BEHAVIORAL AND BRAIN SCIENCES (1998) **21,** 449–498
Printed in the United States of America

Representation is representation of similarities

Shimon Edelman
Center for Biological and Computational Learning,
Department of Brain and Cognitive Sciences,
Massachussetts Institute of Technology,
Cambridge MA 02142
Electronic mail: *edelman@ai.mit.edu* *www.ai.mit.edu/~edelman*

Abstract: Advanced perceptual systems are faced with the problem of securing a principled (ideally, veridical) relationship between the world and its internal representation. I propose a unified approach to visual representation, addressing the need for superordinate and basic-level categorization and for the identification of specific instances of familiar categories. According to the proposed theory, a shape is represented internally by the responses of a small number of tuned modules, each broadly selective for some reference shape, whose similarity to the stimulus it measures. This amounts to embedding the stimulus in a low-dimensional proximal shape space spanned by the outputs of the active modules. This shape space supports representations of distal shape similarities that are veridical as Shepard's (1968) second-order isomorphisms (i.e., correspondence between distal and proximal similarities among shapes, rather than between distal shapes and their proximal representations). Representation in terms of similarities to reference shapes supports processing (e.g., discrimination) of shapes that are radically different from the reference ones, without the need for the computationally problematic decomposition into parts required by other theories. Furthermore, a general expression for similarity between two stimuli, based on comparisons to reference shapes, can be used to derive models of perceived similarity ranging from continuous, symmetric, and hierarchical ones, as in multidimensional scaling (Shepard 1980), to discrete and nonhierarchical ones, as in the general contrast models (Shepard & Arabie 1979; Tversky 1977).

Keywords: affordance; categorization; constancy; distal/proximal stimulus; features; invariance; isomorphism; mental models; perception; representation; similarity; visual shape recognition

1. Introduction and overview

1.1. Motivation

A common assumption underlying theories of vision is that a representation of the world – a geometrical replica (Marr 1982) and possibly also affordances required for a repertoire of actions (Gibson 1966) – should be delivered to the decision-making stage of an intelligent system, natural or artificial. Achieving principled correspondence between the representation and the world is a challenging philosophical and computational problem. On the philosophical level, one would like to know how representation is possible in principle. In vision, for example, one may ask: What is it about the internal state of an observer seeing a cat on a mat that makes it refer to the shape of the cat?

A traditional answer to this question has been, for a long time, *similarity.* According to this view, which originated with Aristotle, an internal entity represents an external object by virtue of resemblance or isomorphism between the two: the representation of a tomato has something of the redness and of the roundness of the real thing.

Echoes of this idea, inherited by Berkeley and Hume from the Scholasts, can be found in present-day sources: "Representation of something is an image, model, or reproduction of that thing" (Suppes et al. 1994, p. 517). Clearly, no one these days believes that a representation of a cat in an observer's brain is cat-shaped (or striped, or fluffy); rather, it is construed as a set of measurements that collectively encode the geometry and other visual qualities of a cat. Nevertheless, the philosophical foundation of the current theories of shape representation is still isomorphism: typically, it is assumed that structural (Biederman 1987) or metric (Ullman 1989) information stored in the brain reflects corresponding properties of shapes in the world, on a one to one basis.

Apart from having philosophical problems (Cummins 1989), this approach also presents a formidable computational challenge if the representation is to be veridical (i.e., if the geometry of each viewed shape is to be faithfully reconstructed from the proximal stimulus [Edelman 1998]). Given the inherent imperfections and distortions introduced by the sensory channels (as manifested in the plethora of visual illusions), it is perhaps not too surprising that human perception of shape falls short of veridicality in a variety of tasks, such as the estimation of local surface ori-



SHIMON EDELMAN is Professor of Computer Science and Artificial Intelligence at the University of Sussex at Brighton. He is the author of over 40 journal articles in the areas of computer vision, visual psychophysics and neural information processing. His book "Representation and Recognition in Vision," is due to be published by MIT Press in the fall of 1998.

© 1998 Cambridge University Press 0140-525X/98 $12.50

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

entation (Koenderink et al. 1996), local curvature (Phillips & Todd 1996), or even object size (Gregson & Britton 1990). It is certainly possible to learn fascinating lessons about the workings of the human visual system from the study of the cases in which it behaves nonveridically, nonlinearly, or downright peculiarly (Gregory 1978; Gregson 1988). Nevertheless, the central goal of this target article – understanding how representation is possible at all – is probably better pursued by considering the cases in which the representations used by the visual system do lead to veridical perception. As we shall see, lessons that can be drawn from these cases suggest a philosophically appealing and formally veridical approach to representation that turns out to be computationally feasible.

1.2. Representation by second-order isomorphism

In the processing of visual shape, some of the more striking instances of veridicality are found in experiments in which the subjects must consider *similarities among shapes* rather than the geometry of individual shapes (Cortese & Dyre 1996; Cutzu & Edelman 1996; Edelman 1995a; Shepard & Cermak 1973; Shepard & Chipman 1970). In these cases, the veridicality of the representation of the similarities among shapes is expressed in the consistency among subjects and, when tested with parametrically controlled stimuli (Cortese & Dyre 1996; Cutzu & Edelman 1996; Edelman 1995a; Shepard & Cermak 1973), in the agreement between the parameter-space patterns formed by the stimuli and their arrangement in a configuration obtained from the subject data by multidimensional scaling (more on this in sect. 7).[1] At the same time, human performance exhibits considerable departures from veridicality in perception (Koenderink et al. 1996; Phillips & Todd 1996), especially in the recognition (Jolicoeur & Humphrey 1998) of shapes (as opposed to the perception and recognition of similarities among shapes).

How do people happen to be better judges of similarities among shapes than perceivers of shape? This state of affairs should be expected if the visual system seeks a *second-order isomorphism* (Shepard 1968) between similarities among shapes and similarities among the internal representations they induce, instead of a first-order isomorphism between the shapes and their representations. Quoting Shepard and Chipman (1970, p. 2), "the isomorphism should be sought – not in the first-order relation between (*a*) an individual object, and (*b*) its corresponding internal representation – but in the second-order relation between (*a*) the relations among alternative external objects, and (*b*) the relations among their corresponding internal representations. Thus, although the internal representation for a square need not itself be square, it should (whatever it is) at least have a closer functional relation to the internal representation for a rectangle than to that, say, for a green flash or the taste of a persimmon." Essentially, this is a call for the representation *of* similarity instead of representation *by* similarity (see Fig. 1).[2]

1.3. A computational theory of veridical representation

To provide a computational basis for the representation of similarity, it is not enough merely to postulate, as J. J. Gibson did, that the relevant information is picked up or resonated to, without specifying the details of the pick-up process (Marr 1982; Ullman 1980). In the case of representation *by* similarity, the pick-up of external information amounts to a reconstruction of the visual world. Although it is quite easy to state, the reconstructionist goal is notoriously difficult to attain computationally, as illustrated by the limited success of Marr's research program in computer vision and by the calls for alternative paradigms (Aloimonos 1990; Bajcsy 1988). Fortunately, as we shall see, reconstruction is not necessary if the representation *of* similarity is taken to be the goal of the visual system.

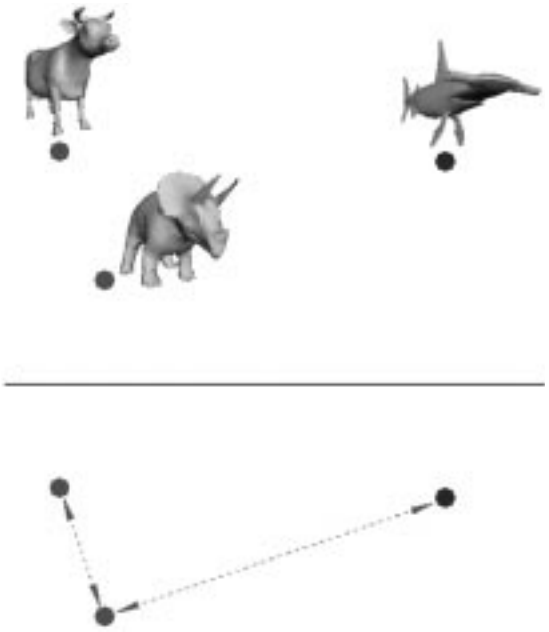


Figure 1. Clustering by natural kinds, and a representation of it that fulfills the requirement of second-order isomorphism, according to Shepard (1968). The disposition of the tokens corresponding to the three shapes in this illustration in the proximal representation space (bottom) reflects the disposition of the shapes in the distal shape space (top); the shapes of the tokens are irrelevant to their representational capacities.

Computationally, the problem of representation can be addressed on several levels (cf. Marr 1976). On the abstract level, the concern is to come up with an appropriate mathematical formulation, one that would make the problem well-posed and tractable. The idea of second-order isomorphism does in fact lead to a well-defined computational notion of representation: according to this idea, to represent a collection of objects means to reflect in a consistent manner any change an object may undergo.

By and large, this notion of representation is conceptually orthogonal to the reconstructionist approach: the tokens standing for objects need not resemble the objects themselves (see Fig. 1). Although representation by second-order isomorphism does reduce to plain reconstruction if the represented quantities correspond to distances among densely spaced points situated on the surface of an object,[3] such a reduction is unwarranted; apart from placing a heavy computational burden on the perceptual system, it serves no useful purpose. As noted by Shepard and Chipman (1970, p. 3), "it only attempts the absurdity of putting off until later the whole process of pattern recognition that must by definition precede the pivotal event in question" (i.e., the delivery of a representation capable of supporting perceptual judgment and categorization).

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

On the algorithmic level, representation by second-order isomorphism calls for ensuring that the similarities between (necessarily proximal) perceived entities correspond in some orderly fashion to the distal similarities between objects. A mechanism tuned to a particular shape provides a convenient way to estimate the similarity between the current stimulus and a reference stimulus, if its response falls off monotonically with the extent of the (distal) deviation of the current stimulus from the preferred one. This monotonic relationship between proximal and distal similarities provides the requisite algorithmic basis for veridical representation: as in nonmetric multidimensional scaling (Kruskal 1964; Shepard 1962), the rank order of the proximal similarities, being the same as the rank order of the distal similarities, allows recovery of the distal configuration of the stimuli in some underlying parametric space (Edelman 1995b).

On the implementational level, the challenge, then, is to identify a mechanism (biological or artificial) capable of responding selectively to certain shapes. A generic connectionist classifier trained on the recognition of a particular class of objects provides the requisite implementational substrate; a particular classification architecture (namely, the regularization networks of Poggio & Girosi 1990) may be preferred on the grounds of biological plausibility.

An adequate computational solution, spanning all three levels, would exert a decisive influence on the philosophical outlook on the problem of representation. At the very least, familiar dogmas would have to be reassessed and the relative merit of competing proposals reevaluated. The developments of recent years in the computational, psychophysical, and neurobiological studies of visual representation suggest that the time for such a revision has come. In the remainder of this article, I survey some of the relevant developments and suggest a way to relate them to some of the current views on the issue of representation in the philosophy of mind.

2. Representation of similarity: Some preliminaries

I now proceed to describe in detail the computational-level approach to representation outlined in the introduction. A standard answer to the central question at this level – what to represent – is, not surprisingly, "shape." The surprise comes with the realization that an alternative answer is both plausible and preferable. The approach expounded below, which is closely related to Shepard's (1968) idea of representation by second-order isomorphism, offers such an alternative answer: *represent similarity between shapes, not the geometry of each shape in itself.*

2.1. Distal shape space

To be able to discuss second-order isomorphism, one must first define the two relevant similarity functions, one for the distal (represented) shapes and the other for the proximal (representing) entities. I begin with the former.

Similarity between objects can be defined via an embedding of the objects into a metric space, where it is then determined by the distance between the points corresponding to each object. Rather than postulating a unique true distal similarity space for shapes, I propose to consider an arbitrary space of the required kind and to show later on that the exact choice of the space is not critical.[4] What should be required of such a distal *shape space?* Under second-order isomorphism, changes of shape, not the shapes themselves, are to be represented. According to this view, changing a shape corresponds to a movement of the point encoding the shape in an appropriate parameter space. To allow metamorphosis within a certain class of objects, all the members of that class must admit a *common parametrization.*

Although modern computer graphics offer a number of approaches to a common parametrization for a very wide spectrum of possible shape morphing (Galin & Akkouche 1996; Pentland & Sclaroff 1991) (see also Appendix A), it is unrealistic to expect that a structure of similarities common to extremely disparate shapes will carry over into a cognitive system (the need to judge the similarity between objects from widely disparate categories arises rarely, if ever). Different object classes may, therefore, be encoded by different sets of parameters.

To some extent, the ease with which a common parametrization can be constructed for a set of objects probably depends on the degree of their membership in the same natural kind (Quine 1969) of shapes (say, quadruped animals) or in the same artificial shape category (office tables). If any shape were equally likely (for a "medium-sized" count noun object), the burden of representing the visual world would be, I suspect, much heavier.

2.2. Proximal shape space

Defining similarity via proximity in an internal metric shape space is somewhat more problematic, as discussed by Gregson (1975, Ch. 4). The main tool at the disposal of a psychologist who wishes to show that representations of a set of stimuli can be taken to form a spatial order is multidimensional scaling (see Shepard, 1980, for a review). Using this technique, it is possible to show that, in a wide variety of perceptual tasks, subjects behave as if they represented the stimuli by distributions of points in an internal similarity space of the kind that is needed here (Nosofsky 1992; Shepard 1987).

A degree of caution is called for when interpreting this state of affairs. First, the applicability of multidimensional scaling is ultimately determined by the relevance of the resulting solution:

> Even though it is always the case that, if we are prepared to tolerate a high enough dimensionality and if we are prepared to tolerate degenerate, clustered, or lumpy configurations, we can get a spatial representation, ultimately, the criterion for accepting a representation is the sense that can be made of it, and the results that can be retrieved or predicted, by rules invariant over the space, from it. (Gregson 1975, p. 134)

Second, one should not assume too lightly that the internal similarity space is metric in the full sense used in, say, differential geometry. In that space, as pointed out by Clark (1993, p. 147), "Distances are monotonically related to similarities, but there is no presumption that sums or ratios of distances are interpretable. There may be no common unit to express distances along different axes." Fortunately, in visual shape processing these concerns seem to be largely mitigated; in section 7, we shall see that both the metric space assumption and the applicability of multidimensional scaling are justified by the human performance data in a variety of shape perception tasks.

The metric-space definition of internal similarity seems to fall short of explaining such prominent phenomena in the

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

perception of similarity as subjectivity, task dependence, and asymmetry (Medin et al. 1993; Nosofsky 1991; Tversky 1977; Tversky & Gati 1978). These shortcomings are only superficial, however. In particular, although the metric-space model makes it possible to speak about objective distal similarity (a prerequisite for a realist ontology of visual shapes), the perceptual system of the observer can warp the objective similarity space, according to his or its idiosyncrasies and to the dictates of the task (Goldstone 1994; Harnad 1987). Furthermore, similarity need not remain restricted by the symmetry it inherits from the underlying distance function; the metric-space model can be considered a starting point for a more realistic definition, of the kind proposed, for example, by Krumhansl (1978). Indeed, as I shall argue in section 5, a distance-based definition of similarity does not preclude modeling a considerable variety of similarity-related phenomena in human perception.

The possibility of a principled quantification of both the distal and the proximal shape similarity addresses the first problem faced by the proposed theory of representation: what to represent. The next question – how to communicate similarity relationships induced by a given distal shape space structure across the gap separating the world from the observer – is addressed in the following section.

3. Representation of similarity: The problem

3.1. Levels of representation of similarity

Let us now consider the process of representation as a mapping from a distal to a proximal metric shape space. One may ask, at this point, what properties the mapping must have for the image of the original shape space to qualify as its faithful representation.

3.1.1. Distinctness. The minimal requirement appears to be that the mapping be one to one, so that distinct points in the original space are mapped to distinct points in the representation space.[5] To realize the implications of limiting the representational requirements to distinctness, note that a major reason for maintaining internal representations is generalization: any system, at any point in time, will have encountered only a finite number of (labeled or rewarded) stimuli; for any other stimulus, the response will have to be generalized, based on memory traces of past experiences with related stimuli (Shepard 1987). A representation whose fidelity is limited to distinctness provides no basis for generalization because it does not contain information concerning relationships among stimuli, beyond the identity of each of them.

3.1.2. Nearest-neighbor preservation. A modicum of generalization capability is afforded by the requirement that the representation mapping preserve the nearest neighbor structure prevailing in the original space. In this case, two points that are nearest neighbors of each other before the mapping remain so after the mapping. This kind of representation preserves the structure of natural kinds, which, in turn, provides a basis for generalization (specifically, all objects more similar to some object O_1 than to O_2 will be represented as such, rather than merely as distinct both from O_1 and from O_2).

3.1.3. Full similarity spectrum preservation. If the identity of the kth nearest neighbor of each point is preserved for some $k > 1$, the resulting representation will be in closer correspondence with the original space. At the limit, when the rank order of all interpoint distances for any finite set of points is fully preserved, the representation mapping becomes a similitude. The original shape-space configuration of the points can then be recovered from the distance rank information, up to rigid motion (Borg & Lingoes 1987; Kruskal 1964; Shepard 1962; 1980). A representation that has this degree of fidelity can support categorization at a number of levels, including determination of the identity of the stimulus (see sect. 5).

This hierarchy is clearly not the only possible way to define the fidelity of the representation mapping. If the representation is to be used mainly for classification, one may require points that are separable under some parametric decision surface in the original space to remain so following the mapping (this is in contrast to distance-based requirements, which are nonparametric). For example, if points in the original shape space tend to form linearly separable clusters, one may require that the clusters remain linearly separable under the mapping. Moreover, one may also require that clusters that are not originally linearly separable become so under the mapping (Cortes & Vapnik 1995). These considerations are beyond the main concern of the present section, which is to specify a *minimal* computational basis for the processes that operate on the representation space. Still, if the original-space configuration of stimuli allows an efficient remapping that makes explicit an underlying structure of linearly separable clusters, this possibility must remain open following the mapping into the representation space. Whereas the lowest-fidelity (distinction-preserving) representation does not necessarily preserve such properties, the highest-fidelity (similarity-preserving) representation clearly does.

3.2. Distal to proximal mapping *M*

In practice, the structure of the world is never perceived directly, but always through the more or less distorting channel of the distal to proximal mapping. If that channel lets some of the original dimensions of variation of stimuli collapse, the resulting representation runs the risk of not satisfying even the distinctness requirement stated above. For example, in achromats, the perceptual dimensions of color are projected out of existence, giving rise to a perceptual system separated from that of a normal person by a gap that cannot be bridged. A more complicated situation may arise when the transformation relating two representations is invertible but highly distorting. In that case, two systems may have widely different but not unbridgeable grasps of the world. A pair of stimuli that normally appear similar to one of the systems may seem dissimilar to the other.[6]

3.2.1. Constraints on the mapping *M*. Let us consider the constraints on the distal to proximal mapping M implied by the requirement that a representation should preserve similarity ranks everywhere in the shape space. A one to one mapping with this property must be a composition of scaling with rotation or reflection (Reshetnyak 1989).[7] Thus, the requirement of global rank preservation is quite restrictive in the class of mappings it allows.

