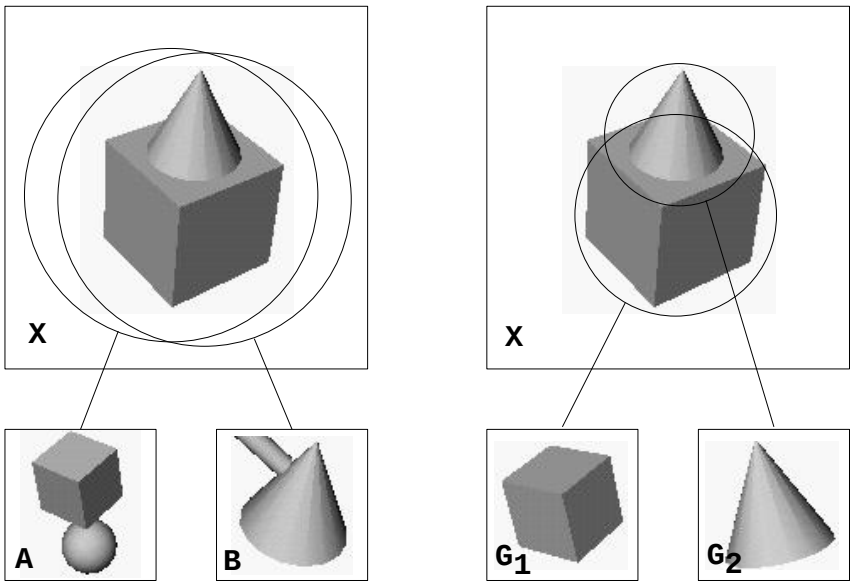

Figure 2: *Left:* Chorus of holistic prototypes; the new object X is represented by its similarities to objects A and B. This representation scheme, which I described in the target paper, can support various recognition-related tasks, working from gray-level images of real objects (Edelman and Duvdevani-Bar, 1997a; Edelman and Duvdevani-Bar, 1997c; Duvdevani-Bar et al., 1998). According to **Latimer**, "it could also provide an explicit, neurally-plausible mechanism for responding directly and accurately to [objects] and their inter-relationships"; **Jüttner** notes that it "transforms images into a rule-based representational format which is open to propositional reasoning" (cf. Barsalou's, 1997, notion of perceptual symbol systems). However, as pointed out by **Intrator**, **Foldiak**, **Goldstone**, **Markman and Yamauchi**, and **Postma et al.**, this scheme does not allow structural decomposition and analysis of shapes. *Right:* Chorus of generic fragments, as suggested by **Postma et al**. This scheme is a simplification (involving image-based fragments) of the standard structural model of representation, such as Biederman's (1987) Recognition By Components (RBC). Neither RBC, nor simplified models such as this one (which does not seek to recover 3D parts and their spatial relationships) has been ever made to work on real images. A compromise approach, which combines the theoretical and practical appeal of Chorus with a certain ability for explicit representation of structure, is illustrated in Figure 3.

Holistic representation (Figure 2) is, therefore, a sensible opening strategy, which can serve as the ba-

sis for the development of more sophisticated analytical approaches. The need for augmenting a holistic similarity-based model with some capabilities for structure manipulation is stressed by **Goldstone** and **Markman and Yamauchi**, who list experimental findings concerning perception and categorization of complex objects and scenes that are best accommodated by a structural model. I agree with their conclusion (drawn also by **Intrator** and by **Eklundh and Carlsson**) that co-existence of multidimensional feature space and structural models is desirable. Such co-existence should not, however, become a goal in itself, lest the difficulties inherent in the purely structural approaches (Edelman, 1997) cancel any potential advantage that may stem from combining structural descriptions with prototype-based shape spaces.

How can one steer a middle way between the holistic feature-space extreme, justly criticized as falling short of replicating human performance in many tasks, and the structural extreme, which has remained a piece of science fiction (albeit an intellectually appealing one) since its introduction more than two decades ago? **Postma, van der Herik and Hudson** claim that a dozen or so reference shapes are unlikely to suffice for distinguishing between each pair of the huge number of naturally occurring shapes. This, however, need not be a problem for a large-scale Chorus-like model. Such a model can have at its disposal hundreds of prototypes modules, of which only a small subset becomes active in any given discrimination task.[4] In comparison, the proposal of Postma et al. to use generic "prototypes" such as Biederman's (1987) "geons" seems to me counterproductive, given the poor track record of geon-based theories in computational vision (Edelman, 1997) and the emerging consensus regarding their shortcomings as models of human object recognition performance (Kurbat, 1994; Tarr et al., 1997; Jolicoeur and Humphrey, 1998).