Locally, the rank preservation requirement is satisfied by any well-behaved (i.e., smooth and invertible) mapping (Cohn 1967). Such mappings are conformal, that is, they

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

preserve angles and, therefore, also the similitude of small triangles (see Appendix B). In particular, a scalene triangle formed by a triplet of points in a distal shape space will be mapped into a triangle with the same ranking of side lengths in the proximal representation space (see Fig. 1).

3.2.2. Component-wise analysis of *M*. How likely is a mapping *M*, implemented by a typical visual system, to meet these requirements for distance rank preservation? Such a mapping can be described generically as a composition of four functions, $M = f_4 \circ f_3 \circ f_2 \circ f_1$, where the first two components, f_1 and f_2, are dictated by the properties of the world and the other two constitute part of the system (see Fig. 2):

Geometry: The function $f_1(\mathbf{p})$ maps the distal parameter-space description $\mathbf{p}$ of the object into its geometry (e.g., the coordinates of the vertices of a fine mesh, suitable for rendering by a graphics system).

Imaging. The function $f_2(\mathbf{p}; \mathbf{z})$ maps the object's geometry into the image on the receptor surface of the visual system. Its dependence on the shape parameters $\mathbf{p}$ is determined by the prior action of f_1 and is written down explicitly for convenience; the dependence on the viewing conditions $\mathbf{z}$ is, however, peculiar to f_2.

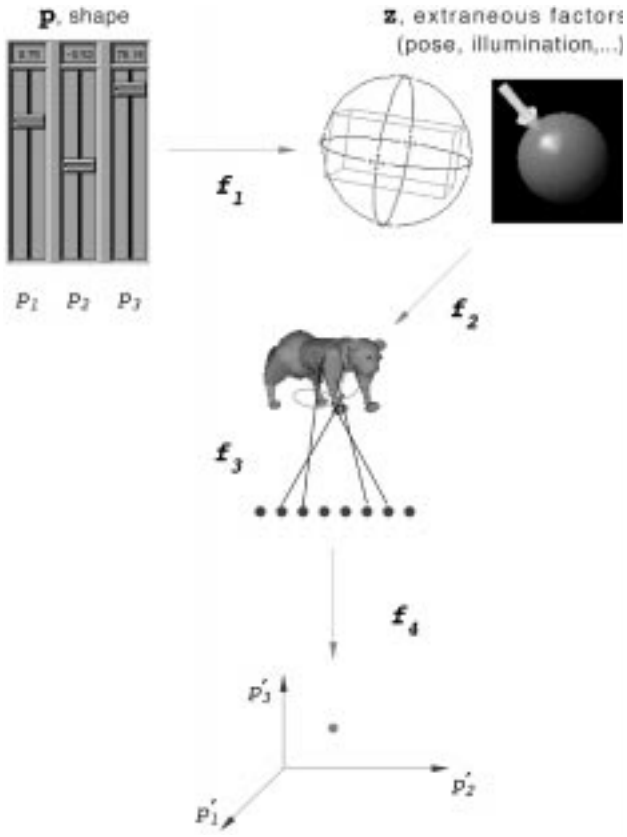


Figure 2. Components of the distal to proximal mapping, *M* (see sect. 3.2.2). In this schematic illustration, the shape of the object is determined by three parameters $\mathbf{p}$ (depicted by three "slider" controls). The appearance of the object is governed by these parameters and by variables $\mathbf{z}$ that represent factors such as orientation and illumination direction. To ensure proper representation of the original parameter space, a typical perceptual system must carry out many measurements and then reduce the dimensionality of the resulting space while getting rid of the extraneous variables $\mathbf{z}$.

Measurements. The function $f_3(\mathbf{p}; \mathbf{z})$ corresponds to the set of internal measurements performed on the image. In a typical model of biological vision, each measurement stage consists of a convolution with a number of filters, followed by the application of a nonlinearity.

Dimensionality reduction. The function $f_4(\mathbf{p})$ maps the measurement space into a low-dimensional representation of the shape space, while removing the dependence on the viewing conditions $\mathbf{z}$. The low dimensionality of the ultimate internal shape space reflects the corresponding characteristic of the distal parameter space; it is also important for reasons of computational tractability (Edelman & Intrator 1997).

Note that the second component of *M* – the view mapping, f_2 – introduces a dependence on variables $\mathbf{z}$ that are extraneous to the shape parameters to be represented. These variables encode the orientation of the object with respect to the observer, to the light sources, and to the other objects in the scene. Their influence must be counteracted by the perceptual system, through the combined action of measurement and dimensionality reduction, $f_4 \circ f_3$, to reduce the likelihood that two nearby parameter-space points (i.e., two similar shapes) are mapped into widely disparate points in the final representation space. Absolute invariance with respect to these variables is not necessary; it is only required that changes in shape space influence the measurements more strongly than view-space changes (Edelman & Duvdevani-Bar 1997b; more on this in sect. 4). Furthermore, not all the dimensions of $\mathbf{z}$ have to be treated by the same mechanism: image-plane translation can be compensated for by a covert shift of attention (Anderson & Van Essen 1987) or an overt one (such as a saccadic eye movement), variation in apparent size – by global scaling using a hard-wired mechanism (Schwartz 1985), and rotation in depth – by learning an appropriate normalizing mapping specific for each object class (Lando & Edelman 1995; Poggio & Edelman 1990).

As pointed out above, the preservation of distance ranks implies that any change in the distal parameter space must be reflected in the final low-dimensional representation (if some of the original dimensions collapse under the representation, distances between points are likely to be distorted). To ensure that as many as possible of the original dimensions of variation among the distal objects are preserved, it is worthwhile to make as many varied measurements as possible. This makes the measurement space (defined by the action of f_3) high dimensional and necessitates subsequent dimensionality reduction (through the action of f_4). In a flexible system, dimensionality reduction would have to involve learning to find informative dimensions, depending on the statistics of the input and (if available) on additional knowledge provided by the environment (for an introduction to this aspect of representation, see, e.g., Intrator 1993).

4. Representation of similarity: A solution

4.1. Representation = measurement + dimensionality reduction

We have seen that veridical representation is theoretically possible insofar as a low-dimensional subspace isomorphic

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

(in Shepard's sense) to a distal shape space may be extracted from the high-dimensional space of measurements performed by the system. This situation is illustrated schematically in Figure 3. The input to an object recognition system – an $n \times n$ image – can be considered as a point in an n^2-dimensional image or *raster* space $\mathcal{R} = R^{n^2}$ (in biological vision, one may think of the space of patterns transmitted by the optic nerve to the brain). The task of a representational system is, given a pattern $\mathbf{X} \in \mathcal{R}$, to determine the location of $\mathbf{X}$ in a proximal shape space $\mathcal{S} \subset \mathcal{R}$.

The problem of locating $\mathbf{X}$ within $\mathcal{S}$ is analogous to the problem of determining the exact location of a point on a terrain, which arises in navigation and in the preparation of topographical maps. In topography, this problem can be solved by triangulation: the location of the point is computed from bearings taken to a number of landmarks whose coordinates are known. Likewise, the location of a point in the shape space can be found from its disposition with respect to a number of reference points known to belong to the same space ("terrain"). This approach leads to a straightforward implementation of representation by second-order isomorphism, as described in the next section.

4.2. A Chorus of prototypes

The main difference between triangulation in topography and in cognitive modeling is the quantity measured to provide the location of the test point. In topography it is easy to measure direction, and in a biologically motivated model, distance (actually, a quantity monotonically related to distance). Consider a generic connectionist classifier, trained on instances of a certain shape class, that corresponds to a reference point or a prototype in the shape space. Note, first, that such a classifier can be made to learn from examples. A simple mechanism shown to be applicable, in particular, to visual object recognition is radial basis function (RBF) interpolation (Poggio & Edelman 1990); other learning frameworks such as multilayer perceptrons trained by back-propagation are also applicable. An RBF module essentially interpolates the view space (see Fig. 3) of the object on which it has been trained, starting from the exemplar views provided during training. As a result, the response of such a classifier is approximately constant over the range of the different viewing conditions.

If the classifier's response also falls off gradually and monotonically with parameter-space distance from the stimulus (the shape on which it has been trained; see Fig. 4), it can be used to pinpoint the location of the test stimulus in the shape space, by a process related to triangulation and to nonmetric multidimensional scaling (Edelman

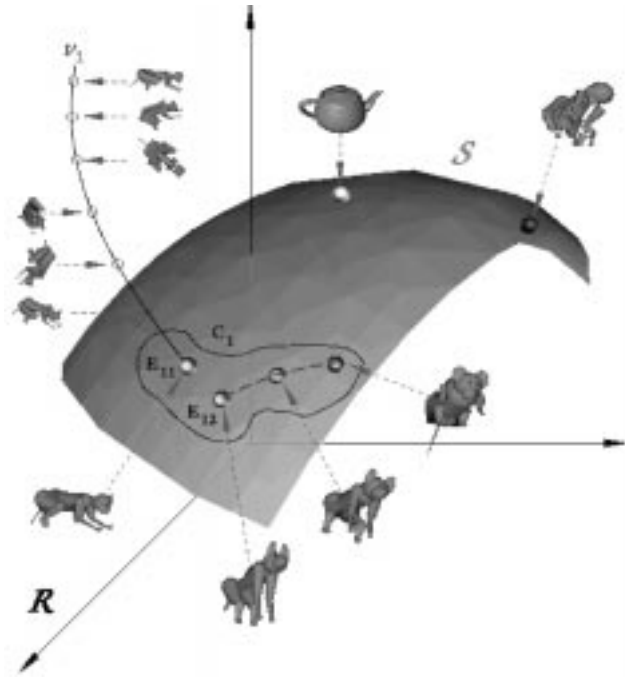


Figure 3. The image space, $\mathcal{R}$ (depicted here as 3-dimensional, to facilitate visualization), and some of its subspaces (see sect. 4.1). Two exemplars, $\mathbf{E}_{11}$ and $\mathbf{E}_{12}$, belong to the same class, C_1 (the class of four-legged animal shapes). Some of the different views of $\mathbf{E}_{11}$ are shown (marked by open circles), along with its view space, $\mathcal{V}_1$. The surface patch represents a part of the shape space $\mathcal{S}$; the view spaces of the individual objects are transverse to it. A morphing sequence originating at $\mathbf{E}_{12}$ and leading to two other shapes is illustrated by the dashed curve contained in $\mathcal{S}$. Movement toward the upper right corner of $\mathcal{S}$ corresponds to a reduction in the resemblance between the resulting image and the images of coherent looking objects.

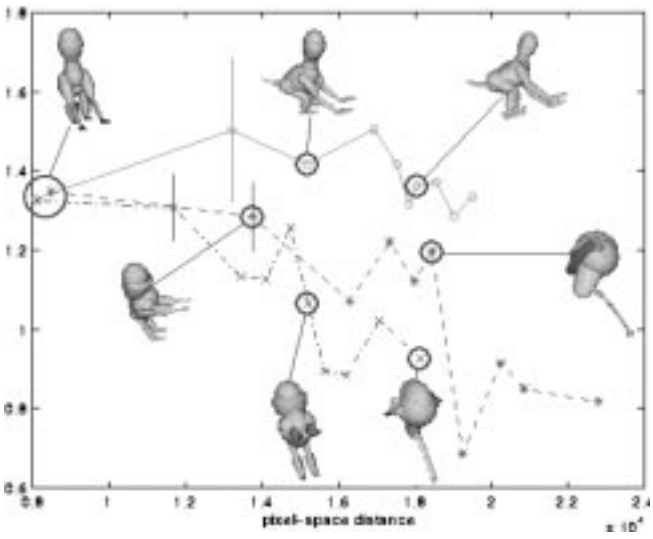


Figure 4. Three kinds of response of a radial basis function module trained on 10 random views of a parametrically defined object to stimuli differing from a reference view of that object (marked by the large circle) in three ways: (1) progressive view change, marked by ○'s; (2) by progressive shape change, marked by ×'s; (3) by combined shape and view change, marked by *'s. The points along each curve have been sorted by pixel-space distance between the test and the reference stimuli (shown along the abscissa); the units along the ordinate are arbitrary (only the relative response in the three conditions matters). Points are means over 10 repetitions with different random view-space and shape-space directions of change; a typical error bar (standard error of the mean) is shown for each curve. Note the insensitivity of the module's output to view-space changes, relative to shape-space changes.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

1995b). Note that a number of classifiers, each tuned to a different reference point, must be activated (just as in triangulation a number of landmarks must be used for each measurement).

An ensemble or a *Chorus* (Edelman 1995b) of k classifiers maps the distal shape space to a proximal representation space, $\mathcal{R}^k$. If the response of each classifier degrades gracefully with the dissimilarity between the test stimulus and the preferred shape, the entire ensemble realizes a mapping M that is smooth and regular. Thus, the distal to proximal mapping is conformal[8] and can therefore serve as a substrate for veridical representation of the original parameter space, as argued in section 3.2.1.

The main reason to use a bank of classifiers rather than raw measurement-space distances to reference points for pinpointing the current stimulus is the possibility of training a classifier to ignore those directions in the measurement space that are irrelevant to the identity of the stimulus (e.g., directions corresponding to changes in the viewpoint parameters **z**). Connectionist modelers have realized in the past that the response change caused by moving the stimulus away from a stored exemplar should depend on the direction of movement if the space of admissible exemplars is a low-dimensional manifold immersed in the representation space. Specifically, moving along a tangent to that manifold should incur a smaller generalization cost than moving in a direction perpendicular to it. This insight has been incorporated into algorithms that train for invariance by differential reinforcement of stimuli removed in the tangent and the normal directions to the target manifold (Simard et al. 1992). In Chorus, invariance is not a goal but rather a precondition that must be fulfilled for the resulting representation to be veridical. Furthermore, absolute invariance is not necessary: it suffices that the structure of categories, as defined by appropriate metrics in the low-dimensional proximal representation space, not be distorted by the irrelevant components of distance, measured along the extraneous dimensions **z**.

Training classifiers for particular stimuli, as it is done in Chorus, can be interpreted as downplaying the irrelevant dimensions by switching from the measurement-space metrics to representation-space metrics induced by the class identities (Baxter 1995). This property of the space spanned by the outputs of classifiers is important for devising better classification schemes. A typical example is vector quantization – a representational scheme in which the location of a point in a multidimensional space is coded by the identity of its nearest neighbor, chosen from a small set of points covering the space. In Baxter's (1995) canonical vector quantization, the distances to the covering points are computed according to the classifier metrics, not the raw vector space metrics.

In comparison with the canonical vector quantization, in Chorus the primary goal is representation, not classification. Accordingly, the computational question to be addressed is not whether the nearest-neighbor structure makes more sense when measured in the classifier space compared with the measurement space but, rather, to what extent the classifier-space distance structure of an arbitrary set of points reflects the corresponding structure in some low-dimensional distal parametrization. A preliminary empirical exploration indicates that classifier-space distances are indeed likely to behave in the desirable fashion (Edelman & Duvdevani-Bar 1997a). The mathematical reason behind this property of Chorus may be its relationship to a powerful method of dimensionality reduction (Bourgain 1985; Linial et al. 1994), in which points belonging to a multidimensional space are embedded into a space of much lower dimensionality while preserving to a large extent the original interpoint distances. In Bourgain's embedding of a finite set of points, the locations of the points in the new space are encoded by their distances from randomly chosen subsets of the original set, which serve as reference entities. Distances to reference points are measured in Chorus too: the response of a classifier trained on a reference pattern constitutes such a measurement, with the added advantage of tuning out the irrelevant dimensions. Thus, the use of classifiers in Chorus makes Bourgain's principle of dimensionality reduction applicable in a situation where "noise" dimensions abound.

5. Uses of similarity

In the preceding section, we saw that the output of a Chorus of classifiers constitutes, under certain conditions, a veridical representation of a distal shape space to which the individual reference classes belong. I will now examine the extent to which this representation can be put to use in modeling the perception of similarity and its role in categorization. In this section, I will show that (*a*) the responses of a number of classifiers acting in parallel can serve as a substrate for carrying out classification at different levels of categorization, depending on the way these responses are processed, and (*b*) if the salience of individual classifiers in distinguishing between various stimuli is tracked and taken into consideration depending on the task at hand, then similarity between stimuli in the representation space can be made asymmetrical and nontransitive, in accordance with Tversky's general contrast model of similarity (Tversky 1977).

5.1. Similarities at different levels of categorization

To understand the potential of the multiple-classifier representation to support shape categorization, it is necessary to consider the requirements of the relevant tasks at the different category levels.

5.1.1. Basic level. At the basic category level (Rosch et al. 1976), we are interested in the *identity* of the class $\mathbf{C}_j$ that is the closest neighbor of the stimulus **X** within the shape space $\mathcal{S}$. In some cases, the identities of several closest neighbors may be required (see Fig. 5, middle). Note that at the basic level the identities of the neighbors should suffice for categorization, whereas at the subordinate level the knowledge of their disposition relative to the stimulus in the shape space may be required.

The major obstacle to be overcome at the basic level is the dependence of the appearance of the stimulus **X** on factors such as illumination and viewpoint in addition to the category membership index j. If $\mathbf{C}_j$ is taken to correspond to the image of a member of j in some canonical orientation, the viewing conditions can be seen to span a *view* space $\mathcal{V}_j$, which is transverse to the class space $\mathcal{C}$, and pierces it at $\mathbf{C} = \mathbf{C}_j$ (see Fig. 3). A general-purpose function approximation module (Poggio & Edelman 1990)

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

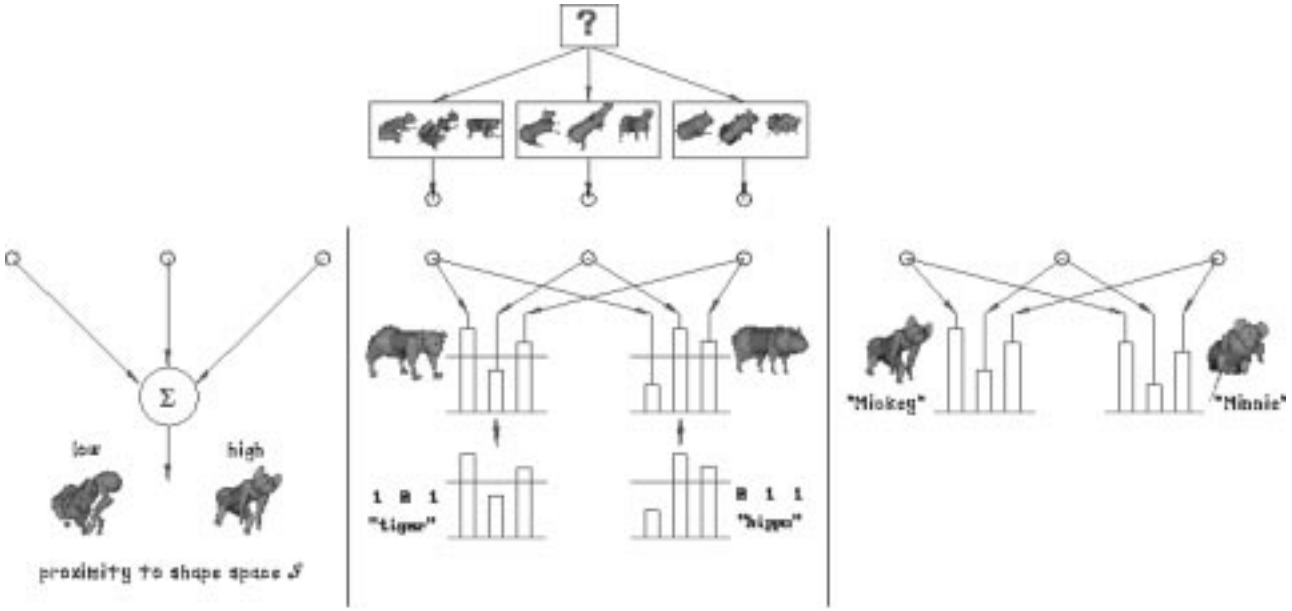


Figure 5. Using the Chorus representation (top) at different levels of categorization (see sect. 5.1). The three panes at the bottom show (left to right): a superordinate level, the basic level, a subordinate level.

trained to implement the "view normalization" mapping $T(j) : \mathrm{V}_j \rightarrow \mathbf{C}_j$ can perform basic-level categorization because its response can be made largely independent of the viewing conditions.

5.1.2. Subordinate level. At the identity level, the task is to determine the exact location of the stimulus in the shape space, rather than its nearest neighbor(s) in the collection of known class prototypes. The central problem here lies in the fine resolution that must be attained despite the residual misalignment left over from the action of the normalizing transformation *T*. This problem can be approached by learning *hyperacuity* in the instance space. In hyperacuity-related visual tasks such as vernier discrimination (Westheimer 1981), spatial resolution better than the spacing of the photoreceptors on the retina is attained by combined action of graded overlapping receptive fields (Snippe & Koenderink 1992). In shape-space localization, the response profile of each of the classifiers in Chorus defines a "receptive field" over the space $\mathcal{S}$. The vector of responses of a number of classifiers (Fig. 5, right) contains the information necessary for pinpointing the location of the stimulus within $\mathcal{S}$, as argued in section 4. Moreover, because of the graded nature of each response profile and the overlap between the different shape-space receptive fields, the localization is likely to be much more precise than what would have been possible if the responses of the classifiers were considered individually, in precise analogy to the spatial hyperacuity.

The required insensitivity of shape-space localization to viewpoint transformations stems from two sources. First, experience shows that hyperacuity can be attained despite considerable random misalignment of the stimulus as a whole, relative to its "home" or training pose, probably due to the shallow and overlapping profiles of the individual receptive fields (Poggio et al. 1992). Second, explicit training for invariance with respect to "irrelevant" transformations can complement the inherent tolerance of the receptive-field system. Importantly, once learned from examples, the normalizing transformation $T(j)$ can work even for stimuli not previously encountered by the system, provided that they belong to the same class as the examples used for training. The simplest approach here is to apply to a novel stimulus a transformation that is the average of the normalizing transformations learned for the class to which the stimulus belongs (Lando & Edelman 1995).

5.1.3. Superordinate level. Consider now two tasks at a less specific level in a hierarchy of recognition tasks. The first of these is to decide whether the stimulus **X** is the image of *some* familiar object. For this purpose, it would suffice to represent the shape space $\mathcal{S}$ as a scalar field over the image space $S(\mathbf{X}) : \mathcal{R} \rightarrow R$, which would express for each **X** its degree of membership in $\mathcal{S}$. For example, one may set $S = \max_i\{p_i\}$ (the activity of the strongest-responding prototype module), or $S = \Sigma_i p_i$ (the total activity, as in Fig. 5, right; cf Nosofsky 1988).

The second task is to characterize a superordinate-level category of the input image, and not merely to decide whether it is likely to be the image of a familiar object. This can be done by determining the identities of the prototype modules that respond above some threshold. For example, if, say, the *cat*, the *sheep*, and the *cow* modules are the only ones that respond, the stimulus is probably a four-legged animal.

5.2. Features of similarity

In Chorus, the response of each classifier p_i is, in a sense, a feature, whose value for a stimulus $\mathbf{A} \in \mathcal{R}$ is signified by the activation $p_i(\mathbf{A})$. Consider the similarity structure induced by this feature space over the universe of stimuli. With the qualifications stated in section 2, one can take the Euclidean distance between the feature vectors corresponding to two objects, $\mathbf{p}(\mathbf{A})$ and $\mathbf{p}(\mathbf{B})$, to be a default measure of the similarity between them: $s_E(\mathbf{A}, \mathbf{B})^{-1} \sim \Sigma_{i=1}^{k} [p_i(\mathbf{A}) - p_i(\mathbf{B})]^2$.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

A uniform scaling in the responses of all prototype detectors $\mathbf{p} \rightarrow c\mathbf{p}$ (as in seeing through fog) should not, however, be interpreted as a change in the shape of the stimulus object. To make the similarity insensitive to such scaling, let us define similarity by the cosine of the angle between $\mathbf{p}(\mathbf{A})$ and $\mathbf{p}(\mathbf{B})$, in the space spanned by the prototype responses (cf Ekman & Lindman 1961):

$$s_a(\mathbf{A}, \mathbf{B}) \sim \sum_{i=1}^{k} p_i(\mathbf{A}) p_i(\mathbf{B}) \doteq \langle \mathbf{p}(\mathbf{A}), \mathbf{p}(\mathbf{B}) \rangle \quad (1)$$

This definition of similarity must, however, be further modified, for at least two reasons. First, s_a is independent of context, whereas perceived similarity depends on the "contrast set" against which it is to be judged. Second, s_a is symmetric, whereas human perception of similarity appears to be asymmetric in many cases (Tversky 1977). To make s_a depend on the context, one can introduce a vector of weights, one per prototype, so that $w_i = w_i(\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots\})$. Thus, comparing **A** and **B** in two contexts, {**A, B** | **C, D, E**} and {**A, B** | **F, G, H**}, may result in different values of similarity between **A** and **B**. To model the asymmetry that frequently arises when subjects are required to estimate the similarity of some stimulus **A** to another stimulus **B,** one may observe, following Mumford (1991a), that subjects in this case behave as if they take "**A** is similar to **B**" to mean "**B** is some kind of prototype in a category which includes **A**. Thus, the stimulus input **A** being analyzed is treated differently from the memory benchmark **B**" (Medin et al. 1993; Mumford 1991a). To give **B** the required distinction, each feature $p_i(\mathbf{B})$ can be weighted in proportion to its long-term saliency $\mathrm{sal}(p_i, \mathbf{B})$ in distinguishing between **B** and the other stimuli.[9] The resulting expression for similarity, which provides for the effects of context and for asymmetry, is

$$s(\mathbf{A}, \mathbf{B}) \sim \sum_{i=1}^{k} w_i p_i(\mathbf{A}) \left(\frac{p_i(\mathbf{B})}{\mathrm{sal}(p_i, \mathbf{B})} \right) \quad (2)$$

Note that this definition has the same form as the additive clustering (ADCLUS) similarity measure of Shepard and Arabie (1979), which, in turn, instantiates Tversky's (1977) discrete contrast model of feature-based similarity. At the same time, it is built on top of a continuous metric representational substrate – the shape space spanned by proximities to prototypes. The degree of compromise between these two approaches to similarity may depend on the demands of the task at hand, via the parameters of Equation 2. At the one extreme, a Chorus-based system may behave as if it maps the stimuli pertaining to a task into a metric space, with the ensuing symmetric similarity and possible interaction among different dimensions; the other extreme may involve discrete all-or-none features, as in the examples surveyed by Tversky (1977).

6. Representation of similarity and other theories of what the brain may be doing

6.1. Making sense of novel objects

A central feature of the Chorus method is its ability to deal with novel objects (cf Fig. 7, p. 463); once these are represented in terms of similarities to some of the reference objects, they can be remembered, recognized, or otherwise processed (Edelman & Duvdevani-Bar 1997a). In theories of vision, this ability has so far been considered the prerogative of structural approaches to representation (Biederman 1987; Marr & Nishihara 1978). In structural approaches, a small number of generic primitives (such as the several dozen geons postulated by Biederman) is used along with spatial relationships defined over sets of primitives to represent a potentially unlimited variety of shapes.

In principle, even completely novel shapes can be given a structural description, because the extraction of primitives from images and the determination of spatial relationships is supposed to proceed in a purely bottom-up, or image-driven, fashion. In practice, however, both these steps have so far proved impossible to automate, for reasons that may be nonaccidental (Edelman & Weinshall 1998). The few computer vision systems currently capable of unconstrained recognition from gray-scale images either ignore the challenge posed by the problems of categorization and of representation of novel objects (Murase & Nayar 1995) or treat categorization as a by-product of recognition (Mel 1997).

In comparison with all these approaches, Chorus treats familiar and novel objects equivalently, as points in a shape space spanned by similarities to a handful of reference objects. The viability of this method is attested to by the pilot implementation of Edelman and Duvdevani-Bar (1997a), which achieved recognition performance on par with that of state of the art computer vision systems despite relying only on shape cues where other systems use shape and color or texture or both (Mel 1997; Murase & Nayar 1995; Schiele & Crowley 1996). This performance was achieved with a low-dimensional representation (10 dimensions, compared to hundreds in other systems) whose extraction from raw images did not require the problematic computation of a structural description. The use of entire reference objects as high-level features suggests a link between Chorus and the studies of similarity and generalization in feature spaces carried out by Shepard and others.

6.2. Similarity and memory-based generalization

Shepard's (1968; 1984) notion of second-order isomorphism is closest to the present one among the prior approaches to the understanding of representation. Interestingly, the computational approach to second-order isomorphism in Chorus is related to other work of Shepard – his law of generalization, which points out that the likelihood of obtaining the same response to two stimuli decreases exponentially with their separation in a psychological space, as defined, for example, by multidimensional scaling (Shepard 1987).

Shepard's law of generalization can be implemented in a straightforward manner in a connectionist framework by constructing tuned units that exhibit radially symmetric exponential decay around the location of the preferred stimulus in a feature space (Hanson & Gluck 1993; Shepard & Kannappan 1993). However, it is rather more interesting computationally to note what happens when the radial "receptive field" of an exponential-decay unit is turned into an ellipsoidal one by training the unit to ignore changes along some of the feature-space dimensions. In particular, if viewpoint-related changes in the appearance of a three-dimensional shape to which the unit is tuned come to be ignored

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

(e.g., through learning), the unit becomes a device capable of measuring the *shape-space* distance between the current stimulus and the optimal one. From here, as we saw in section 4, it is just one step to an implementation of the idea of representation by second-order isomorphism; all one need do is have a number of tuned units acting in parallel.

A computational mechanism that is particularly suitable for implementing the tuned units is the regularization network (Poggio & Girosi 1990). The simplicity of learning from examples in such networks and the relatively straightforward way they can be mapped onto the neurobiology of the brain prompted Poggio to revive the old notion of the function of the brain being largely that of a flexible memory, capable of learning from examples, and of similarity-based classification (Poggio 1990; cf Hebb 1949; Marr 1970). It is important to realize, however, that by themselves neither these nor many other learning-based approaches in the literature can solve the problem of representation as posed in the introduction. The reason is that representation is not a problem of associating (whether by learning or otherwise) a proper output with a given input, simply because what counts as "proper" differs from task to task (unless the world is represented by its replica, a choice that merely postpones the hard decisions by one stage). Thus, although different views of the same object should clearly be associated with a constant response or mapped into a canonical view (Poggio & Edelman 1990), there does not seem to be a useful universally valid specification of the proper response to a novel shape, for example, one that is a parametric blend of two familiar shapes. Consequently, in a representational scheme learning must be augmented by generalization (a process whereby useful responses can be generated for novel stimuli). Thus, Chorus adopts the basic learning strategy by letting units become loosely tuned to certain familiar shape classes (invariantly over dimensions that are irrelevant to shape, such as viewpoint), *and* it makes the existing tuned units collectively represent novel shapes in a manner that allows them to be localized in an underlying low-dimensional shape space.

6.3. The new Pandemonium

The tuned modules of which Chorus is composed can be considered as "holistic" feature detectors, where the *i*th feature of the stimulus is its similarity to the *i*th reference object.[10] The concept of a feature detector originally developed under the influence of the discovery of "bug detectors" in the frog retina (Lettvin et al. 1959); this was linked to the notion of behavior-releasing mechanisms borrowed from ethology (Barlow 1979). Its generalization to higher perceptual functions such as shape recognition was subsequently attempted. A well-known proposal for an object recognition scheme based on feature detectors – the Pandemonium (Lindsay & Norman 1977; Selfridge 1959) – consisted of a three-level hierarchy: feature demons (responsible for the detection of lines, corners, etc.), cognitive demons (responsible for entire objects), and a master demon (responsible for the recognition decision). The limited influence of the Pandemonium model on computer vision (as opposed to psychological theories of shape processing) can be traced to two shortcomings.

The first problem with the Pandemonium is the choice of all-or-none primitive features, such as edges, corners, and so on. This choice, which clearly violates Marr's (1976) principle of least commitment, is likely to lead to the loss of valuable information at an early processing stage; in the framework of section 2, it can be seen to render the distal to proximal mapping nonsmooth, lessening the likelihood of veridical representation. This situation can be remedied if probabilistic features are used instead. According to the probabilistic approach, sensory coding is "the process of preparing a representation of the current sensory scene in a form that enables subsequent learning mechanisms to be versatile and reliable" (Barlow 1990; 1994). Specifically, a representation is useful for learning if it includes records of recurring and co-occurring events. In Barlow's probabilistic Pandemonium, the response strength of a demon would be proportional to $-\log P$, where P is the probability of occurrence of the feature the demon detects (cf Intrator & Cooper 1992) .

The second problem with the Pandemonium lies at the level of decision-making (the master demon), where the stimulus is essentially described by the identity of the strongest-responding cognitive demon. This winner-take-all decision (another violation of the principle of least commitment) does provide some information about the stimulus (namely, the identity of a reference stimulus to which the current one is the most similar) while discarding much more; the representation it provides only qualifies as nearest-neighbor preserving, according to the terminology of section 3. Chorus improves on this by retaining the responses of a number of cognitive demons.

6.4. Top-down effects and representation as explanation

A number of recent theories postulate an interplay between bottom-up and top-down influences in the processing of perceptual information (Carpenter et al. 1991; 1992; Hinton et al. 1995; Mumford 1991b; 1992; Ullman 1995). Evidence from neurobiology (surveyed, e.g., by Ullman 1995) strongly suggests that information can flow from the higher to the lower cortical areas and to the thalamus. The computational role of the top-down direction of flow of information may be clarified if one assumes that the goal of perceptual processing is to find a good (e.g., minimum description length) "explanation" for the stimulus (Dayan et al. 1995; von Helmholtz 1964). Intuitively, it seems unquestionable that a human observer is capable of parsing even the most complicated scenes into the constituent objects in such a manner that every pixel eventually receives a label attributing it to this or that component. Such processing of scenes (as opposed to objects presegmented from their natural background) is a serious challenge for feedforward schemes such as Chorus.

The notion of representation as explanation does not contradict the idea that similarities between stimuli are to be represented, although in certain cases, such as scene processing, these two approaches offer largely orthogonal views on the problem of representation. On a conceptual level, the representation of a scene may well be a part of a cognitive schema (Rumelhart 1980) in which it is embodied, and may therefore be encoded in terms of similarities to related schemata. Perceptually, however, scenes that fit the same schema (e.g., city street) are too diverse for the similarities to be informative, unless the computation of similarity involves explicit alignment of corresponding components (Markman & Gentner 1993) or ignores shape de-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

tails altogether. In the latter case, only gross violations of the schema structure, such as the appearance of a sofa levitating above a sidewalk (Biederman et al. 1982), are registered.

With some ingenuity, the theory behind Chorus may actually be interpreted in terms of the idea of representation as explanation. Specifically, the activity of the reference-object modules may be taken to model the probability distribution associated with the structure of the visual stimulus. In the case of single objects, this interpretation does not seem to be too problematic: a stimulus that is attributed both to the *camel* and the *leopard* modes in the probability (or explanation) space is simply taken to be a *giraffe*. In comparison, in the case of scenes (or, more generally, of objects that share common parts, which, in turn, come to be represented independently), an explanation of the stimulus requires an account of the spatial arrangement of the components and not only of their identities. A natural approach to this problem is suggested by Riesenhuber and Dayan (1997), who propose to combine global configural and local template-like representations in a scheme that is driven by a top-down interpretation process (see also sect. 9.2).

In addition to dealing with compound objects and scenes, a Chorus-like scheme may benefit from top-down flow of information in deciding which stimuli are to be retained as reference objects, in gathering the statistical salience data for each reference object (sect. 5), and in control-related chores such as the computation of the target for the next fixation (cf Koch & Ullman 1985). By and large, however, Chorus embodies an attempt to find out how far a mostly bottom-up approach to representation can be taken. Perceiving the hidden causes of things is a feat worthy of Sherlock Holmes, and the human visual system seems to be capable of it, given enough time and a challenging task such as separating figure from ground in an underexposed photograph (Mumford 1994, p. 133). In less extreme situations, including a variety of controlled experimental conditions, the performance of a perceptual Dr. Watson ("merely" making sense of the stimulus, as detailed in the next section, instead of accounting for each and every pixel, as expected from a Holmes) seems to be a goal both worthy of pursuit and more readily attainable.

7. Perception of similarity

According to the proposed theory of representation, to make sense of a stimulus means to locate it in a low-dimensional psychological space that *(a)* is inhabited by similar stimuli and *(b)* stands in a principled relationship to a low-dimensional physical space, such as a common parametrization of the stimulus set. The main tool in testing the predictions of this theory is multidimensional scaling (MDS), a computational procedure for embedding a set of points, one per stimulus, into a metric space in such a manner that the interpoint distances conform as closely as possible to perceived similarities (proximities) between the points, as measured in some psychophysical procedure (Kruskal & Wish 1978; Shepard 1980).

7.1. Background

Normally, MDS is used in an exploratory mode, as follows. After the data are collected, the stimuli are embedded into a low-dimensional space and the resulting configuration is inspected. The analysis is considered successful if the dimensions of the (psychological, or proximal) embedding space are correlated with some (physical, or distal) variables involved in the generation of the stimuli and if the configuration of the stimuli in that space is meaningful. Among the examples of this procedure given by Shepard (1980), one finds the application of MDS to the processing of perceived similarities between Morse signals (the data were obtained by asking unskilled subjects to decide whether two consecutively sounded signals were the same or different). The two dimensions of the embedding space in that example correspond to the number of components and the proportion of dots and dashes. Another example is the near-circular arrangement of colors in two dimensions, obtained by MDS from a table of judged similarities between color patches; this result supported Newton's suggestion to represent hues by points on a circle.

In the domain of shape perception, MDS has been applied in the analysis of perceived similarities among relatively simple two-dimensional (2D) figures (rectangles, random irregular polygons), but the most spectacular results have been achieved in two studies that involved more complex shapes. In the first of these studies, subjects were requested to judge (from memory) the pairwise shape similarity of 15 of USA states (Shepard & Chipman 1970). The 2D configurations obtained by MDS were surprisingly consistent across subjects and also made sense geometrically (i.e., states of similar elongation and shape were grouped together). Shepard and Chipman pointed out that the findings of *(a)* very much the same configuration whether the states were pictorially displayed or only imagined, along with *(b)* the relationship, in both cases, between the recovered configuration and the actual cartographic shapes support the idea of a second-order isomorphism between internal representations and their corresponding external objects.