Intrator's suggestion to use prototypical (statistically defined rather than generic) shapes as "parts" seems to be nearer the mark, if only we can manage to avoid the need for temporal binding of parts — a traditional handicap of the structural approaches. One possible way to do that is to resort to binding by retinotopy (Edelman, 1994), a concept illustrated in Figure 3. In this approach, structure is represented explicitly, but in an image-based rather than object-centered manner. Functionally, this is only a small concession: a full-blown structural description must in any case be extracted anew for each distinct aspect of the object (if it can be extracted at all); image-based structure is aspect-specific by its nature. Computationally, however, the latter is much 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 holistic Chorus scheme is the introduction of attention-like control over the location and the size of the retinal receptive field of each module (which can be done in a hard-wired fashion, as depicted in Figure 3). In other words, the Chorus of prototypes can be turned into a Chorus of fragments, when necessary. This, however, is at present only a conjecture; theoretical analysis and computational experiments currently under way in my laboratory should decide whether or not this approach can endow Chorus with the ability to represent structure without giving up its practical appeal and its straightforward interpretation in terms of familiar mechanisms of biological information processing.

4 Specific vs. abstract similarity

Andresen and Marsolek contend that in Chorus the representation of similarity on an abstract level (as between the words "rage" and "RAGE") must be preceded by its representation on a more concrete level. Furthermore, they note that subjects in priming experiments exhibit double dissociation between the levels: in some conditions, concrete or specific but not abstract visual representations are activated, while in others only abstract representations are primed. They conclude that a distinct system dedicated to abstract representations must exist alongside a specific, Chorus-like one. Their first premise is, however, invalid: the activation of a concrete-level representation does *not* necessarily precede that of an abstract-level one, if the representations are distributed. This point is best illustrated not on totally disparate shapes such as "rage" and "RAGE"

[4] This corresponds to combining Barlow's idea of a sparse code with Tanaka's estimate of 1300-2000 object-tuned modules in the inferotemporal cortex in the monkey.

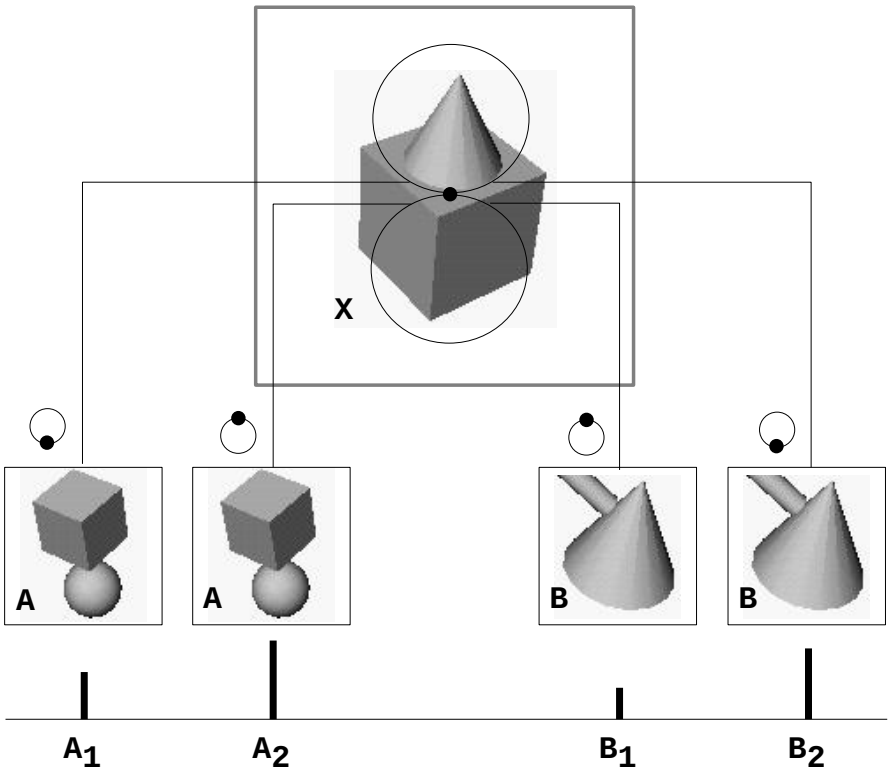

Figure 3: Chorus of prototypical fragments. In this proposed scheme, each object-specific module comes in several varieties, distinguished by the location of the module's receptive field (indicated schematically by the open circle) relative to the fixation point (indicated by the thick dot). For example, module A_1 responds optimally when the fixation is just above a stimulus resembling object A. Likewise, module A_2 prefers the object to be just 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. Here, however, these activities carry additional information concerning the structure of X. For example, the activities of A_2 and B_1 together characterize the shape of the lower fragment of X, while the activities of A_1 and B_2 together determine the shape of its upper fragment — without recourse either to generic parts, or to any kind of binding mechanisms (beyond co-activation and retinotopy). This scheme is even closer to Barsalou's (1997) perceptual symbol system idea.

(for which similarity is solely a matter of convention and should be encoded by a "lateral" association link between two equal-status prototypes), but rather on concepts that are part of a hierarchy, such as giraffe and quadruped. In Chorus, several modules whose activity patterns normally signify the presence of some kind of quadruped animal may fire and cause the higher levels to decide that a quadruped is present, without any of the specific quadrupeds being detected (because the *pattern* of the module activations does not happen to coincide with any of the patterns corresponding to the specific quadrupeds). Thus, while a separate "abstract" representation system suggested by Andresen and Marsolek may exist, its existence remains a conjecture yet to be proved.