In the second study, the stimuli (2D closed contours) were created parametrically in such a way that the set of shapes formed a toroidal configuration in the parameter space (Shepard & Cermak 1973). The perceived similarities paralleled closely the parameter-space distances among the stimuli. Shepard and Cermak also reported some interesting patterns of clustering that subjects imposed on the stimuli when prompted to consider possible categorical labels (such as "fish" or "jet plane") that could be applied to the (originally unmarked) 2D contours; these findings support the assertion, made in section 2.1, that a metric-space representation of similarity does not contradict the possibility of category-related effects and, in fact, can provide the requisite substrate for the emergence of those effects.

7.2. Explorations of shape space

To obtain more direct support for the second-order isomorphism idea, it is necessary to exert control over the original configuration built into the stimuli; the success of the recovery of that configuration from subject data can then be quantified and judged statistically. This corresponds to an application of MDS in confirmatory rather than exploratory mode – an approach that can only be pursued with shapes that are generated with computer graphics and are controlled parametrically.

The veridicality of representation of parametrically de-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

fined three-dimensional (3D) shapes in human subjects has been tested in two recent studies (Cutzu & Edelman 1996; Edelman 1995a). In each of a series of experiments, which involved pairwise similarity judgment, delayed matching to sample, and long-term memory recall, subjects were confronted with several classes of computer-rendered 3D animal-like shapes arranged in a complex pattern in a common parameter space. Response time and error rate data were combined into a measure of perceived pairwise shape similarities, and the object to object proximity matrix was submitted to nonmetric MDS. In the resulting solution, the relative geometrical arrangement of the points corresponding to the different objects invariably reflected the complex low-dimensional structure in parameter space that defined the relationships between the stimulus classes (see Fig. 6).[11]

The ability of the subjects to represent the low-dimensional pattern of similarities among stimuli did not extend to nonsense objects, as indicated by the results of control experiments involving "scrambled" shapes (Cutzu & Edelman 1996). The stimuli in these experiments were obtained by translating the parts of the animal-like shapes to a common center, resulting in starlike nonsense objects. For these objects, the similarity between true and MDS-recovered configurations was consistently lower than for animal-like shapes.

Computer simulations showed that the recovery of the low-dimensional structure from image-space distances between the stimuli was impossible, as expected. In comparison, the psychophysical results were fully replicated by a Chorus-like model, patterned after a higher stage of object processing, in which nearly viewpoint-invariant representations of familiar object classes (but, presumably, not of nonsense objects as in the control experiments; cf. Bulthoff & Edelman 1992) are available; a rough analogy is the inferotemporal visual area (e.g., see Logothetis et al. 1995; Tanaka 1993; Young & Yamane 1992). As pointed out in section 4, such a representation of a 3D object can be formed easily if several views of the object are available by training a mechanism such as a radial basis function network to interpolate a characteristic function for the object in the space of all views of all objects (Poggio & Edelman 1990). A number of reference objects (in Fig. 6, the corners of the parameter space *cross)* were chosen, and a separate RBF network was trained to recognize each such object (i.e., to output a constant value for any of its views, encoded by the activities of the underlying receptive field layer; cf. Fig. 4). At the RBF level, the similarity between two stimuli was defined as the cosine of the angle between the vectors of outputs they evoked in the RBF modules trained on the reference objects (Equation 1). The MDS-derived configurations obtained with this model showed significant resemblance to the true parameter-space configurations (see Fig. 6, right).

7.3. Further predictions

The experiments mentioned above and the accompanying simulations indicate that the human visual system is capable of forming an internal representation of a set of stimuli that is second-order isomorphic to the original and, furthermore, that a simple implementation of the Chorus scheme can exhibit a comparable capability for veridical representation. Although the psychophysical findings support the idea of representation by second-order isomorphism, they are compatible with a number of possibilities of implementing the appropriate distal to proximal mapping other than Chorus. In fact, given the claim that a veridical representation is obtained generically if the mapping is smooth (sect. 2), one should look into the data for traits that are peculiar to Chorus and are not easily explained either by a reconstructionist interpretation (which seems unlikely, in view of the results of the control experiments) or by alternative mappings. Specifically, it should be possible to

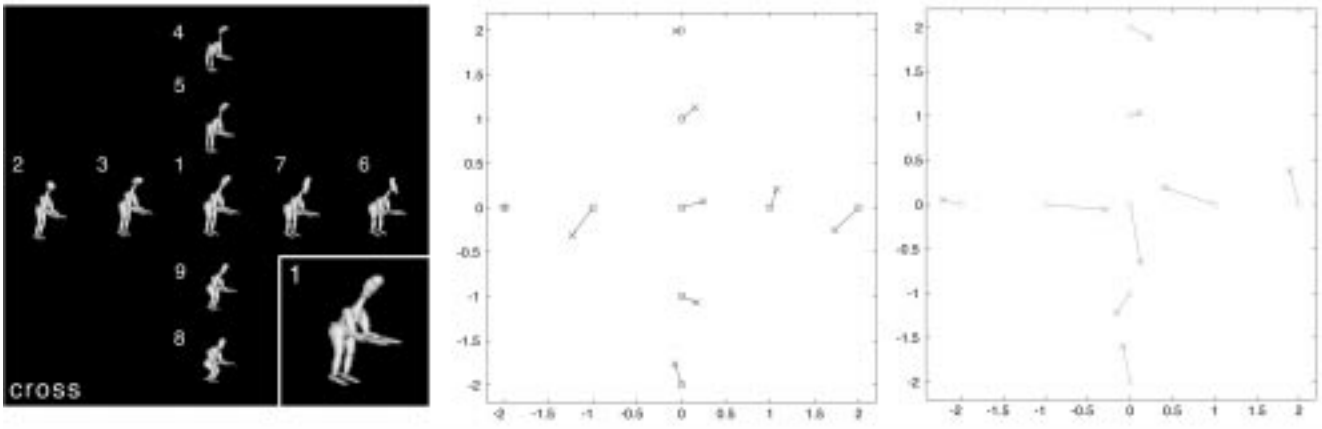


Figure 6. *Left:* Parameter-space configuration used for generating the stimuli in one of the experiments described by Cutzu and Edelman (1996a). *Middle:* 2D MDS solution for all subjects. Symbols: ○, true configuration; ×, configuration derived by MDS from the subject data, then Procrustes transformed (Borg & Lingoes 1987) to fit the true one. Lines connect corresponding points. The coefficient of congruence between the MDS-derived configuration and the true one was 0.99 (expected random value estimated by bootstrap [Efron & Tibshirani 1993] from the data: 0.86 0.03, mean and standard deviation; 100 permutations of the point order were used in the bootstrap computation). The Procrustes distance between the MDS-derived configuration and the true one was 0.66 (expected random value: 3.14 0.15). *Right:* 2D MDS solution for the RBF model. Coefficient of congruence: 0.98 (expected random value: 0.86 0.03); Procrustes distance: 1.11 (expected random value: 3.14 0.17).

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

1. Predict, for each subject, the distortion in the MDS configuration for one parameter-space pattern, given the distortion of another pattern. A better prediction is expected from the Chorus model, compared with a generic warping scheme that does not rely on distances to reference points.

2. Quantify the importance of parameter-space distances from the stimulus to preset reference points. A stronger effect of the change of these distances is expected, compared with a parameter-space movement that preserves the relative distances to the reference points; preliminary results compatible with this prediction have been reported by Edelman et al. (1996).

3. Test the nature of the reference shapes using priming. Stronger priming is expected for familiar shapes (including the so-called "impossible" objects) relative to less familiar ones. In comparison, the generic reconstructionist hypothesis (Biederman 1987), according to which representations are constructed "on the fly" by putting together universal primitives, seems to predict uniform priming for possible objects and less priming for the "impossible" ones.

8. Neurobiology of similarity

The approach to representation based on a smooth distal to proximal mapping, and its implementation by the bank of classifiers, leads to explicit predictions regarding the mechanisms of object processing at the higher levels of the primate visual system. Specifically, one expects to find there units responding preferentially to certain objects, with the response falling off monotonically with dissimilarity between the stimulus and the preferred object while staying nearly constant over different views of the preferred object (cf. Fig. 4).

Although reports of cells in the monkey inferotemporal cortex that respond preferentially to faces by now span decades (Gross et al. 1972; Perrett et al. 1989), cells tuned to general objects have been found only recently. In particular, Tanaka and his group reported the desired selectivity for specific (mostly 2D) objects in recordings from the inferotemporal (IT) cortex of anesthetized monkeys (Fujita et al. 1992; Kobatake & Tanaka 1994; Tanaka 1992; 1993; Tanaka et al. 1991). The interpretation of such findings has traditionally been hampered by the unknown nature of the optimal stimuli for the discovered cells: if a cell responds as vigorously to a brush as to a face, it cannot be properly considered a face detector. Rather than attempting the impossible (i.e., ruling out all the stimuli that the cell does *not* like), Tanaka developed an ingenious method for narrowing down the range of features that are both present in a given stimulus and effective in eliciting a response from the cell. This method has yielded the first evidence of the parallel between the functional organization of the IT cortex, where cells responding to similar shapes are arranged in columns running perpendicular to the cortical surface, and the primary visual cortex, where the columnar structure reflects orientation selectivity and ocular dominance.

Although the columnar organization of the IT cortex has been interpreted in terms of an alphabet of "elementary" features, it seems to be equally compatible with the notion that entire objects are represented, as called for by the Chorus model (Tanaka 1993). Under this interpretation, the several hundred columns that can be squeezed into the available cortical area correspond to so many classes of "reference" stimuli. If the tuning properties of the columns are such that any stimulus likely to be encountered activates a number (say, three or four) of columns, the entire system should have a considerable representational power. Moreover, this power would grow if the system were plastic enough to attune itself to novel object classes, as may indeed be the case (Kobatake et al. 1992; Rolls et al. 1989).

More recent data support this interpretation of Tanaka's findings: working with awake monkeys, Logothetis et al. (1995) reported recordings from cells tuned to specific views of 3D objects (other than faces) on which the monkey had been trained. A small proportion of the object-tuned cells found by Logothetis et al. each responded to a limited subset of the objects, irrespective of view. Together with the previous reports of a hierarchical two-stage approach to (relative) invariance in the face cells (Perrett et al. 1989), these findings suggest that a cell that responds to a certain shape nearly independently of viewpoint (corresponding to a prototype cell in Chorus) may do so by integrating the responses of several cells each of which prefers another view of the same shape, as suggested in section 4 (Edelman & Weinshall 1991; Poggio & Edelman 1990).

None of the experiments described above involved parametric manipulation of the stimulus shape – a crucial component in testing the predictions of the theory of representation proposed here. In another study, where such manipulation was attempted, the stimuli were complex, parametrically defined, periodic 2D patterns (Sakai et al. 1994). In that study, the cellular response was found to decrease monotonically with parameter-space distance between the test stimulus and the preferred pattern to which the cells were tuned. With parametrically controlled 3D stimuli, it should be possible to look for cells that behave in a manner similar to the RBF module whose response is illustrated in Figure 4. The specific predictions are as follows:

1. The cell will respond equally to different views of its preferred object, but its response will decrease with parameter-space distance from the point corresponding the shape of the preferred object (three such cells have been reported by Logothetis et al. 1995).

2. The responses of a number of cells, each tuned to a different reference object, will carry enough information to classify novel stimuli of the same general category as the reference objects.

3. If the pattern of stimuli has a simple low-dimensional characterization in some underlying parameter space (as in Fig. 6, left), it will be recoverable from the ensemble response of a number of cells, using multidimensional scaling.

9. Discussion

9.1. Similarity: The raw and the processed

In shape perception, the foremost information-processing challenge has traditionally been to achieve object constancy, that is, to perceive the object's shape despite wide variations in its visual appearance caused by changes in illumination and in the object's position with respect to the observer. The proponents of constancy observe, with Heraclitus, who pointed out that one cannot step into the same river twice, that people literally never see the same object

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

twice: objects are scaled up or down, translate, rotate, articulate, deform, are lit or shadowed, and are occluded by other objects or obscured by fog.

This observation is both true and misleading. Stressing the influence of the viewing conditions on the appearance of objects tacitly assumes that it is the exact shape of the object that a representational system should attempt to recover. However, as students of categorization know well, an intelligent agent is much better off representing an object on a number of hierarchical levels of abstraction (with the option of attending to high-resolution details, if the object happens to be present in front of the observer, and if the task demands it) than storing a high-resolution replica of the object and facing the problem of separating the chaff (pixel-level information) from the wheat (classification information) every time a new instance of that same object class is encountered.

When considered with the goal of proper representation of similarity in mind, the problem of variability of object appearance assumes a somewhat different aspect. At the computational level, instead of seeking absolute *invariance* with respect to the extraneous view-related parameters, a system can settle for mere *tolerance*, as determined by the interplay of within- and between-category similarities. At the implementational level, the availability of learning modules that can be trained to compensate for the variability in object appearance shifts the focus from the easier problems in vision (of which invariance seems to be an example) to the more challenging ones, such as making sense of objects not previously seen. The Chorus scheme, built around a theory of representation of similarity, and implemented by a bank of trainable modules tuned to reference objects, embodies both the computational and the implementational-level lessons stated above.

9.2. Some challenges

The holistic treatment of objects, adopted by the present theory, results in representations that are easily learnable from examples, but must be further worked upon if required to support inferences concerning hierarchical structure. For example, one can perceive the numerals on the face of a bent clock in Dali's *Persistence of Memory* as shapes in themselves, as well as seeing them as parts of the whole. It may be possible to address this requirement, to some extent, by coupling mechanisms that are selective for scale and retinal location with those that are selective for shape (Edelman 1994). A well-founded approach to such a coupling, built around a recently developed computational mechanism called the Helmholtz Machine (Dayan et al. 1995), has been implemented and tested (on stylized face images) by Riesenhuber and Dayan (1997).

According to the reasoning of Dayan et al., complex underconstrained perceptual tasks require intimate cooperation between bottom-up, or data-driven, processes and top-down, or expectation-driven, ones. Their arguments resemble those of other proponents of the Helmholtzian strategy, mentioned briefly in section 6.4, and are related to Grenander's notion of Pattern Theory (opposed to and complementing mere pattern recognition), as recently advocated by Mumford (1994). Returning to the example of Dali's painting, one can observe that people are aware not only of the clocks that appear in it but also of their twisted and bent shapes. Indeed, making sense of this painting may require knowledge of the possibility of objects bending without losing their identity.[12] The extension of the Chorus framework to deal with this and similar cases will have to await future work; one possible direction such a development could take would be based on the ideas of class-based processing (Lando & Edelman 1995; Moses et al. 1996), and of example-directed metamorphosis (Beymer & Poggio 1996).

9.3. Philosophical implications

Some of the philosophical implications of the Chorus scheme were mentioned briefly by Edelman (1995b); here, I discuss at greater length the place of the proposed theory in the current philosophical debate on the nature of representation, stressing its relationship to the increasingly influential idea of the world as an external memory.

9.3.1. Locke's conformity and Shepard's second-order isomorphism. In describing the implementation of Chorus (sect. 4), I have suggested that the modules tuned to specific shapes can be considered as feature detectors, spanning a feature space in which each dimension codes similarity to a particular object class. The idea of a feature detector as a basic ingredient of a representational system can be traced back to John Locke, who was among the first to fully realize the infeasibility of Aristotelian representation by resemblance. Because the firing of a feature detector is an event that is internal to the representational system, this immediately raises the problem of grounding (cf. Harnad 1990) the representation in reality:

> 1. Objection. "Knowledge placed in our ideas may be all unreal or chimerical." . . . If our knowledge of our ideas terminate in them, and reach no further, where there is something further intended, our most serious thoughts will be of little more use than the reveries of a crazy brain. . . .
>
> 2. Answer: "Not so, where ideas agree with things." (Locke 1690, Book IV, Chapter IV, sects. 1,2)

The principle on which Locke based his answer to the grounding problem is that of "conformity," postulated to prevail between the representations and their objects. As is well known, Locke distinguished between simple and complex ideas, each kind with its own grounds for conformity. Consider first the former, somewhat less problematic, kind. The argument here was that "the idea of whiteness, or bitterness, as it is in the mind, exactly answering that power which is in any body to produce it there, has all the real conformity it can or ought to have, with things without us. And this conformity between our simple ideas and the existence of things, is sufficient for real knowledge" (Locke 1690, Book IV, Chapter IV, sect. 4). In terms of feature detectors, this is a statement of belief in the availability of reliable detectors for immediate perceptual qualities.

The finding of cells tuned to well-defined features such as patterns of motion (Movshon et al. 1985; Newsome & Pare 1988), 2D shapes (Kobatake & Tanaka 1994; Tanaka et al. 1991), or faces (Gross et al. 1972; Perrett et al. 1982) supports this part of Lockean doctrine, and, in fact, suggests that it may be extended from "simple" features to entire objects. The impact of this evidence seems to have been limited by a persistent concern that the feature detectors do not "really" detect the features they happen to be tuned to (Cummins 1989; Dretske 1981; Fodor 1987).[13] Nevertheless, it has been suggested (Albright 1991) that philosophi-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

cal worries regarding the possibility of Lockean conformity in the functioning of feature detectors found in the brain should be quelled to some extent by the successful manipulation of the organism's perception of a feature through the injection of current in the vicinity of the appropriate detector pool in the cortex (Salzman et al. 1990).

More important, in light of the possibility of veridical representation of distal changes by proximal ones, as in Shepard's (1968) theory of second-order isomorphism, the philosophical lure of settling the question regarding what this or that individual feature detector "really" detects is significantly reduced. Moreover, the problematic distinction between simple and complex ideas suggested by Locke can be given up: in Chorus, the "feature detectors" can be tuned to arbitrarily complex objects, yet serve as primitives just as learnable[14] and as immediately perceivable as Locke's simple ideas. At the same time, if second-order isomorphism can be made to work, Locke's "conformity" acquires a new concrete meaning: the order and the connection of ideas is identical to the order and the connection of things.[15]

9.3.2. A new angle on compositionality. According to this view, a representational system need not possess a *combinatorial* mechanism for creating complex "ideas" out of simple ones. In vision, the hypothesis of the combinatorial structure of concepts takes the form of part-based theories of object representation (Biederman 1987; Bienenstock & Geman 1995). The debate between theories that involve dynamically bound generic parts and prototype-based theories parallels the classical dispute between Empiricist and Rationalist theories of concepts, in which the main argument against prototype-based theories is their alleged failure to support compositionality and productivity (Fodor 1981, p. 296). That argument, however, hinges on a logicist approach, which does not recognize any way of combining simple concepts into complex ones, short of logical/syntactical connectives.