5 Similarity under a prescribed metric

The notion of objective similarity, discussed earlier, presupposes the existence of a unique "natural" metric on the distal shapes. **Hahn and Chater** argue that even if such a metric exists, subjects are not necessarily bound by it, and may judge as similar objects that share arcane features such as "pixel to pixel alternation" but differ in every corresponding pixel (for example, 010101 and 101010). A related point is made by **Palm**, who distinguishes between "external sensory similarity" and "functional similarity" (shared, for example, by various chairs, all of which can be sat upon, without being visually similar). **Postma, van der Herik and Hudson** draw attention to the need for invariance with respect to transformations such as translation and scaling, which leave shape unchanged, yet affect strongly what the target article calls the measurement-space appearance of the objects (as illustrated by the same pair of patterns, 010101 and 101010, one of which is a cyclic translation of the other). I am less concerned about the kind of similarity mentioned by Hahn and Chater, because I believe that it is of secondary importance in everyday perception, where it is clear what the natural metric is.[5] After all, it requires a certain sophistication on the part of the observer to realize that two pictures are the same in that they contain the same number of black pixels, or that two character strings are the same because they spell the same word, or that two sets of particle tracks in a Wilson cloud chamber are the same because they both correspond to β-decay events. Despite **Benson**'s comments, who (as far as I can gather) criticizes the lack of representation of this kind of abstract distinctions and similarities in Chorus and calls for "linguistic terms" and "additional semantic information," I prefer to keep this cart *behind* the horses.

In contradistinction to non-obvious relations (either abstract or concrete), proper representation of similarity under common transformations such as translation and scaling is a real concern, which indeed is mentioned in the target article. This issue, however, is more complicated than **Postma, van der Herik and Hudson** would have it, if only because human recognition is *not* completely invariant either to translation or to scaling, *pace* (Biederman and Cooper, 1991). Specifically, recent research shows that the degree of invariance depends on familiarity with the patterns, on global similarity between the objects to be discriminated, and on their compositional structure (Dill and Edelman, 1997). Thus, a "blanket" approach to invariance via a global transformation (even a space-variant one, as proposed by **Bonmassar and Schwartz**) does not seem to be appropriate in modeling human performance. A more credible approach is suggested by neurophysiological findings (Tovee et al., 1994; Ito et al., 1995), where cell responses, even if invariant under a certain amount of translation and scaling, only pertain to particular stimuli, refuting the possibility that invariance arises out of some global and universal mechanism (more on this below).

6 Similarity in context

As noted by **Eisler**, **van Leeuwen** and **Tovee**, similarities depend strongly on the context of the comparison (what Eisler calls "the pertinent universe" and **Jüttner** refers to as the choice of the map, which is prior to the choice of landmarks, *a propos* my analogy between categorization and navigation in a shape space). A

[5] Cf. the argument I made above in favor of objective shape spaces.

similar point has been made by Mumford (as discussed in the target article), by Tversky, and many others. As in the discussion of abstract similarities, here too I propose to treat the perception of geometric similarities defined over triplets of shapes as the basic phenomenon, and to use a model of that phenomenon (namely, the Chorus scheme) as a starting point in the development of more comprehensive and more sophisticated approaches. Specifically, as suggested in the target article, the modules comprising Chorus can be assigned salience-related weights, with the salience being determined by the context in which the comparisons are carried out. At present, it is not known to what extent this approach will be able to replicate psychophysical data on the perception of similarity; an extensive simulation study designed to address this issue is clearly required.

7 Top-down effects

Several of the commentaries question the rationale of choosing a basically feedforward architecture, such as that of Chorus, to model object recognition processes in human vision. **Grossberg**, in particular, states that "a major intellectual watershed separates feedforward models from self-organizing feedforward/feedback models." I tend to agree, but, important as it may be, the choice of architecture of the model cannot *precede* the development of a theory of the problem. This methodological issue is a source of much controversy in vision research. Marr (1982) argued that implementation of a model should follow (certainly not precede) the development of the theory. In contrast, connectionist modellers believe that the two should be allowed to interact. In the present case, the logical order is rather clear: feedback models, such as Grossberg's Adaptive Resonance Theory (ART), or Mumford's (1994) bottom-up / top-down scheme deal with the problem of categorization, which can be approached in a principled manner only following a resolution of the logically prior Problem of Representation (Cummins, 1989). This latter problem has to do with the very possibility of securing a principled relationship between the world and its representation. ART, which attempts to capture dynamically the categorical structure of a stream of data, is neutral with respect to the nature of this relationship: the data are (proximal) measurements such as images, and nothing is assumed or deduced about their distal causes.

The neutrality of ART and of the like models with respect to abstract computational-level issues such as veridicality and the Problem of Representation may suggest that they are compatible with the idea of second-order isomorphism and that they can support this mode of representation as well as (and possibly better than) the Chorus scheme. I assume this is what **Grossberg** had in mind when he wrote in his commentary that "ART models self-organize 'second-order isomorphisms' using either unsupervised learning, supervised learning, or both." There are, however, certain obstacles to be overcome before ART can be used in this manner. First, the feedback nature of ART makes the analysis of the possible relationship between distal and proximal entities more difficult than for a purely feedforward model: whereas second-order isomorphism requires merely that the distal to proximal mapping be smooth, in ART the mapping is iterated, and it is not clear what requirements it should fulfill, and what is the interaction between iteration and veridicality. Second, in the context of representing (not yet categorizing) novel stimuli, an ART-based approach such as that of the system described in (Bradski and Grossberg, 1995) is actually detrimental, because it forces the attribution of the current stimulus to one of the familiar categories (or the creation of a new category), whereas it may be preferable to represent it within the existing framework (e.g., in terms of similarities to existing categories, as it is done in Chorus). Hence my preference for feedforward models for the time being.

The turn of recognition-related tasks such as categorization comes when the Problem of Representation is solved. **Palm** doubts the ability of Chorus to perform segmentation and categorization, which, he claims, can be made much easier by allowing for top-down influences in one's model. Without such influences, Palm claims, the feedforward Chorus is essentially limited to interpolation among stored examples. Whereas Chorus indeed does not deal with the problem of segmentation, it has been shown effective in discrimination and categorization of objects unfamiliar to it, achieving about 85% correct performance over a database of

50 such objects (Edelman and Duvdevani-Bar, 1997c).

The power of interpolation among stored examples obviously depends upon the nature of the information available in each example, and on what the system does with it. In the most recent application of the Chorus scheme, the examples were entire view-spaces[6] of reference objects (Edelman and Duvdevani-Bar, 1997b; Duvdevani-Bar et al., 1998). Interpolation among these allowed the system to estimate the view-space for a novel object, and to use that estimate subsequently to carry out a variety of visual tasks (e.g., to recognize a novel view or to determine the pose of an object previously seen from only one vantage point).[7]

8 What Chorus really does

Of the commentators who raise computational issues, **Bonmassar and Schwartz** are the only ones to misunderstand thoroughly the target article. The first of their misunderstandings has to do with multidimensional scaling (MDS), which is not "a particular form of clustering" (Kruskal, 1977), but rather a kind of distance-preserving dimensionality reduction. Their second misreading of the target article is that Chorus uses MDS "to effect classification" — in fact, Chorus does not use MDS at all (which is why, incidentally, the remark that the target article does not specify a neurally plausible implementation for MDS is irrelevant). The information concerning the shape-space location of the stimulus is present in the activities of the reference-shape modules, insofar as these covary monotonically with the appropriate distal similarities. An experimenter studying the model (or the brain) can use MDS to extract that information and to embed it into a 2D space; the model itself need not do that. If there are 1000-2000 reference-object modules (of which only a very small proportion fires for any given stimulus), these can be mapped directly onto a similar number of "output lines" (leading to association or action modules), for example, by a linear matrix switch of the kind described in (Willshaw et al., 1969). One may hypothesize that the CA1 and CA3 circuits in the hippocampus (Hasselmo, 1995) constitute a "crossbar" matrix switch of this type. Note that straightforward input-output association is impossible if the dimensionality of the signal is on the order of 1000000 (as it is in the primary visual cortex, or V1) rather than 1000 (as in the inferotemporal, or IT, cortex). Thus, Bonmassar and Schwartz's statement that "there is a basic mathematical equivalence between clustering based on 'similarities' and clustering based on direct feature vector representation" is mistaken: neither clustering nor other processing (e.g., association) of the raw feature vectors would work, because of the high dimensionality, and because of the predominance of irrelevant dimensions (as noted in the target article, section 3.2).

The third misunderstanding by **Bonmassar and Schwartz**, which crops up repeatedly in their commentary, is centered on a mistaken characterization of Chorus as relying on "simple linear 'interpolation' between shifted versions of a prototype." Bonmassar and Schwartz conflate here two issues: that of multiple-view interpolation by the prototype modules, and that of translation invariance. The former is certainly not a linear phenomenon (Poggio and Edelman, 1990; Bülthoff and Edelman, 1992). In fact, the main assumption behind the use of radial basis functions (RBFs) in the implementation of the prototype modules is that of a *smooth* relationship between the effect of the variables over which the module must generalize (i.e., the viewpoint) and its required output (a constant, for a given object). As a result, the RBF mechanism can dampen the effects of any smooth transformation or deformation of the input, including the "space-variant nature of V-1 representation" stressed by Bonmassar and Schwartz, given enough exemplars to work with. Furthermore, if the visual system is capable of foveation (fixating the object to be recognized), only a limited form of translation invariance is required. Specifically, invariance has to hold over an area equal to the apparent size of the object (to support recognition when different parts of the object are fixated), rather than over the entire

[6] A view-space of an object is the low-dimensional trajectory ascribed in the measurement space by the point corresponding to a view of that object, as it undergoes a parametric transformation such as rotation in depth. The dimensionality of the view-space manifold is determined by the number of parameters in the transformation.