In Fodor's (Rationalist) interpretation of Empiricism, a system equipped with, say, three object-specific modules, tuned to the shapes of a tuna, a cow, and a car, has only three (indivisible) visual concepts: *tuna, cow,* and *car.* In fact, however, such a system turns out to be capable of representing a variety of other shapes, some of which are quite unlike the shapes for which dedicated modules are available (cf. Fig. 7). Here and elsewhere in cognitive modeling, the logicist approach insists on indivisible primitives and logical connectives, effectively forcing a violation of the principle of least commitment. As a result, logicists cannot but predict a representational capacity that falls far short of the empiricist predictions based on coarse coding, which, in this example, means falling short of the experimental observations. In contrast, if the stimulus is compared simultaneously to a number of graded prototypes, instead of being subjected to a Pandemonium-like all-or-none logical/syntactic analysis, the productivity problem vanishes, along with the premise for Fodor's argument.

9.3.3. The world as its own representation. In a passage intended to deflect criticism from the proponents of fuzzy-set interpretation of the notion of a prototype, Fodor (1981, p. 297) admitted that prototype theories may be able to handle the combinatorics of defining the *extension* of terms, but not their *sense.* Extension, however, may be all there is to a representation.



Figure 7. Representation of novel objects in terms of similarities to familiar ones. The plot is a two-dimensional rendition (produced by MDS) of a 10-dimensional space spanned by the outputs of 10 prototypes modules in a pilot implementation of the Chorus scheme. Each point in this plot corresponds to a view of an object; views belonging to the same object cluster. The objects on which the modules have been trained are indicated by the small icons; the larger icons point to three novel test objects. Note that representations of similar objects (e.g., the quadrupeds) reside near each other; moreover, the novel quadruped (the giraffe) has been grouped with its likes. Because of the poor resolution of the front end of this system (implemented by a bank of 250 Gaussian receptive fields, each about one-tenth of the size of the stimulus images), objects that resemble each other at a coarse scale are sometimes confused (e.g., the manatee, or the sea cow, has been placed near the van; in the full 10-dimensional space, the manatee, which was a novel test object, was found to be similar to the tuna, the cow, and the automobile wagon, in that order). For a description of this system, see Edelman and Duvdevani-Bar (1997a).

Indeed, the idea of second-order isomorphism places the burden of representation where it belongs – in the world. In Chorus, the ensemble of feature detectors responds (J. J. Gibson would say resonates) to the environment (while extracting task-specific information) without reconstructing it internally. By merely mirroring proximally the similarity structure of a distal shape space, Chorus embodies the ideas of those philosophers who argued that "meaning ain't in the head" (Putnam 1988, p. 73) and that "cognitive systems are largely in the world" (Millikan 1995, p. 170), circumvents the severe difficulties encountered by the reconstructionist approaches in computer vision, and may explain the impressive performance of biological visual systems, which, in any case, appear to be too sloppy to do a good job of reconstructing the world geometrically (O'Regan 1992). Thus, in an important sense, Chorus lets the world be its own representation.

9.3.4. Qualia. If the world is its own representation, how are we to explain phenomenological qualia (Goodman 1977), such as the redness of a tomato or the shape of a pear, as perceived subjectively? The Aristotelian representation

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

by similarity solves the qualia problem appealingly by equating these perceptual qualities with the physical qualities of the corresponding percepts (i.e., the internal representations). Thus, a shift toward the view of representation *of* similarity carries with it a price. The standard version of the problem of qualia actually seems to be exacerbated: on the face of it, it is more difficult to explain the apparent richness of the perceived world if one denies that the shape of each of the constituent objects is in itself fully represented.

A partial solution to this problem is suggested by the realization that the apparent richness of the perceived world is, to a considerable extent, apparent (Dennett 1991). The source of this illusion may lie in the immediate availability of the information in the world, which acts as an "external store" (O'Regan 1992).[16] A growing number of psychophysical experiments support this view (Blackmore et al. 1995; Grimes 1995; O'Regan 1992; Pollatsek et al. 1984; Rensink et al. 1995). In these experiments, subjects are typically found to be unaware of moderate or, at times, major changes in the visual stimulus during the "blanking" period associated with a saccade or induced artificially by presenting two stimulus frames in succession with a short-duration gray-field mask interposed between them. For example, changes such as the disappearance (or appearance) of pieces of furniture in a room scene or the sudden growth (by a significant fraction) of the tallest building in a city skyline scene may go unnoticed. This suggests that under normal viewing conditions (i.e., without scrutiny) much less information than previously assumed is taken away from each scene.[17]

Although Dennett's insights do reduce the acuteness of the qualia problem to a degree, they do not appear to be able to do away with it. In particular, we are still left with the need to explain why and how a tomato looks round and red to the observer, who represents directly only the differences between tomatoes and, say, pears and oranges (as opposed to the shape and the color of the tomato). An explanation here may, however, be less elusive than commonly thought: an accomplished account of qualia in psychophysiological terms has been formulated recently around the notion of a quality space (analogous to the shape spaces discussed earlier in this paper), reconstructed from an observer's responses, using multidimensional scaling (Clark 1993). Adding to the thoughts of Carnap and Goodman a great deal of data from psychology and physiology, Clark shows that, in principle, it is not impossible to characterize a perceptual experience in objective terms, starting from relative similarity defined over tuples of objects – the very notion that constitutes the foundation of the second-order isomorphism theory (see Appendix D).

9.4. Concluding remarks

I have presented a theory of shape representation based on Shepard's notion of second-order isomorphism between the similarity structure of the internal representation space and that of the world of objects. The highlights of the proposed theory are as follows:

1. *Formal veridicality:* Representations are grounded in physical reality. This is expressed by a correspondence between proximal and distal similarities, which, under certain conditions, allows for formal veridicality.

2. *Unifying approach:* The representational substrate is a feature space spanned by similarities to reference objects. The feature-space approach offers the possibility of a smooth integration between the processing of shape and other visual dimensions. Furthermore, it provides a common representational substrate for cognitive tasks at different levels of categorization.

3. *Learnability:* Representations can be learned from examples, using well-understood computational mechanisms.

4. *Empirical support:* There is a natural mapping of representation of similarity onto well-defined neurophysiological mechanisms (ensembles of tuned units). This mapping is indirectly supported by psychophysical data, and by a functional-level simulation in an artificial neural network model.

5. *Philosophical appeal:* The proposed theory takes a clear stand on philosophical issues that have been intensely debated for a long time. It also offers an opportunity to increase the productivity of the debate, by encouraging the consideration of relevant arguments from adjacent disciplines.

To conclude, let us return to the Riddle of Representation, as posed in the introduction: By virtue of what does the representational state of a human observer seeing a cat on a mat refer to that cat (Cummins 1989)? A slightly different formulation of this riddle – what is common to two humans, a robot, and a Martian, who all see a cat on a mat? – may actually point toward a solution. It seems likely that the only thing that *can* be common to these four representational systems is the cat itself, sitting "out there" on the mat. One way to implement the idea of the world as its own representation is by constructing a system that has at its disposal tunable modules that can be trained to respond to cats or dogs or any other object. Such a system will represent a cat when it sees one (by virtue of firing of the appropriate modules) and will also be able to dream of a cat or imagine one (if the modules are made to fire in the absence of an immediate sensory stimulation). Moreover, if a selection of modules (not more than a few hundred), each tuned to a different class of stimuli, is available, the system should also be able to represent (through the response of a small subset of the modules at a time) many more stimuli, in addition to those actually stored in memory.

APPENDIX A. Formalization of distal shape spaces

The idea that objects belonging to a given natural kind can be given a common parametrization has independently led to the emergence of the concept of a shape space in a number of applied disciplines ranging from biological morphometrics to computational molecular biology. In addition, concepts related to shape space have been defined in different mathematical disciplines, such as statistics, complex analysis, and algebraic geometry.

Perhaps the most straightforward approach to the construction of a low-dimensional shape space is based on the notion of "landmarks" – fiducial points affixed to the object whose location determines the object's shape (Bookstein 1991). An orderly study of the geometry of shape spaces defined by locations of points has been initiated only recently, by Kendall (1984; 1989), who pointed out that the notion of a shape must include a specification of the transformations which, by definition, leave the shape invariant. In Kendall's shape spaces, where objects are rigid configurations of points, it is natural to define shape modulo the action of the orthogonal group of transformations (i.e., rigid motions plus reflection). From this it follows that dissimilarity between two sets of points is to be measured by the Procrustes distance, which is defined by the sum of squares of residual distances between corre-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

sponding points remaining after applying an optimal orthogonal mapping that matches one set to the other (Borg & Lingoes 1987).

An interesting consequence of allowing for a Procrustes transformation before computing shape-space distance is that it makes the topology of the space nontrivial. Consider the simple example of the space of all triangles in a plane, and a particular member of that space: the equilateral triangle. Start deforming this triangle by moving one of the vertices inward, along the perpendicular to the opposite side; this deformation corresponds to a movement of the corresponding point in the shape space. At some stage, the chosen vertex will cross over the opposite side (at which point the triangle will degenerate into a line) and will continue moving outward. Finally, an equilateral triangle will be re-formed; this triangle is a rotated version of the original one and therefore equivalent to it under the Procrustes metric. Hence, continuous movement along a straight line in the triangle-vertex space corresponds to a movement along a closed line in the shape space. It can be shown that this space is also not flat and that it contains singularities (one of which is the triangle whose three vertices coincide); furthermore, the local Riemannian metric that takes these properties into account determines a global metric that is identical to the Procrustes distance (Carne 1990; Le & Kendall 1993).

In some cases it may be desirable to define shape modulo a group of transformations that is less restrictive than the orthogonal group, or, in other words, to allow deformation.[18] In that case, a suitable framework for the definition of a shape space is provided by the theory of Riemann surfaces (Krushkal' 1979). Specifically, any two surfaces (shapes) of a given genus related by a *conformal* mapping can be considered as equivalent (belonging to the same class), with a *quasiconformal* mapping (see Appendix B) taking one shape class into another. The resulting shape space (known as the Teichmüller space) has a Riemannian metric defined by the deviation of the quasiconformal mapping from conformality (Krushkal' 1979). The Teichmuller space can be parameterized by a small set of real numbers that provide a possible coordinate system for the resulting shape space (Sundararaman 1980).

APPENDIX B. Quasiconformal mappings

In two dimensions, a mapping realized by an analytic function with a nonvanishing Jacobian in a given region is conformal there (Cohn 1967). In other words, any well-behaved function that maps a portion of the plane to itself is bound to preserve angles on a small scale (and hence also ratios of side lengths of small triangles; see Fig. 8). In higher dimensions, conformality is very restrictive. As proved by Liouville in 1850, already for $n = 3$ there are no mappings that are everywhere conformal from R^n to itself except those that are composed of finitely many inversions with respect to spheres, or Mobius transformations. These constitute a finite-dimensional Lie group that includes the group of rigid motions in R^n and is only slightly broader than that group (Reshetnyak 1989). This means that enforcing conformality in a mapping between high-dimensional spaces amounts to enforcing global isometry or global preservation of distances (by analogy with the 3D Euclidean space, mappings that satisfy this constraint are called rigid motions).

A considerably broader class of mappings emerges if the requirement of conformality is replaced by that of quasiconformality. A regular topological mapping is quasiconformal if there exists a constant q, $1 \leq q < \infty$, such that almost any infinitesimally small sphere is transformed into an ellipsoid for which the ratio of the largest semiaxis to the smallest one does not exceed q (Reshetnyak 1989). Intuitively, a conformal mapping is locally an isometry (i.e., a rigid motion; see Fig. 8); a quasiconformal mapping is locally affine (i.e., a combination of motion with shearing deformation). Under such a mapping, the ranks of distances between points are preserved approximately, on a small scale (Väisälä 1992, p. 124). The relevance of quasiconformality to the representation of real-world shapes stems from the realization that distance ranks need not be preserved globally across the entire shape space; they need only be preserved within shape classes (just as the common parametrization that is the basis for the definition of distal similarity is required to hold within, but not to extend across, the boundaries of natural kinds).

APPENDIX C. Distal to proximal mapping and the possibility of different parametrizations

Consider the effect of the geometry mapping, f_1, defined in section 3.2.2. The properties of this mapping are to be defined with re-

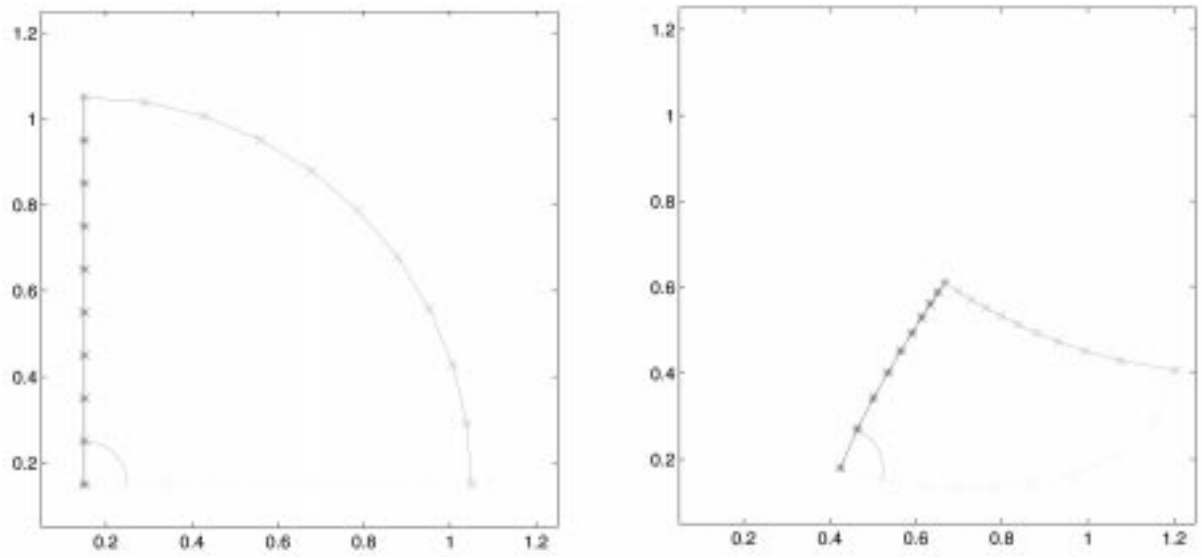


Figure 8. Illustration of the concept of conformal mapping, discussed in Appendix B. *Left:* Two similar "triangles" formed by two straight line segments and two circular arcs, all meeting at right angles. *Right:* The same two "triangles" under the action of the conformal mapping $\mathbf{z} = \sqrt{\text{atanh}(\mathbf{x})}$ (the choice of the function here is arbitrary and is merely intended to illustrate the concept of conformality). For the small triangle, the isosceles shape as well as all the angles are preserved. At a larger scale, the triangle is distorted, although the angles remain right.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

spect to a family of possible parametrizations of the distal shape space rather than with respect to some illusory true and unique parameterization. Let $\mathcal{P}$ be the set of all parameterizations related to a given one p_0 by some conformal mapping T. The set $\mathcal{P}$ is an equivalence class (Väisälä 1971); moreover, because the composition $T \circ M$ is conformal if M is, veridical representation of some $p \in \mathcal{P}$ is equivalent to the representation of any other $p \in \mathcal{P}$. Now, a conformal mapping M will give rise to a proper (i.e., second-order isomorphic) representation of object clustering under all parameterizations belonging to *some* class $\mathcal{P}_x$. The nature of that class will depend on the nature of the mapping (which can emphasize some distances among objects at the expense of others, with or without altering the distance ranks).

A system that is a product of natural selection is expected to have evolved a mapping suited to the representation of those aspects of its habitat that are most important for its survival and behavior. Thus, along with veridical representation, it is also possible that two perceptual systems implementing different mappings will have incompatible (or even conflicting) pictures of the world. Note that this effect cannot be distinguished from that of different parametrizations (discussed above).

APPENDIX D. More on qualia

A simplified version of Clark's (1993) qualia account can be formulated on the basis of the present approach, for example, by considering the redness of a ripe tomato as a counterpoint to the greenness of an unripe one and the shape of a pear as a contrast to the shape of an apple. Obviously, a shape, a color, or some other quality considered in isolation can be represented in any manner whatsoever; it is the introduction of other objects that makes representation challenging. Now, a progressive reduction in the level of illumination would force the observer to switch gradually to scotopic vision, effectively losing not only the ability to discriminate between the two tomatoes on the basis of their color but also all the color qualia. Likewise, a gradually ripening green tomato would, by any sensible account of qualia, be perceived as an equally gradual turning of the quale of greenness into that of redness.

This suggests that it may be more productive to consider qualia such as "redness versus greenness" and "pear-shape versus apple-shape" as primitive, and redness or pear-shape as derived (by a process computationally equivalent to multidimensional scaling). The "redness versus greenness" quale may then be *identified* with the feature-space support for telling apart ripe and unripe tomatoes; although this reduction seems to hold only in the context of tomato discrimination, it is easily extended to apply to any other pair of stimuli, by projecting the difference between their feature-space representations onto the paradigmatic "red versus green" distinction. In shape perception, an analogous argument can be constructed using, for example, the distinction between a pear and an apple; morphing a pear into an apple is the shape-space counterpart of the color shift induced by the ripening of the tomato in the color example. In summary, it seems sensible to accept the notion that qualia are qualia of similarities; this rules out the awkward situation in which a quale can be anything at all and points toward a potentially fruitful way to address the problematic issues associated with qualia experimentally.

ACKNOWLEDGMENTS

The title of this paper is a paraphrase of W. V. O. Quine: "To be is to be the value of a bound variable" ("On What Is," in *From a logical point of view,* pp. 1–16. Cambridge, MA: Harvard University Press, 1953). I thank A. Aertsen, G. Cottrell, F. Cutzu, P. Dayan, S. Duvdevani-Bar, N. Intrator, D. Lloyd, D. Mumford, A. O'Toole, T. Poggio, R. Shepard, and S. Ullman for useful discussions and suggestions. I am grateful to Sharon Duvdevani-Bar for Figures 4 and 7, and to Florin Cutzu for Figure 6. Supplementary material for this article (including papers cited as "in press") can be found at http://www.ai.mit.edu/~edelman/archive.html.

NOTES

1. Agreement between patterns derived from visual and haptic perceptual data has also been reported (Garbin 1990).

2. The idea of representation by second-order isomorphism has been advanced, under various guises, in a number of fields in cognitive science. Typically, the researchers in these fields take for granted the implausibility of representation by similarity, that is, by first-order isomorphism. Consequently, the theories mention merely "isomorphism," it being implied that the isomorphism holds between structures (and is, therefore, "second-order," in Shepard's terms) and not between individual entities (Gallistel 1990; Holland et al. 1986; Palmer 1978). Second-order isomorphism has been advocated recently by Cummins (1996), who calls it "The Picture Theory of Representation." This descriptor is rather unfortunate, because in vision research pictorial representations are strongly associated with the Aristotelian notion of representation by similarity, or first-order isomorphism.

3. As pointed out by S. Ullman (personal communication).

4. The problem of alternative parametrizations is addressed in Appendix C.

5. It is difficult to impose this requirement over all possible objects unless the dimensions along which objects can vary are known in advance. Thus, any perceptual system is prone to the error of omission caused by the necessarily finite set of measurements that span its internal representational space.

6. Two examples are the "other race" effect in face recognition (Brigham 1986) and the distinction between the sounds *l* and *r,* as perceived by a native speaker of Japanese versus a native speaker of English.

7. One should keep in mind that scaling and other transformations mentioned in the present context pertain to configurations formed by objects in the shape space, and not to the objects themselves.