[7] The setting of interpolation weights in this example is, strictly speaking, a top-down operation, albeit of a different kind than the top-down processing stream in models such as ART.

visual field. This invalidates Bonmassar and Schwartz's claim that "[Chorus] would require storage of a large number of eye-position prototypes."

How can this translation invariance be achieved? At the time of writing of the target article, I believed that a space-variant mapping proposed by Schwartz and Cavanagh and developed further by Bonmassar and Schwartz may actually be part of the solution, not part of the problem. Specifically, foveation, followed by the complex logarithm mapping, followed again by a covert shift of attention (McCulloch, 1965) to the centroid of the resulting signal can result in approximate size invariance. This approach would also keep the problem of translation invariance within manageable limits, to be dealt with by mechanisms such as interpolation (Bradski and Grossberg, 1995). However, a review of the neurobiological literature (see chapter 6 in Edelman 1999), and the results of recent studies on the sensitivity of human object recognition to translation, convinced me that a global mapping (even a space-variant one) is not a good model of the primate visual system insofar as translation invariance is concerned. On the one hand, translation invariance exhibited by cells in the IT cortex is limited to receptive fields that can be rather small and is specific to the class of shapes to which the cell is tuned (Tovee et al., 1994; Ito et al., 1995). On the other hand, in human subjects the transfer of shape discrimination across just a few degrees in the parafovea is imperfect if the shapes are defined by the spatial configuration of several common parts, but is nearly perfect if the objects share the part structure and differ only parametrically (Dill and Edelman, 1997). In comparison, if translation and other invariances were the result of a global mapping, the same degree of invariance would be expected for any shape — in contradiction to the neurobiological and the psychophysical data. The upshot from this discussion is that Bonmassar and Schwartz's commentary is rather tangential to the issue at hand, and that the problem of size/translation invariance must still be considered as open.

9 Complexity and scalability

The commentary by **Eklundh and Carlsson** raises the important question of computational complexity that is not adequately treated in the target article. How many prototypes are necessary for representing the shapes of objects corresponding to the 30000 or so count nouns (Biederman, 1987) presumably known to an adult speaker of English? Eklundh and Carlsson state that "with an increasing number of categories the number of similarities to be represented grows combinatorially." This observation is true but irrelevant to the complexity of representation: Chorus aims at (1) representing the objects in terms of their similarities to a *fixed* number of reference shapes, while (2) preserving the similarities among objects to the largest possible extent. Because the dimensionality of the representation space is fixed, the real concern is whether it suffices to deal with the increasing number of objects (a problem whose size is obviously linear in the number of objects), rather than with the number of object relations such as similarities (whose number grows much faster). Experiments with an implementation of Chorus (Edelman and Duvdevani-Bar, 1997a; Edelman and Duvdevani-Bar, 1997c) indicate that the number of prototypes (reference shapes) necessary for supporting a certain level of recognition performance grows slower than the number of objects. These results, however, were obtained with only about 50 objects; further and more extensive experiments are necessary to determine whether computational complexity is a real concern here.

10 Learnability

Another computational concern — that of learnability — is raised by **Williamson**. He argues that despite a certain biological and computational appeal of the radial basis function (RBF) network used in Chorus, the standard algorithms used for training RBF networks are biologically implausible. Williamson proposes an alternative implementation for an object-specific module of the kind required by Chorus; his Gaussian ARTMAP network is related to **Grossberg**'s ART, and is endowed with an online learning algorithm. Now,

because the Chorus model is motivated by functional considerations (derived from the second-order isomorphism theory), the object-specific modules that serve as its building blocks can be implemented by a variety of architectures, as demonstrated in a related study on the extraction of veridical low-dimensional representations from image data (Intrator and Edelman, 1997). Thus, because on the algorithmic level Chorus is a generic model, the introduction of any additional architecture capable of fulfilling the required function broadens the support for the model as a whole. On the more abstract computational level and on the level of biological implementation, the situation is, however, not as simple. First, a mixture model such as Williamson's Gaussian ARTMAP inherits from ART the predisposition for single-cause explanations for the input, at the expense of impartial representation (which would let the input belong neither to this nor to that category); I have already mentioned this characteristic of ART in my reply to Grossberg's commentary. Second, as Williamson notes, Gaussian ARTMAP, being a probability mixture model, does not automatically enforce as much smoothness as may be required by the second-order isomorphism theory (unlike the RBF model, where smoothness is a major goal in the learning procedure). Furthermore, from the standpoint of biological implementation, the RBF learning algorithm is not as implausible as suggested by Williamson, especially if learning is limited to the estimation of the linear weights between the hidden layer and the output (Edelman and Weinshall, 1991). An in-depth comparison between the biological plausibility and other merits of certain versions of RBF networks on the one hand, and of versions of ART such as Gaussian ARTMAP and its EM (Expectation-Maximization) learning algorithm is beyond the scope of this paper.

11 Neurobiology

Only a few of the commentators bring lessons from neurobiology to bear on the discussion. Some of these are highly disputable, as exemplified by **Foldiak**'s statement that sensory processing in the brain involves dimensionality expansion, not reduction, presumably because "V1 contains about 100 times as many neurons as the optic nerve does, and higher visual areas maintain similar numbers." The mistake here is the assignment of one neuron per dimension. On the one hand, this *must* be the strategy of the visual system at the level of the visual input to the brain (i.e., in the optic nerve), simply because at that level there is no way in which the system can "know better" than to assume that each input line corresponds to an independent dimension. On the other hand, in the rest of the visual system the issue becomes that of effective, not nominal, dimensionality. For example, if all the input lines are perfectly correlated, then the effective dimensionality is equal to one. If the correlations between neuronal responses in the higher areas were as "surprisingly low" as described by Foldiak, it would be impossible to recover the category of the visual stimulus from mass-response data such as the fMRI signal, the optical signal measured using voltage-sensitive dyes, or the more old-fashioned evoked potential field: all these would resemble high-dimensional noise. Just as in V1 the most important dimensional characterization of the representation is in terms of the *functional architecture* (i.e., the columnar structure, the cytochrome oxidase blobs, etc., as defined by Hubel, Wiesel, Livingstone and others), so in IT the dimensionality of the representation is more likely to correspond to the number of column-like modules discovered by Tanaka et al. (Fujita et al., 1992; Tanaka, 1996), and not to the number of neurons there. The notion of functional architecture and Tanaka's findings (not cited by Foldiak) are also relevant in qualifying Foldiak's statement that the metaphor of a visual alphabet, which suggests a small set of symbols, is implausible because "sensory neurons have a huge variety of response properties." Already in V1, only a few of the possible dimensions of the image (namely, oriented energy at a subset of locations) are represented; in IT, the code is at least as low-dimensional.

Not all theoretical neurobiologists are as happy as they should be about the dimensionality reduction that occurs in the visual processing stream. In particular, **Bonmassar and Schwartz** argue (*contra* Foldiak) that vision cannot be veridical because "V1 discards more than 99.99% of the information available at the level of retinal (optical) image." This argument, however, is based on a further and rather unwarranted assumption that all 1000000 or so dimensions are required for describing the various distinctions among distal stimuli

that must be veridically represented in the first place. In addition to being pessimistic about the possibility of veridical representations, Bonmassar and Schwartz are rather conservative in their description of the current understanding of the process of recognition in the brain (they write that "we know very little about any aspect of trigger feature representation in IT at the present time"). I attribute this gloomy outlook to their somewhat outdated view of the psychophysics and the neurophysiology of object recognition. Regarding the function of IT cortex, Bonmassar and Schwartz choose to refer only to Schwartz et al., 1983, and neglect to mention the data amassed in the last decade and a half (cited in the target article). The psychophysical findings of veridical representation of shape spaces, from (Shepard and Cermak, 1973) to (Cutzu and Edelman, 1996), are ignored by them altogether. Against this background, the target article's account of the function of IT cortex may indeed appear as "deus ex machina."

Whereas much more is now known about the IT cortex than a decade or so ago, some of the crucial issues concerning the function of this area are the subject of an intense controversy. One of these is the question of the grain of the representation there: do IT cells prefer entire objects or frequently occurring object fragments in their response patterns (Tanaka, 1993)? In his commentary, **Tovee** calls the latter the "visual alphabet" hypothesis, claiming that the target article adopts it as the neural basis for the Chorus model. In fact, in the target article I adopted an opposite, holistic stance (see, e.g., section 9.3.2), with the purpose of finding out whether this route, which is computationally much more convenient than the compositional one, can lead to sufficiently powerful representations. My conclusion, supported by computational experiments (Edelman and Duvdevani-Bar, 1997a; Edelman and Duvdevani-Bar, 1997c), is that the holistic approach to representation advocated by Tovee is feasible. Additional considerations, such as the need for an explicit representation of structure in some tasks (discussed above), suggest, however, that the holistic approach should be supplemented by another one, based on object fragments or a "visual alphabet." Future experiments should determine whether an extension of Chorus along these lines (as sketched in Figure 3) is computationally feasible and biologically relevant.

12 Methodological and meta-theoretical issues

The combination of theoretical considerations with results of computational experiments and neurobiological evidence, as attempted in the target article, is especially important in connection with two issues raised by **Jüttner**. The first of these is the equivalent performance of quite different models of similarity perception in the experiments of Unzicker et al., in press. As stated in the target article (and reiterated elsewhere in this response), the computational requirements of the second-order isomorphism theory are generic, and cannot be used for specifying a particular model architecture. The reasons for preferring the Chorus scheme, and, in particular, a Chorus of RBF modules, have to do with concrete issues such as implementational parsimony, learnability, and, ultimately, biological evidence (the latter is decisive as far as the relevance of second-order isomorphism as a model of visual representation in the brain is concerned). The second remark made by Jüttner refers to Anderson's (1978) plea for "indeterminacy concerning the representations as long as the processes operating on them remain unspecified." Again, bringing to bear considerations from all the relevant disciplines, including neurobiology, reduces this indeterminacy: the presently available biological data certainly constrain the processes of vision, if not yet determine them unequivocally (in disembodied theorizing, in comparison, anything goes).