8. Strictly speaking, it is quasiconformal (as is any diffeomorphism restricted to a compact subset of its domain; Zorich 1992, p. 133), which means that it can be considered conformal on a small scale (see Appendix B).

9. The computation of salience can be carried out by a method such as Littlestone's (1988) WINNOW.

10. The holistic nature of these features stems from the possibility of a reference shape being an entire object, rather than, say, a generic part.

11. For further details, see Cutzu and Edelman (1998). This finding has recently been replicated psychophysically in the monkey (Sugihara et al. 1998).

12. For a striking report of the malleability of object representations that emerge in a developing cognitive system, see the work of Landau et al. (1988). They found that children's assumptions about which deformations an object can undergo while retaining the same count-noun name depended on the object's appearance: deformations of furry convoluted objects (as compared with a single example view) were tolerated to a much larger extent than deformations of angular artifact-like things.

13. Compare the debate about whether the simple cells in the mammalian primary visual cortex are really line detectors or local Fourier analyzers (Hubel & Wiesel 1959; Maffei 1978).

14. By ostension, as in "this is a cat" (pointing to a cat) (see Quine 1969).

15. "Ordo et connexio idearum idem est ac ordo et connexio rerum" (Spinoza 1677, II, p. 7).

16. Compare Berkeley (1710, sect. 45): "Upon shutting my eyes all the furniture in the room is reduced to nothing, and barely upon opening them it is again created."

17. See also Biederman et al. (1974). In memory research, this point seems to be more widely accepted, in the form of the schema theories (Bartlett 1932; Rumelhart 1980). For some notes of caution, see Cavanagh (1995) and Koriat and Goldsmith (1995).

18. Consider, again, Dali's *Persistence of Memory:* We perceive the thing suspended from the tree branch as a deformed clock rather than an uninterpretable shape; this shows that there can be

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

perceptual equivalence between some shapes that are related by deformations rather than transformations.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Chorus of *k* prototypes or discord of contradictory representations?

David R. Andresen and Chad J. Marsolek

Department of Psychology, University of Minnesota, Minneapolis, MN 55455.
andr0196@maroon.tc.umn.edu; chad.j.marsolek-1@umn.edu
levels.psych.umn.edu

Abstract: The human visual system is capable of learning both abstract and specific mappings to underlie shape recognition. How could dissimilar shapes be mapped to the same location in visual representation space, yet similar shapes be mapped to different locations? Without fundamental changes, Chorus, like other single-system models, could not accomplish both mappings in a manner that accounts for recent evidence.

Edelman posits a shape representation system in which the similarity of distal stimuli corresponds with the distance between their points in an internal representation space. For example, given their distal similarity, the word forms "rage" and "*rage*" would be mapped to points that are very near in representation space. This is an interesting approach to shape representation, one that captures desirable properties of how neural network models can accomplish such representation.

However, a problem arises when one considers how such a system could accomplish both abstract and specific representation. Similar visual forms, such as "rage" and "*rage,*" should be mapped to different (albeit near) locations in situations where the differences signal something important about the inputs. However, the forms "rage" and "RAGE," which are fairly dissimilar, should be mapped to the same location in situations where the differences should be ignored, as in reading for meaning. How could Chorus accomplish the latter mapping? If the system were to map "rage" and "RAGE" together, then it would be unable to also accurately represent the physical similarity between "rage" and "*rage*" in the same representation space; the latter distal stimuli are even more similar than the former.

It appears that Edelman dismisses such abstract visual representations: "it is unrealistic to expect that a structure of similarities common to extremely disparate shapes will carry over into a cognitive system" (sect. 2.1). However, empirical evidence indicates that, in some situations, disparate shapes are mapped to a common location in visual representation space.

In a recent study (Bowers 1996), subjects first read words presented in either all lowercase (e.g., "rage") or all uppercase letters (e.g., "RAGE") and also heard other words presented auditorily. In a subsequent test phase, they identified words presented very briefly in all lowercase letters, most of which were primed from previous processing. All words were composed of letters with highly dissimilar lower- and uppercase visual structures (e.g., rage/RAGE). Most important, same-case primed and different-case primed words were identified with equal accuracy, yet both were identified significantly more accurately than auditorily primed words. Assuming that such priming reflects structural changes in the relevant shape representations, dissimilar shapes must have been mapped to the same location in representation space (to account for equivalent same- and different-case priming), and this must have been a visual representation space (to account for greater visual than auditory priming).

The problem of accounting for both abstract and specific visual representation is not limited to the domain of word forms; it applies to objects as well. For example, grand pianos and upright pianos may be relatively dissimilar distal stimuli, yet they should be mapped to the same location in visual representation space to facilitate access to common postvisual information. At the same time, two exemplar grand pianos may be similar distal stimuli, yet they must be mapped to different locations in order to distinguish them. If Chorus were to map grand and upright pianos to the same location, it would be unable to represent accurately the similarity between Edelman's grand piano and Shepard's grand piano.

A recent study indicates that both abstract and specific object representation occur in the visual system (Marsolek 1997). Subjects first viewed line drawings of objects and printed words that named other objects, in the central visual field. In a subsequent test phase, they named line drawings of objects presented very briefly in the left or right visual field. Some test objects were the same as those viewed previously, some exemplars differed from those previously viewed, some corresponded to the previously viewed words, and some had not been primed in any way. When test objects were presented directly to the left hemisphere (i.e., in the right visual field), same-exemplar primed and different-exemplar primed objects were named with equal accuracy, and both were named significantly more accurately than word-primed objects. In contrast, when test objects were presented directly to the right hemisphere (i.e., in the left visual field), same-exemplar primed objects were named significantly more accurately than both different-exemplar primed objects and word-primed objects (and accuracy in the latter two conditions did not differ).

Thus, when subsystems in the left hemisphere were given advantages in processing test objects, behavior indicated that dissimilar shapes were mapped to the same location in representation space (same- and different-exemplar priming were equivalent), and this must have been a visual-object representation space (both same- and different-exemplar priming were greater than word priming). In addition, when subsystems in the right hemisphere were given advantages in processing test objects, behavior indicated that dissimilar shapes were mapped to different locations in representation space (same-exemplar priming was greater than different-exemplar priming).

This double dissociation disconfirms theories that posit a single, undifferentiated system for visual object recognition. We hypothesize instead that two relatively independent neural subsystems operate in parallel to subserve object recognition: an abstract subsystem that operates more effectively than a specific subsystem in the left hemisphere than in the right and maps even visually dissimilar objects together when they are associated with common postvisual information, and a specific subsystem that operates more effectively than an abstract subsystem in the right hemisphere than in the left and maps visually similar objects to separate locations when they are visually and meaningfully distinctive (Marsolek & Burgund 1997). These subsystems probably use contradictory internal processing strategies, one in which parts are explicitly represented as such for an abstract subsystem (Marsolek 1995), and another in which parts are not represented explicitly for a specific subsystem (Marsolek et al. 1996). Given that the behavioral expression of abstract versus specific representation depended on which hemisphere received higher quality input before the other, at least relatively independent neural circuitry must have implemented abstract and specific representations.

These results also disconfirm theories that posit separate subsystems for abstract and specific representation while also positing that the two operate in sequence (either abstract representation precedes specific representation or vice versa). Evidence for abstract representations, without accompanying evidence for spe-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

cific representations, was obtained in left-hemisphere presentation. Evidence for specific representations, without accompanying evidence for abstract representations, was obtained in right-hemisphere presentations. Hence, the results cannot be explained by suggesting, for example, that Chorus supports specific representations and that a subsequent process like that hypothesized for superordinate-level categorization (examining only prototype modules above some threshold may indicate that the input is a four-legged animal [sect. 5.1.3]) accomplishes abstract representation. If so, specific priming should have accompanied any observation of abstract priming; the specific representation for an input should have been computed (with all prototype modules producing responses) before the abstract representation could have been computed. Yet, this prediction did not hold in left-hemisphere presentations.

Where does this leave us? Assuming separate subsystems are involved, Chorus could possibly account for the kind of processing involved in a specific subsystem, but fundamental changes would be needed for it to account for abstract representation. If Chorus as it stands accounts for specific representation, why would it be incomplete? Does an abstract subsystem underlie such a crucial component that ignoring it leaves a substantial gap in our understanding of object recognition? Apparently so. Basic- and entry-level object-naming effects alone (Jolicoeur et al. 1984; Rosch et al. 1976; also evident in Marsolek, submitted) indicate that dissimilar visual objects (e.g., grand and upright pianos) tend to be categorized at a level corresponding to the output from an abstract subsystem (e.g., piano) more often and more readily than at a level corresponding to the output from a specific subsystem (e.g., an exemplar grand piano). Explanations of both abstract and specific abilities must be developed for a complete understanding of human visual-form recognition.

Seeing wood because of the trees? A case of failure in reverse-engineering

Philip J. Benson

University Laboratory of Physiology, Parks Road, Oxford, OX1 3PT, United Kingdom. **philip.benson@physiol.ox.ac.uk** **www.physiol.ox.ac.uk/~pjb**

Abstract: Failure to take note of distinctive attributes in the distal stimulus leads to an inadequate proximal encoding. Representation of similarities in Chorus suffers in this regard. Distinctive qualities may require additional complex representation (e.g., reference to linguistic terms) in order to facilitate discrimination. Additional semantic information, which configures proximal attributes, permits accurate identification of true veridical stimuli.

The human perceptual system is adept at warping (at least) visual space through nonlinear distorting channels. Warps of sensory space are present in adaptation and sensitivity enhancement (they might also subserve mechanisms involved in remapping and plasticity). Adaptation to stimuli distorts their representational mechanism or category. Is this sufficient for rejecting representations limited to distinctness (sect. 3.1.1)? I think it is, but for a reason very different from the one provided by Edelman (sect. 9.1), who denies that similarity is distinctness. Edelman's metric is problematic because of its bias toward recovering stimulus configurations using multidimensional scaling (MDS) – a method with some inherent problems. Distinctiveness is much more powerful than Chorus can represent.

What Edelman and his colleagues believe is that the proximal code is within-category conformal (cf. sect. 2.1). Chorus works because Chorus works with its input (Fig. 7) and is permitted to disregard n^{th}-dimensional attributes of the distal token. Those attributes may, for much of machine or biological vision, suffice for representation under experimental conditions. But when an outlying stimulus is submitted to normalisation "sphering" (here, and as in component analyses), accurate classification is bound to fail. The dimensionality reduction process, $f_4 \circ f_3$ (sect. 3.2) may omit salient features or their interaction. The n^{th}-dimension might just be the one that captures the most important quality of "interestingness," as in projection pursuit methods (Friedman & Tukey 1974). As in Markovian networks, it is quite likely that latent variables are beneficial in the representational process. These may not be directly accessible for interpretation or decision making. However, such variables are veridical terms – verisimilitudinous, in fact. To impose conformal redundancy seems to presuppose prior knowledge about possible stimuli, and what Chorus represents seems to be second rate. Accessibility (even) to similarity codes seems to preclude considering that there might be a genuine difference between phenomenal consciousness (what you see, or rather, what's out there) and access consciousness (what you get). [See Block: "On a Confusion about a Function of Consciousness" BBS 18(2) 1995.] There are plenty of occasions in which no amount of retuning will give you access to those attributes. In my multidimensional cognitive hierarchy I need the trees in order to establish varieties of qualities of their wood. I might be better off referring to those distal tokens, for which I have no visual representation, in another complex relational modality, possibly linguistic. But there remain occasions on which this will fail. It may happen when there are no adequate visual or verbal characteristics of an intrinsic quality of the stimulus, yet behaviour is most definitely affected by its presence or absence.

An example will illustrate my point. Prototypes are most probably a figurative consequence of category formation; they do, however, serve as the means by which distinctness or "interestingness" is embodied. Prototypicality might thus be considered a residual feature of cognitive processing caused by the activation of similar neighbouring exemplars (general arousal, in other words). A proximal stimulus that has been (or is naturally and phenomenally) enhanced in some way should make the system exhibit facilitation along the attribute dimension(s), heightening distinctiveness. Improved within- or between-category identification of this stimulus can be explained in two ways. First, the attribute is more readily selected because the breadth of tuning or geometry of proximal modules (sects. 1.3; 2.2) is affected by the same nonlinearities in the sensory system. Second, and alternatively, the intrinsic relationship between features caused by distinctness (and not coded for proximally) enhances categorisation. The latter explanation is most likely in concessions to failed categorisation. The real fly (cf. sect. 7.3) in the ointment of Chorus theory is its inability to *retrieve* representations of distinctness.

The upshot of this is that when I know I am dealing with a tree, I might not know why it has more of "wood" (invariances notwithstanding), but knowing it does is useful and interesting and it appeals to me for some reason. Trees or cats on mats are quite unlikely to be common to internal representational systems, no matter how plastic the tuneable modules are. In the worst possible case I would not be able to tell whether what I was looking at was veridical. Although I could acknowledge the effect of a novel distinct difference on my reaction to it, without access to supporting higher-order representations, perhaps linguistic, I would not be able to save on cat food.

ACKNOWLEDGMENT

Preparation of this manuscript was supported by grants from the UK Medical Research Council, the Oxford McDonnell-Pew Centre for Cognitive Neuroscience, and the James S. McDonnell Foundation.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

Representation is space-variant

Giorgio Bonmassar[a] and Eric L. Schwartz[b]

[a]*Department of Biomedical Engineering, Boston University, Boston, MA 02215.* [b]*Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215.* **giorgio@bu.edu; eric@bu.edu**

Abstract: Under shift, caused for example by eye movement, or by relative movement of the subject or object of perception, the cortical representation undergoes very large changes in "size" and "shape." Space-variance of cortical representation rules out models that fundamentally require linear interpolation between shifted patterns (e.g., Edelman's model) or rigid shift of an invariant retinal stimulus corresponding to shift at the cortex (e.g., the shifter theory of van Essen). Recently, a computational solution of "quasi-shift" invariance for space-variant mappings has been constructed (Bonmassar & Schwartz 1997a; 1997b).

Edelman's work addresses an important gap in the computational discussion of neural representation which to date has largely been carried out on a verbal level. His position is that representation is a record of similarities to stored prototypes rather than direct representation in the form of templates, or feature vectors. Rather than learning all possible prototypes (similarities), a "small" number are stored, with interpolation of new stimuli providing generalization. Edelman uses a particular form of cluster analysis (multidimensional scaling) to effect classification. No neurally plausible means of implementing multidimensional scaling in the brain is provided, and no comparison with other similar forms of clustering, or indeed, no statistical pattern recognition in general, is supplied. It seems to us there is a basic mathematical equivalence between clustering based on "similarities" and clustering based on direct feature vector representation. We will focus instead on the issue of linear interpolation of learned prototypes, which we identify as the key contribution of this model.

Representation in the brain is expressed, we believe, in a wide variety of cortical loci. The majority of cortical visual areas are topographically organized. Spatial representations in the brain (in the form of topography and columnar spatial patterns) are themselves a form of representation, and one that obviously does not depend on "similarities" between prototypes, but is an example of direct, template-based representation.

The spatial structure of visual stimuli is represented in V-1 as a topographic map whose fidelity is sufficiently detailed to account for visual acuity. This map is approximated by the complex logarithmic (log-polar) map (Schwartz 1977). No other analytic form for V-1 topography has yet been presented, and the approximation of the two-dimensional topographic map by the simple one-parameter fit in the form of (complex) log $(z + a)$, with a representing the extent of "foveal" representation (a is roughly 0.5 degrees), is considered to be a "good" approximation by most workers in the field (e.g., see Dow et al. 1985; Tootell et al. 1985; Van Essen et al. 1984). Recently, a more general conformal map has been numerically generated from 2-deoxyuglucose data obtained from primate V-1, and the error bounds for this fit are in the range of 15–20% (Schwartz 1994). Although this numerical conformal map has no simple analytic representation, it is similar in its properties to the complex log, and we will use the complex log as a convenient way of modeling the spatial properties of early primate visual representation.

The nonlinear spatial structure of V-1 representation poses an unavoidable problem for the basis of Edelman's model: simple linear "interpolation" between shifted versions of a prototype fails because the human visual representation is strongly space-variant, and both the size and "shape" of the V-1 representation of a stimulus undergoes very large changes. This is demonstrated in Figure 1, which shows the behavior of letters under shifts or, equivalently, of eye movement. The cortical representation of these shapes is strongly distorted under shift (i.e., eye movement). That interpolation of the same letter but with different eye positions could not possibly work is evident. Edelman's model would require storage of a large number of eye position "prototypes," multiplying the combinatorial explosion already present owing to the other geometric symmetries. Of course, one can invoke (as does Edelman) the deus ex machina of IT (inferotemporal) cortex here to somehow unravel this problem, but we know very little about any aspect of trigger feature representation in IT at the present time. There is some evidence that IT trigger features are invariant under size, translation, and rotation transforms (Schwartz et al. 1983), but IT receives its representation ultimately from V-1, and therefore inherits the space-variant nature of V-1 representation.

In our lab, we have considerable experience building machine vision systems based on complex logarithmic image representation (reviewed in Schwartz et al. 1995). We can confidently state from experience that linear interpolation of view, as used by Edelman, grossly fails to allow a system built on space-variant design principles (e.g., the human brain) to function. The reader can verify this directly from Figure 1.

Recently, we have developed a computational solution to this problem by devising a new form of Fourier transform (the exponential chirp transform) that provides quasi-shift invariance, as well as size and rotation invariance that are consistent with the difficulties imposed by V-1 representation (Bonmassar & Schwartz 1997a; 1997b). Edelman has published a psychophysical study in which perfect translation invariance in human vision is called into question. We can explain this quite simply in terms of "quasi-invariance," which is defined precisely in our papers on the exponential chirp cited above. This discussion indicates a fundamental terminological and conceptual problem in the perceptual literature. Geometric invariance is not possible in a space-variant sys-

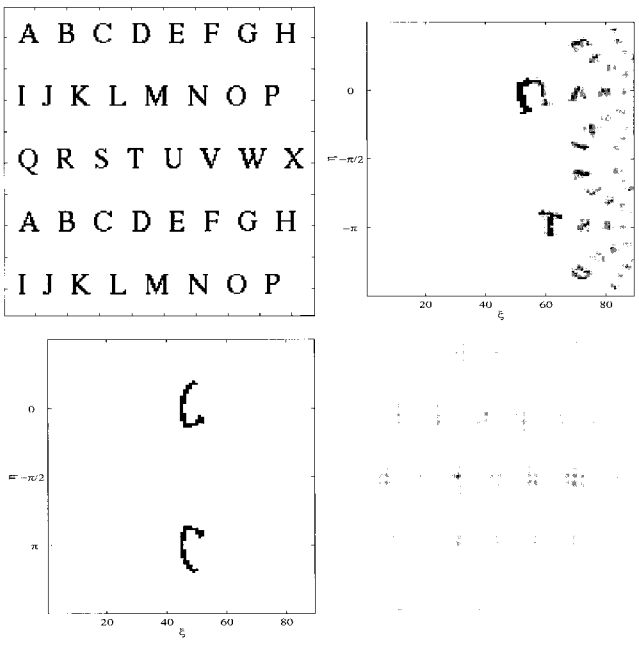


Figure 1 (Bonmassar & Schwartz). The result of applying the space-dependent cross-correlation to a (197 × 194) image of letters with an image of the letter "T" at the fixation point. *Top left:* the original image of letters; *top right:* its space-dependent representation; at *bottom left:* the log-polar (space-dependent) image of the letter "T" (split by the vertical meridian into a "left" and "right" hemisphere segment. These two last space-dependent images are used by the ECT (Exponential Chirp Transform) algorithm to compute the space-dependent cross-correlation, as shown in the bottom right of the figure. Clearly visible is the sharp peak located in the position of the letter "T" in the original image (cortical) space.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

tem. Similarly, there is (and can be) no "veridical" representation in the brain, since V-1 discards more than 99.99% of the information available at the level of retinal (optical) image (Rojer & Schwartz 1990).

The symptoms associated with space-variance in human vision provide a fatal problem for models based on simple linear interpolation (Edelman) or simple "linear shift" to account for the problem of eye movement (the Olshausen-Anderson-Van Essen "shifter theory" cited by Edelman as a solution to the problems introduced by eye movement). Linear shift, or linear interpolation, cannot be invoked as a modeling tool in the primate visual system because of the strongly nonlinear nature of V-1, and later cortical representation. Linear shift of a cortical pattern does not correspond, in an isomorphic sense, to linear shift of a retinal pattern! Models that require this feature (e.g., the shifter theory, the linear interpolation aspects of Edelman's model) cannot be correct.

We are constantly surprised that models purporting to explain biological vision ignore the most basic spatial structure of the visual system. However, it is always useful to be able to falsify models, particularly in fields such as this, in which most models are "not even wrong." For the present, we can assert, with strong confidence, that models depending fundamentally on the ability to shift linearly or interpolate cortical representations of visual stimuli, and that as a result ignore the space-variant structure of the primate visual system, are, to paraphrase W. Pauli, "even wrong."

Distal similarity, shape referents, subjective world, and redundancy

Hannes Eisler

Department of Psychology, Stockholm University, 106 91 Stockholm, Sweden. **he@psychology.su.se www.psychology.su.se**

Abstract: The concept of distal similarity that plays a crucial role in Edelman's theory of representation is called into question in this commentary on theoretical as well as empirical grounds. A possible confusion between shape and (knowledge of) its referent, the problem of the subjective world, redundancy, and large individual differences in subjective space encountered in contrived universes are discussed.

I concur with Edelman that the recognition of objects, based on their shape, builds on similarity; however, I can see some problems with some of his arguments. First, there is a problem with the concept of "distal similarity," between shapes or between other stimuli, colors, for example. As pointed out in Eisler (1960), similarity refers to psychological attributes; for pairs of stimuli, any definition is arbitrary. To give substantive content to the concept of distal similarity, it should be possible to measure it without recourse to asking observers.

I was surprised to find "jaggedness" to be a pertinent property of the shapes of states of the United States in Shepard and Chipman's (1970) experiment. In an experiment (Eisler & Roskam 1977b) on the similarity of patterns consisting of pairs of luminous points positioned in the first quadrant, we expected two dimensions from the physical arrangement, either extension in x- and y-directions, or vector length and angle. But a third dimension emerged: "cornerness" for the points that were farthest away from the origin. Of course, this attribute of percept space could be considered a property of stimulus space, contributing some to the distal similarities, but how would we know in advance?

A third example is a study of the visual perception of texture such as the surface of bricks or cloth (Eisler & Edberg 1982). An attempt to use "stylized textures" (regular patterns of circular areas varying in number, diameter, and distance) as easily quantifiable referents to real textures, suitable for a texture chart for architects, failed. The attributes obtained from multidimensional scaling (MDS) or similarity judgments of real textures could not be captured by the stylized patterns. Proponents of the idea of distal similarities seem to have fallen into the "physicalistic trap," clinging to physical measures rather than psychological (Eisler 1982). I assume that the perception (or experience) of similarity (note that I do not use "subjective similarity," since objective similarity does not exist) is direct, perhaps using a "smart" perceptual mechanism Runeson (1977) by applying Landahl's (1945) physiological model. This idea would also be consistent with the findings of von Grünau et al. (1994).

The second problem is a confusion of pure shape with (knowledge of) its referent. Edelman mentions "quadruped animals," and depicts in Figure 1 the distance between a cow and a triceratops (the tail of which I sadly miss, by the way) as smaller than that between either and the third legless animal. But are these distances determined by the shapes alone or by presence and absence of quadrupedality? In that connection I would like to note the importance of the (subjective) world which is only superficially mentioned by Edelman as a class or category. The universe may be clear from the sample used or defined by instruction. Similarities depend strongly on the pertinent universe (see Sjöberg & Thorslund 1979).

It may be worth mentioning the rather amazing stability of subjective space; it is not only MDS that can reveal its structure. In the above-mentioned experiment on luminous points (Eisler & Roskam 1977a; 1977b), five different estimation instructions (of which one was similarity) for the same stimulus set "tapped" the space with congruent results (cf. also Eisler 1982). This does not demonstrate conclusively whether the mapping is inherently built on similarities, but in any case the space could be constructed from similarity judgments. Stability does presuppose a "natural" universe, however (cf. Eisler 1982). The rather contrived circle-and-spoke figures used by Shepard (1964) showed large individual differences in their isosimilarity contours; certain subjects not only collapsed dimensions, as mentioned by Edelman, but attended to only one of the two dimensions (Eisler & Lindman 1990).

Finally, let's return to the tailless triceratops. I could see what it was from the head alone; that it had four legs I knew before I saw them. This is a problem of redundancy of shapes: How much of a given shape is necessary for its placement in a subjective space, based on similarity? This calls to mind Thurber's drawing of a room with hunting trophies – heads of deer and antelopes – on the wall, all covered with small distinct patterns. "They were shot by George's uncle – the one that lost his mind." But one could recognize the animals.

Appearance is more than shape, illumination, and pose

Jan-Olof Eklundh and Stefan Carlsson

Computational Vision and Active Perception Laboratory, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, 100 44 Stockholm, Sweden. **joe@nada.kth.se, stefanc@nada.kth.se www.nada.kth.se/**

Abstract: Although we find the idea of representation by similarities attractive as such, we have two main objections to the specific proposal of Edelman. First, he does not consider complexity issues in terms of storage and speed of recall for recognition. Related to this, the appearance of objects depends on far more factors than just shape, illumination, and pose. This requires an intermediate shape abstraction process that extracts category-specific shape properties from the mixed appearance of images.

Edelman argues for visual representations supporting second-order isomorphisms: "i.e., correspondence between distal and proximal similarities among shapes, rather than between distal shapes and their proximal representations" (Abstract). This idea and its implications are attractive (limited need for scene reconstruction, the world serves as its own representation).

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

The target article deals primarily with the issue or representation, but as is pointed out in section 1.3, this should be considered with regard to the problem of pattern recognition and categorization. It is when we consider Edelman's suggestions from that perspective that we come up against its limitations.

First of all, recognition can be trivially regarded as establishing a correspondence between incoming stimuli and stored representations, as described in the pattern recognition paradigm. In particular, visual stimuli images can be represented in the form of a high-dimensional vector of image intensities. Recognition is then equivalent to partitioning this high-dimensional space and associating the incoming stimuli to their corresponding sectors. That the appearance of a shape depends on various external factors can then be accounted for in principle by extending the partition for that specific shape. The dimensionality and thereby the complexity of the recognition system will then necessarily grow with the number of shapes and objects. This is a problem that any computational theory has to to address.

The target article suggests a learning step that reduces the dimensionality of the representation, but it still fails to discuss complexity or scaling issues, although they are inherent in the discussion of various methods (e.g., for achieving interpolation of views). The crucial question of complexity of storage and recall versus the number of shapes or categories stored is not addressed. With an increasing number of categories, the number of similarities to be represented grows combinatorially and the manifold on which the representations live becomes increasingly complex so as to capture isomorphisms. In general, one can say that although the pattern recognition paradigm is noncontroversial, it is of limited usefulness unless scaling and complexity issues are taken into account. This was the argument that made Marr (1976) embark upon his work on scene reconstruction. The argument still holds, whatever conclusion one draws.

The proposed representation is based on the appearance of the objects in images. Edelman discusses the fact that appearance is not invariant but depends on illumination and pose. This seems to be an oversimplification, since there are far more factors that can effect the appearance of a shape. In doing away with any need for representing parts or properties explicitly, it is not clear how the framework can deal with similarities on the basis of the constituent parts, or with the fact that the categories of interest at a particular instance are task dependent. What about a set of bottles with different labels? How is the shape of the bottle abstracted from the appearance of the bottle and the label together? In Edelman's system, there is no clue to this except the extension of the "bottle manifold" in the representational space to include all various kinds of labels. This will eventually lead to combinatorial explosion.

The basic property of the system responsible for these problems is the global nature of the representation. Edelman has not convinced us that we can do without an intermediate representational step, based on nonglobal shape properties that would allow for the abstraction of a shape from its mixed appearance. This is what most research in computer vision is founded on and studies of lesions in the human visual system also indicate that both types of models are needed (e.g., see Farah 1994). [See also Farah: Neuropsychological Inference with an Interactive Brain" BBS 17(1) 1994.]

What is wrong with prototypes

Peter Földiák

Psychological Laboratory, University of St Andrews, St Andrews KY16 9JU, United Kingdom. **peter.foldiak@st-andrews.ac.uk**
psych.st-and.ac.uk:8080/~pf2

Abstract: Representing objects and concepts as points in low-dimensional shape space defined by distances to other complete object exemplars or prototypes, expressed as single numbers, misses the key advantages of representation in terms of hierarchically constructed, meaningful features of the environment. Generalisation along statistically significant, near-independent, sparse, cooperative features that stand directly for various aspects of a concept is essential.

Edelman's target article comes from a long line of papers placing object/concept prototypes or exemplars at the centre of representation. The representation of a prototype corresponds to a complete object, without any structural, componential, or featural description. Relationships between items here are purely functions of distances in metric spaces in which these prototypes are located. The following serious problems follow from this.

1. The relationship between two concepts is usually much more complex than what a single number, the "similarity," can represent.

2. In prototype models, generalisation is simply a function of distance. Real generalisation and analogies depend on the correspondence of only some or even just one of the properties of the objects or concepts involved and are not affected by even large differences in other aspects. A red cherry can be more similar to a red bus than to a yellow carrot. Prototype-space would predict otherwise. Such judgements may also depend on the context, while prototypes do not allow the consideration of only certain aspects of a concept.

3. In a metric space the triangle inequality implies that if points A and B are close and points B and C are close, then A and C will also be close. Furthermore, if A and B are close and B and C are far, then A and C are also far. The full richness of relationships between a large collection of complex items can hardly be successfully embedded in such a space, especially in low dimensions. For instance, a red cherry is similar to a red bus, and a red bus is similar to a green bus, but the red cherry and the green bus should be maximally distant.

4. Prototypes are unsuitable for representing composite concepts by similarity to their components. A guppy may be a poor example of "fish" and a poor example of "pet" but it is a highly typical example of pet fish (Hampton 1993).

5. Sensory processing in the brain involves dimensionality expansion, not reduction (Barlow 1972; Field 1994). V1 contains about 100 times as many neurons as the optic nerve does, and higher visual areas maintain similar numbers. Despite this, correlations between neurons are surprisingly low even in higher areas and even in restricted experimental situations (Gawne & Richmond 1993), so it is unreasonable to assume small numbers of highly redundant "modules." The additional representational capacity of a high-dimensional representation can be used to increase selectivity and to make the representation sparse (Földiák 1990; Olshausen & Field 1994). The metaphor of a visual "alphabet" is also misleading as it suggests a small set of symbols. In fact, sensory neurons have a huge variety of response properties.

6. Neurons in higher visual areas can show remarkable specificity, and can stand for complex combinations of lower level features; nevertheless, their selectivity is far broader than what would be necessary for a prototype coder. They also ignore or generalise over far larger number of aspects of stimuli than what would make it helpful to think of them as representing complete, "holistic" prototypes. Even high-level cells code only certain aspects of an object and are not "pontifical cells" (Barlow 1972). The output of the suggested holistic classifier is not in any useful sense a feature as it does not signal any aspect of the object.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

7. Would the representation of the colour red by the activity of all red objects classifiers by efficient, and would it produce the correct generalisations? According to the prototype scheme the properties associated with redness all generalise far less to a red object with an unusual shape. Shape here should be irrelevant.

8. The suggestion that the new stimulus "giraffe" is represented by similarity to "camel" and "leopard" is an example of the inadequacy of the scheme. There is no way to know whether such a "giraffe" is an ungulate with spots or a predator with a hump. How would the prototype scheme represent a pink submarine unambiguously?

The introduction of elliptical basis functions to restrict the selectivity of the proposed classifiers to limited aspects of objects could help solve some of these problems, but the further we go in that direction the less holistic and prototype-like and the more feature-based the scheme becomes. Feature-based models representing items as sparse (Barlow 1959; Földiák 1990; Olshausen & Field 1994), cooperative (Dayan et al. 1995; Hinton 1992), low-redundancy (Barlow 1989; Bell 1996; Földiák 1990; Schmidhuber 1992) features can go a long way toward solving all the problems mentioned above. Such representations should not only be distributed and sparse, but they should also consist of features that directly correspond to meaningful statistical regularities, "suspicious coincidences," or "sensory cliches" (Barlow 1989) of the environment, while still being more structured and general than ones consisting of classifiers of individual exemplars. They not only tell us the degree of similarity between two concepts, but the overlap between their representations specifies the nature and aspect of the relationship. They provide biologically plausible, multiple-cause models (Földiák & Young 1995; Saund 1995) of the stimuli as opposed to the chorus of single-cause prototypes suggested.

Objects, please remain composed

Robert L. Goldstone

Psychology Department, Indiana University, Bloomington, IN 47405.
rgoldsto@indiana.edu cognitrn.psych.indiana.edu/

Abstract: The holistic representation of objects as coordinates in a psychological space should be supplemented with decompositional processes that break objects down into components. There is strong psychological evidence for object decomposition, and structured representations are also needed because of their computational efficiency. Structured and unstructured representations can be unified by a process that extracts regularities at multiple levels of an object.

Edelman's target article presents a coherent and persuasive account of at least one-half the task of representing similarities between objects. The article focuses on the representation of objects by their coordinates in a relatively low-dimensional space shared by other objects. Each object is represented holistically rather than decomposed into features. This holistic treatment of objects is a powerful technique, providing efficient representations of object similarities, particularly when combined with representations derived from interobject relations and blends of (whole) objects. There are, however, psychological and computational reasons for believing that objects are also represented by their decomposition into features.

A substantial body of psychological evidence supports decompositional accounts of object recognition. While Edelman cites recent research showing cells of inferotemporal (IT) cortex responding to whole objects, the earlier work on neurons that respond selectively to specific object properties such as color, orientation, edges, and motion provides some evidence for early featural decomposition (Hubel & Wiesel 1968). At a functional level, cognitive psychology provides additional evidence for decomposition. Garner (1974) reports evidence that shape and color features can be selectively attended without interference from each other, but some other stimulus properties cannot. Treisman and Gelade (1980) argue that features are registered separately, giving rise to efficient and parallel searches for individual features and the automatic splitting apart of different features that occur in the same object. In general, not all parts of an object are equally tightly connected to each other. Some object parts influence each other strongly and are fused, whereas other parts are naturally isolated and correspond to different psychological features.

Computationally speaking, structured representations are often highly efficient and parsimonious. Imagine a domain in which all the objects contain 5 parts selected from a vocabulary of 15 parts, and each part is related to every other part in one of eight ways. For example, a wristwatch could be encoded as a *watch connected* to two *straps,* one strap *attached* to a *buckle,* the other strap *attached* to a *prong,* and the prong *inserted* in the buckle. If each of the objects that could be represented by this componentially described system were represented holistically, then $(15*8)^5$, or about 24 billion whole-object representations would be required. Actual objects will certainly occur very sparsely in this space of features and relations. Still, to ignore the componential structure of the objects is to forfeit the opportunity of adopting a representational system with only 15 + 8 elements. The additional mechanisms required for building and processing structured representations are often more than compensated for by their economy, particularly as objects become increasingly complex.

As suggested by the above combinatorics, applying whole-object representations beyond toy domains quickly results in very high-dimensional spaces. Edelman's Figure 3 is misleading in that it suggests that an arbitrary object, such as a teapot, can be represented in the same low-dimensional space as the quadrupeds. If one wishes to reconstruct whole objects simply by identifying their coordinates or neighbors in a space, then an extremely large space would be required to represent all the objects we commonly recognize. The problem with the suggestion (sect. 2.1) that different object classes should be encoded by different parameters (i.e., in different spaces) is that objects belong to different classes at different times, and similarities between objects can be determined across classes. For some purposes, the shape similarity of dogs to wolves, cats, cows, dolphins, balls, tacks, and even shelves is relevant and is generally computable. The notion of parameter spaces that are tuned to particular object domains is important, but it must be supplemented with processes that can compute the similarity of any shapes.

One of the major limitations of multidimensional scaling approaches is their lack of structure. Even if the dimensions can be interpreted, there is no mechanism for dimensional interactions, or for representing relations between dimensions. Structured representations are likely to be particularly important when objects are composed of easily separable features, parts that have previously been encoded, or articulated segments. In these cases, structural descriptions provide an elegant and compressed representational code. Short codes can be used to token features that may be associated with quite complex configurations. For example, a single code can be built for an entire complex letter if it is involved in many words. It is true that these codes violate Marr's (1982) principle of least commitment, but the benefits of information compression necessitate discarding some raw information. Furthermore, if effective featural codes are constructed, then the raw information can be faithfully recovered by activating the feature.

A full account of object representation must exploit the complementary advantages of both holistic and structured representations. Unstructured whole-object representations are particularly useful when object parts are difficult to isolate, when there are many complex interactions between the parts, and when a particular object occurs frequently. The combined advantages of structured and unstructured representations can be achieved by a system that develops features at the most informationally efficient level. Such a system might create a single holistic representation for an entire word if it occurs frequently enough, but would alter-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

natively represent it compositionally if it is less frequent and the letters have been well learned. The same perceptual learning mechanism that can imprint on a whole object can also imprint on an element within the object if it occurs reliably across objects. A general imprinting mechanism of this sort unifies structured and unstructured representations under the assumption that both approaches work by detecting regularities and creating compressed codes for these regularities.

Metric assumptions are neither necessary nor sufficient to describe similarities

Robert A. M. Gregson

Division of Psychology, School of Life Sciences, Australian National University, Canberra ACT 0200, Australia. **robert.gregson@anu.edu.au**

Abstract: Alternative models of similarity judgments that do not rest on metric space assumptions are known to be better descriptions of actual human behaviour but are ignored by Edelman. The internal spaces he postulates are a convenient fiction for artificial intelligence, but not compatible with what is now known about psychophysics at both behavioural and neurological levels of perceptual processing.

Edelman (sect. 1.1, para. 4) begins his argument by misleadingly citing a work of mine (Gregson 1988) that has absolutely nothing to say about similarity. As I have published work very critical of the metric-space assumptions he espouses (Gregson 1975; 1976; 1979; 1980; 1984; 1985; 1993; 1994), the reader who wishes to follow the argument without preconceptions is advised to consult a source in which I discuss similarity explicitly, with examples. Gregson (1995) is the most recent text, and the first to explore explicitly the compatibility of similarity judgments with nonlinear psychophysical modelling. This does draw on results in multidimensional psychophysics (Gregson 1992) but can be treated as self-contained.