Latimer's commentary provides a crucial philosophical angle on the ideas expressed in the target article. Nevertheless, two of the meta-theoretical questions he poses along the way seem to me to obscure rather than clarify things. The first of these is the purported irrelevance of *representation*, which Latimer describes as a ternary relation, involving the thing represented (A), the thing representing (B), and an observer, to whom B represents A. It has been fashionable for some time to argue from this definition that talking about

representations is the same as postulating a homunculus.[8] The homunculus, however, need not be brought into consideration at all: B represents A *to the rest of the system*, if representation is functionally justified in Millikan's (1984) sense, and, even better, if an external intervention at the presumed locus (or "causal nexus") of representation (such as the injection of current in the appropriate place in the cortex; cf. Salzman et al., 1990) affects the situation in the manner compatible with the representational account.

My second remark on **Latimer**'s commentary concerns his questioning of the holistic nature of Chorus. For better or for worse, Chorus acquires and uses images of prototypical or reference objects without analyzing them into parts. Latimer seems to claim that this still does not mean that Chorus is holistic, because the images are ultimately composed of pixels which later play a role in computations of similarity. I see this argument (stated at much greater length in Latimer and Stevens, 1997) as a quibble because it leaves the most important thing unsaid: exactly *how* do pixels play a role in subsequent processing makes all the difference. In the case of Chorus, values of hundreds of pixels are conflated and the information in them is redistributed and transformed each time the activity of a receptive field at the measurement-space level is computed; further on, even more extensive convergence takes place. If this still qualifies Chorus as a model based on (pixel-level) parts, then something is wrong with Latimer's nomenclature.

13 And now, something completely different

The remaining two commentaries yet to be discussed come from a theoretical fringe, defined by an adherence to the arsenal of arguments from nonlinear dynamics (**Gregson**) and, in particular, from chaos theory (**van Leeuwen**). The word "fringe" here is not a facetious epithet, but a description of the relationship between nonlinear phenomena and their local approximations: the very status of the former as a generalization of the latter implies conceptual priority of the latter in the normal progress of scientific understanding. Gregson himself admits that "spaces which are metric only in a local neighborhood, but have no global properties implying constraints on monotone distance-separation relations, can be defined" (paragraph 4) and that "element-wise matchings between corresponding partitioned subsets of stimulus attributes ... can sometimes be locally reconciled with metric space mappings" (paragraph 5). Chorus, which aims at representing the local metric structure of distal similarities (see appendix B of the target article), suits these two descriptions well. It also happens to be mathematically tractable, applicable in practice, and capable of explaining a long list of results in the psychophysics and physiology of the representation of real 3D shapes.[9] Consequently, I believe that both its possible deficiencies in modeling the perception of "geometrical patterns" (Gregson's euphemism for a handful of dots or lines), and its inadequacies in solving structural analogy problems or modeling creative design (pointed out by van Leeuwen) can be safely classified as higher-order effects, to be taken care of in the next revision.

14 Conclusions

In summary of this response, I propose to distinguish between concerns grounded in technical issues such as scalability, computational complexity or compositionality, and criticism of the stance of the target paper on matters of principle, such as veridicality.

I consider the issues of compositionality and the representation of structure as technical for a simple reason: whereas the capability to represent novel objects was traditionally the prerogative of structural models based on the principle of compositionality, it is now demonstrably within reach of alternative approaches such

[8] This argument is especially popular with the neobehaviorists who wish to equate intelligence with a bundle of reflexes (Brooks, 1991).

[9] These have been cited and discussed in the target article, and will not be repeated here. In comparison, I could not discern the relevance of Gregson's only reference from neurophysiology — an fMRI study (Cohen et al., 1996) which lists cortical areas activated in a mental rotation task — to the issues he raises elsewhere in his commentary.

as Chorus. This capability thereby became a matter of *technology*, not principle. Admittedly, Chorus does not represent structure explicitly. This, however, seems to have been a small price to pay for a provably working scheme (Edelman and Duvdevani-Bar, 1997a), in a field where structural approaches such as that of (Marr and Nishihara, 1978) remained a disembodied inspiration to psychologists (Biederman, 1987), never shown to work on more than a dozen hand-labeled line drawings of stylized two-part shapes (Hummel and Biederman, 1992). Moreover, there appears to be a way to extend Chorus to deal with structure explicitly, as proposed in Figure 3. The viability of this proposal is, too, a technical issue, which should and will be resolved by computational experiments; there is no point in trying to settle it by philosophical arguments.

The issue of veridicality of representation is a harder nut to crack (which should not, perhaps, be surprising, considering that it has been around since before Plato). I believe, however, that some headway is possible even here, at least as far as the representation of shape is concerned. A full discussion of the mathematical underpinnings of this belief, centered on the concepts of natural and unique parameterization of shapes, is beyond the scope of the present paper. Suffice it to say here that philosophers would be well-advised to team up with mathematicians in dealing with these issues — unless they are satisfied with the psychologists' workaround for the problem of distal similarity, namely, the imposition of an observer bias.

References

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

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In *The mechanisation of thought processes*, pages 535–539. H.M.S.O., London.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22:577–660.

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

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

Bookstein, F. L. (1996). Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, 58:313–365.

Borg, I. and Lingoes, J. (1987). *Multidimensional Similarity Structure Analysis*. Springer, Berlin.

Bradski, G. and Grossberg, S. (1995). Fast-learning VIEWNET architectures for recognizing three-dimensional objects from multiple two-dimensional views. *Neural Networks*, 8:1053–1080.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47:139–160.

Bülthoff, H. H. and Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Science*, 89:60–64.

Carne, T. K. (1990). The geometry of shape spaces. *Proc. Lond. Math. Soc.*, 61:407–432.

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., Bookheimer, S. Y., Rosen, B. R., and Belliveau, J. W. (1996). Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain*, 119:89–100.

Cummins, R. (1989). *Meaning and mental representation*. MIT Press, Cambridge, MA.

Cutzu, F. and Edelman, S. (1996). Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Science*, 93:12046–12050.

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

Duvdevani-Bar, S., Edelman, S., Howell, A. J., and Buxton, H. (1998). A similarity-based method for the generalization of face recognition over pose and expression. In Akamatsu, S. and Mase, K., editors, *Proc. 3rd Intl. Symposium on Face and Gesture Recognition (FG98)*, pages 118–123, Washington, DC. IEEE.

Edelman, S. (1994). Biological constraints and the representation of structure in vision and language. *Psycoloquy*, 5(57). http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?5.57.

Edelman, S. (1995). Representation, Similarity, and the Chorus of Prototypes. *Minds and Machines*, 5:45–68.

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

Edelman, S. (1999). *Representation and recognition in vision*. MIT Press, Cambridge, MA.

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

Edelman, S. and Duvdevani-Bar, S. (1997b). Similarity-based viewspace interpolation and the categorization of 3D objects. In *Proc. Similarity and Categorization Workshop*, pages 75–81, Dept. of AI, University of Edinburgh.

Edelman, S. and Duvdevani-Bar, S. (1997c). Visual recognition and categorization on the basis of similarities to multiple class prototypes. A.I. Memo No. 1615, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

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

Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360:343–346.

Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition*, 52:125–157.

Hasselmo, M. E. (1995). Neuromodulation and cortical function: Modeling the physiological basis of behavior. *Behav. Brain Res.*, 67:1–27.

Hume, D. (1748). *An enquiry concerning human understanding*. The Internet. available electronically at URL http://coombs.anu.edu.au/Depts/RSSS/Philosophy/Texts/EnquiryTOC.html.

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

Intrator, N. and Edelman, S. (1997). Learning low dimensional representations of visual objects with extensive use of prior knowledge. *Network*, 8:259–281.

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, pages 69–123. Cambridge University Press, Cambridge, UK.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, 16:81–121.

Kruskal, J. B. (1977). The relationship between multidimensional scaling and clustering. In Ryzin, J. V., editor, *Classification and clustering*, pages 17–44. Academic Press, New York.

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

Le, H. (1991). On geodesics in Euclidean shape spaces. *J. Lond. Math. Soc.*, 44:360–372.

Le, H. and Kendall, D. G. (1993). The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. *The Annals of Statistics*, 21:1225–1271.

Markman, E. (1989). *Categorization and naming in children*. MIT Press, Cambridge, MA.

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

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.

McCulloch, W. S. (1965). *Embodiments of mind*. MIT Press, Cambridge, MA.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. MIT Press, Cambridge, MA.

Murphy, G. L. and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92:289–316.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23:94–140.

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

Quine, W. V. O. (1960). *Word and object*. MIT Press, Cambridge, MA.

Quine, W. V. O. (1969). Natural kinds. In *Ontological relativity and other essays*, pages 114–138. Columbia University Press, New York, NY.

Salzman, C. D., Britten, K. H., and Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346:174–177.

Schyns, P. G., Goldstone, R. L., and Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, 21:1–54.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1:54–87.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323.

Shepard, R. N. and Cermak, G. W. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, 4:351–377.

Smith, L. B., Gasser, M., and Sandhofer, C. M. (1997). Learning to talk about the properties of objects: a network model of the development of dimensions. In Medin, D., Goldstone, R., and Schyns, P., editors, *Mechanisms of Perceptual Learning*, pages 220–256. Academic Press.

Tanaka, J. and Gauthier, I. (1997). Expertise in object and face recognition. In Medin, D., Goldstone, R., and Schyns, P., editors, *Mechanisms of Perceptual Learning*, pages 85–125. Academic Press.

Tanaka, K. (1993). Column structure of inferotemporal cortex: "visual alphabet" or "differential amplifiers"? In *Proc. IJCNN-93*, volume 2, pages 1095–1099, Nagoya.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19:109–139.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., and Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, 8:282–289.

Tovee, M. J., Rolls, E. T., and Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert monkey. *J. of Neurophysiology*, 72:1049–1060.

Watanabe, S. (1985). *Pattern recognition: human and mechanical*. Wiley, New York.

Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, 222:960–962.