Edelman's suggestion in that paragraph that I was concerned with the human visual system's behaving "downright peculiarly" is about as misleading as can be. Having rejected metric-space notions precisely because I think they embody the wrong algebra for mapping behaviour that is quite normal, ubiquitous, and in no way peculiar, but just characteristically human, I have tried to treat similarity judgments as operations within the nonlinear dynamics of perceptual processes, and executed in time. This has involved using geometric patterns as stimulus materials because they are obvious and easy to manipulate, as many other workers in Europe have found (referenced in Gregson 1975; 1994; 1995), but odour mixtures and even series of musical tones have also been used (Gregson & Harvey 1992). I have no sympathy with similarity theory locked onto the transformations of one particular sensory-perceptual modality. One might, for example, even argue cogently that what distinguishes some judgments of olfactory mixtures from others in visual pattern perception is quite different processing of multidimensional similarities.

Tversky and Gati's (1982) critique is, I assume, well known to North American readers (even though Edelman does not cite it), but the diversity of subsequent models that do not assume metric axioms is unfortunately overlooked, though they were all required for good reasons, including a careful regard for real data properties. I have distinguished (Gregson 1995, p. 186) between metric-space models, vector models (from the Stockholm group), polymorphous models, set theoretic models (which include disparate models from Sweden, Australia, and the USA), and cascades in nonlinear psychophysics. In passing, I note that the algebra Edelman offers in section 5.2 is not strictly accurate as a summary of some of the set-theoretic models used and the asymmetry problem had been handled quite differently by both Ekman and myself. The metric-space idea and its counterpart assumptions in some multidimensional scaling algorithms have been discarded by workers because they assume too much and describe too little.

Setting aside technical naïvety, the trap is to assume that the monotonic distance idea for ordering similarity relations in an inner platonic ideal space (sects. 2.2 and 3.2.1) supports the Minkowski metric ideas that are global and invariant over neighbourhoods of the system's momentary reference points. It has been known since at least the 1920s that spaces that are metric only in a local neighbourhood, but have no global properties implying constraints on monotone distance-separation relations, can be defined and their properties resemble features of similarity mappings identified independently by Tversky, Eisler, and myself (Gregson 1995, p. 202).

The other problem that arises when we can construct simple counterinstances is the jump from (1) similarity based on element-wise matchings between corresponding partitioned subsets of stimulus attributes and (2) to matching on relational patterns within vectors of elements. The first can sometimes be locally reconciled with metric space mappings; but the second requires some hierarchies of similarity types, for example, jumping from first-order to relative similarities, or moving from pairs to quads of stimuli in a given comparison.

The accumulation of psychological evidence since the publication of Shepherd's original ideas has shown that the metric models, where discriminable from other models in their predictions, are inferior. In artificial intelligence we may assume what we like in order to see where it leads, and for mathematical tractability some will take the easy way out. In modeling real behaviour we must always respect not only the fine-grained structure of data but also what the brain actually does. The evidence on the latter has been slower coming, but it now tells against metric internal representations of relationships. A recent example is the neurological work of Cohen et al. (1996). Edelman's conclusion (4) of section 9.4 is where we can disagree most profoundly, with respect to the relevant psychophysics as well as the physiology.

Representations need self-organizing top-down expectations to fit a changing world

Stephen Grossberg

Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215. **steve@cns.bu.edu**

Abstract: "Chorus embodies an attempt to find out how far a mostly bottom-up approach to representation can be taken." Models that embody both bottom-up and top-down learning have stronger computational properties and explain more data about representation than feedforward models do.

Adaptive Resonance Theory (ART) models self-organize "second-order isomorphisms" using either unsupervised learning, supervised learning, or mixtures of both. This self-organizing capability is needed to learn in the real world. Regularization networks are not self-organizing in this sense. They cannot do fast stable learning in complex changing environments. These properties depend upon learned top-down expectations, matching of bottom-up data with these expectations, and mismatch-driven search for new representations (Carpenter & Grossberg 1991; Grossberg 1980; 1987). These mechanisms allow ART to automatically "ignore those directions . . . that are irrelevant to the identity of the stimulus" by focusing attention on critical features while suppressing irrelevant features. This ART matching rule has been supported by many psychophysical and neurobiological data (e.g., Grossberg 1995; Grossberg & Merrill 1996). ART matching also allows a dynamical control of attentive vigilance through a process of "match tracking" that automatically determines how general learned rep-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

resentations become to match world statistics (Carpenter & Grossberg 1991). Other models in which bottom-up and top-down processes are used (e.g., Back Propagation and the Helmholtz Machine) do not yet have these properties.

Edelman criticizes winner-take-all decisions because they violate the "principle of least commitment," but such decisions can quantitatively simulate categorical perception data (e.g., Grossberg et al. 1997a). ART systems such as masking fields (Cohen & Grossberg 1986), ART-EMAP (Carpenter & Ross 1995), Distributed ARTMAP (Carpenter 1996), and Gaussian ARTMAP (Williamson 1996) also show how distributed codes may improve recognition, and how the distribution reflects data uncertainty. Gaussian ARTMAP in particular is a self-organizing RBF (radial basis function) production system.

Self-organizing view-invariant 3-D object categories fuse view-specific categories in ARTMAP systems (e.g., Bradski & Grossberg 1995), as in the IT data reviewed in section 7.2. The 3-D categories occur in the Map Field, wherein outputs from multiple categories, whether of different letter fonts or different object views, are adaptively fused.

Edelman's measurements and dimensionality reduction stages are typically called vision and learned recognition stages. Although ART top-down matching occurs in the vision system, even as peripherally as the LGN (Gove et al. 1995; Grossberg et al. 1997b), vision uses principles and circuits different from those of the recognitions system. Edelman describes measurement as "a convolution with a number of filters, followed by the application of a nonlinearity," including light source compensation and figure-ground separation. Cortical models of visual perception, called FACADE models, suggest additional mechanisms (e.g., Arrington 1994; Chey et al. 1997; Francis & Grossberg 1996; Gove et al. 1995; Grossberg 1994; 1997; Grossberg et al. 1997b; Grossberg & Todorovic 1988). For example, parallel processing streams for boundary representation (interblob stream) and surface representation (blob stream) compute complementary computational properties. Feedback between these streams assures their mutual consistency and initiates figure-ground pop-out. Diffusive filling-in completes surface representations from signals that discount the illuminant.

Edelman summarizes a sensible approach to representation, but one that is limited by its feedforward character. ART models self-organize stable representations that achieve second-order isomorphism to arbitrarily large and changing environments, but only by using learned top-down expectations, attention, and memory search. FACADE models have clarified a lot of data about vision, but only by introducing new concepts about how complementary streams of boundary, surface, and motion processes achieve mutual consistency and coherence using other types of feedback. A major intellectual watershed separates feedforward models from self-organizing feedforward/feedback models. This watershed needs to be crossed for a deeper understanding of how humans autonomously form representations of the real world.

The notion of distal similarity is ill defined

Ulrike Hahn and Nick Chater

Department of Psychology, University of Warwick, Coventry CV7 4AL, United Kingdom. **u.hahn,n.chater@warwick.ac.uk**

Abstract: We argue that the notion of distal similarity on which Edelman's reconstruction of the process of perception and the nature of representation rests is ill defined. As a consequence, the mapping between world and description that is supposedly at stake is, in fact, a mapping between two different descriptions or "representations."

Edelman has shown experimentally that people can extract the underlying parameters used to generate a set of novel stimuli. From the results of multidimensional scaling, he conjectures that the internal space that people recover represents these parameters. This implies that nearby points in the original parameter space are near in the mental space, and it is short step from this to saying that similarity is preserved between the two spaces. Such results are not surprising where the dimensions of variation in the objects are subjectively obvious (e.g., the length and orientation of line segments), and in such cases this correlation between parameter space and mental space is frequently found. But it is impressive with Edelman's stimuli, where the underlying dimensions of variation are far from obvious and interact in a complex way to produce the visual image.

Edelman moves from these results to a general theory of perception founded on similarity. He presents this as an alternative to a "reconstructionist" approach. The goal of perception is assumed to be preserving similarities between things in the environment, rather than building an internal representation of environmental structure. Edelman's target article is important and should act as a valuable stimulus for future research. We believe, however, that there are three difficulties with this viewpoint as a general program in perception.

(1) The notion of "distal" similarity seems ill-defined. Goodman (1972) pointed out that any two objects have infinitely many common and distinctive features, thus "objectively" everything is equally similar to everything else. Watanabe (1985b) illustrates that even choosing for a set of objects only those predicates that are extensionally distinct (which for a finite set of objects is a finite set of predicates) still leaves all between-object similarities equal, unless differential weights for predicates are introduced. This is not just a philosophical nicety. In Edelman's experiments, stimuli are generated artificially by varying a set of parameters; thus nearness in parameter space may be chosen as a reasonable measure of similarity.

But the natural world has not been generated by manipulating a small number of underlying parameters. Variation in natural objects can be considered along a limitless number of dimensions. By choosing (and assigning differential weights to) any subset of these dimensions, all manner of "distal" similarities can be generated. Objects may be compared by overall color, by outline shape using any number of shape representation systems, by nearness to the observer (or to Pluto!), by weight, by perimeter length, and so on, indefinitely. Moreover, any of these measurements can be combined in arbitrary ways (e.g., perimeter length times weight) to produce new measures that can be used to give new dimensions.

Any set of any dimensions seems equally good as a distal measure of similarity. It might be suggested, for example, that physics could supply constraints on what can count as an underlying dimension, but it should be clear that this still leaves an infinite number of possible dimensions along which objects in the environment might be assessed; moreover, it will rule out many psychologically critical dimensions (e.g., the dimensions that define facial structure) since these do not relate to physical quantities. In short, it does not make sense to say that two things are similar without specifying in what way they are similar (Goodman 1972); to specify this, however, requires a cognitive agent to *define* which dimensions of distal variation matter and which do not; then the relation between an "objective" distal similarity structure and the similarity structure in the internal space of an agent breaks down. This means the claim that the perceptual system preserves an objective distal similarity structure loses its sense. Edelman, rather than dealing with objective properties of the world, is dealing with *two different descriptions* or representations – an experimenter-intended one (the underlying parametrization) and one formed by participants (the internal similarity spaces).

The situation seems analogous to the general philosophical difficulty with the correspondence theory of truth: there is no "mind-independent" way to specify which *facts* the world consists of, so the claim that true statements correspond to these facts is circular. In exactly the same way, there is no "mind-independent" way to specify which are the *similarities* in the world, so the claim that

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

similarities in mental space correspond to these external similarities is circular. But if there are no distal similarities, there can be no second-order isomorphism on which to build a theory of representation. The debate about the correspondence theory of truth as stated by us is a philosophical classic. The point we are making – that there is no "picture" relationship between statements and world – is widely accepted (see Strawson, Ayer, Wittgenstein II) even within logical positivism (for example, Neurath).

(2) Perception frequently appears to involve classifying very different patterns as similar. For example, the sequences 101010101010 and 010101010101 appear similar, even though they differ at each spatial location. Similarly, a photograph and its negative will be judged similar, even though they differ in every pixel value. Or again, different pictures of the same face, or different tokens of the same phoneme, will seem very similar, even if, under some obvious physical description, they appear completely different. The point is that the perceptual system identifies the common structure in both stimuli. How does this relate to Edelman's claim that distal structure is preserved in the internal representation of similarity? Using some obvious physical interpretation of the stimulus, the objects are very different, yet they are judged to be very similar, violating Edelman's theory. But using, instead, a perceptually appropriate description for measuring "distal" similarity (e.g., that the stimuli above are both examples of alternating patterns: descriptions in terms of the structure of a face or the identity of a phoneme), the similarities between the distal world and the mind are preserved, but only at the cost of circularity.

(3) Finally, we suggest that the reconstructive approach to perception may not be an *alternative* to Edelman's similarity-based view of perception. Instead, a reconstruction of the perceptual world may be required to explain why the similarities are judged as they are. For example, with Edelman's artificial figures, the parameters of variations may be of interest as part of a specification of the structure of those figures – indeed, only by attempting to reconstruct those figures does it seem possible to realize that there *are* only a small number of underlying parameters of variation (i.e., the recipe for reconstructing each figure is the same, apart from parametric variation). Thus, the parametrization used as a basis for internal similarity judgments may be *based on* the attempt to reconstruct the figure. For example, it is not clear why two pictures of the same face will be judged to be similar unless the same underlying 2/3D structure has been reconstructed (at least partially) for both. Thus, we would argue that the reconstructionist view of perception may be an important component in an account of similarity of relevance to Edelman's empirical results.

Representation of similarities and correspondence structure

Nathan Intrator

School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv 69978, Israel. **nin@math.tan.ac.il www.math.tan.ac.il/~nin**

Abstract: Apart from the computationally appealing properties of representation by similarities, it is possible to extend this form of representation when needed to include object parts as well as the correspondence between subobject parts.

Edelman provides a solid theory about object representation and its consequences. The idea of representing an object as a vector of distances from several other reference objects is very appealing on computational grounds and demonstrates a simple and probably robust dimensionality reduction. It further suggests a simple algorithm for hierarchical clustering, in which whenever a "sufficiently different" object appears, it may be registered as a new prototype, and when an object that is "not very different" appears and its class label is unexpected, it is again registered as a new prototype.

I would like to elaborate on the issue of "holistic" features versus the feature-representation that correspond to subparts in objects (sect. 6.3). It is a fundamental question in object representation not only whether there is a need to represent objects as wholes or as combinations of features, but also whether the exact topographic relation between subobject parts is essential. There is no doubt, for example, that there is a big difference between a phone that is on or off the hook, although this may be a very small difference in object space. This example demonstrates the need for an explicit feature-based representation with topographic correspondence, but as it would be difficult to argue that there is a prototype for an off-the-hook phone, holistic representations may coexist with more elaborate feature-based representations. If these representations do coexist, then it is likely that those based on prototypes are more specific but computationally simpler and are hence used for very repetitive (everyday) tasks, or tasks that require fast responses. The more elaborate representation is appropriate when the correspondence between object parts is important, for example, to represent walking or running.

One could argue that when a certain part of an object appears to have higher weight for purposes of recognition or discrimination, then that object part can be represented as a prototype or a distance vector from prototypes. The correspondence between object parts carries information that is very important and useful for classification and discrimination (Geman et al., forthcoming). In the case of representation by similarity, the exact relation between subparts and the object (the binding together of object parts) can be encoded via temporal structure such as synfires (Abeles 1981).

The representation of objects as a vector of distances from several prototypes suggests a very simple method for mental object manipulations, in which creating a mental representation of a certain object simply requires stimulating one (or more) of the prototype cells representing an instance of that object.

In summary, it appears that the simple object representation proposed by Edelman is compatible with the need for binding between subparts. Future psychophysics will clarify whether object representation via subparts coexists with holistic representation and whether the binding problem can be addressed by holistic representations and temporal structure.

Representation of similarities – a psychometric but not an explanatory concept for categorization

Martin Jüttner

Institute for Medicinal Psychology, University of Munich, Munich, D-80336, Germany. **martin@imp.med.uni-muenchen.de lrz.muenchen.de/~u7fo1bg/www/**

Abstract: The representation of similarities is a viable concept for a cognitive extension of visual psychophysics to the recognition of shapes, bringing issues such as similarity and categorization back into that field. However, as a framework it appears too general to place constraints on a particular process model for categorization. In particular, a preference for Chorus-like schemes with respect to structure-oriented approaches is unwarranted.

Edelman's conception can be regarded, on a theoretical level, as an extension of classical multidimensional scaling (MDS) to the recognition of shapes. To evaluate the potentials and limitations of such an undertaking, it is useful to recapitulate one of the basic motivations for MDS: it has been observed repeatedly that the probability that a learned response to any stimulus will generalize

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

to any other is not an invariant monotonic function of *physical* stimulus difference (Shepard 1987). This missing invariance eventually led to the radically different view of MDS as a way to reverse-engineer the problem of generalization. Rather than starting with physical stimulus properties, the response data produced by an observer were used to reconstruct a psychological space where distances would be monotonically related to generalization. Hence, the price paid for gaining a universal law of generalization was the loss of the specificity of the psychophysical mapping between the physical stimulus world and the psychological space where that law applies.

At this point, Edelman tries to reestablish the missing link by reconsidering Shepard's (1968) idea of second-order isomorphism and by evaluating the conditions necessary to preserve the similarity structures of distal (physical) feature space in their proximal (internal) representations. Such an approach is commendable in its own right because it returns classification, similarity, and recognition to the domain of visual psychophysics. Over a long period, models of spatial vision were concerned mainly with predicting detection and discrimination thresholds of certain stimulus patterns without explaining "how things look" (Shapely et al. 1990). The need for a paradigm shift toward a more cognitive perspective in psychophysics can be illustrated by our own work on classification learning in foveal and extrafoveal vision. Here it became clear that the perceptual dimensionality of (proximal) representations in extrafoveal vision is distinctly reduced but not that of foveally acquired representations (Jüttner & Rentschler 1996; Rentschler et al. 1994). Remarkably, this characteristic feature of extrafoveal vision has proved to be free of the well-known deficits in spatial resolution. It also calls for an extension of Edelman's distal-to-proximal-mapping scheme with an additional component accounting for retinal eccentricity.

To this extent the "representation of similarities" certainly provides a useful psychometric tool for understanding categorization. Problems arise, however, if this principle is used as explanatory concept, that is, as a justification for a particular *process* model for categorization. Here Edelman promotes his Chorus concept, in which a relatively small number of individual classifiers is tuned to a particular shape prototype and the relative activation triggered by a given test stimulus determines the similarity and/or classification response. There are a number of arguments against such a tight coupling between the psychometric concept of similarity and this particular implementation.

The first argument pertains to an empirical finding. In the above classification experiments with foveal and extrafoveal vision we recently evaluated a number of prominent classification models from the cognitive literature (Unzicker et al., in press). Among these was one implementation of a regularization network that also plays a central role in the implementation of Chorus. In our comparison, we measured the extent to which the similarity structure immanent in the observers' classification response could be replicated by the various models using the method proposed by Cutzu and Edelman (1996). Despite distinct differences in their theoretical assumptions, the models' performance was surprisingly equivalent, in particular for foveal viewing. Such a result suggests that similarity as such does not place constraints on a particular process model for categorization.

Second, in practical applications of a Chorus-like classification scheme, the computation of similarities has to be preceded by two decisions: which feature dimensions to use and which prototypes to consider. In this respect, Chorus relies on additional top-down information concerning the classification context, or, to take up Edelman's triangulation analogy, successful navigation requires knowing not only the bearings with respect to some landmarks but also to which map that sort of information applies.

There is another respect in which the limitations of the Chorus scheme become obvious. Like the classical Pandemonium model, Edelman's new version faces severe problems in more realistic (i.e., texture-defined) environments where it becomes exceedingly difficult to decide which of the texture, colour, or contour-determined patches actually define the "object" whose similarity coordinates are to be determined. It seems rash to oppose structure-oriented approaches as Edelman does. After all, it is the problem of scene understanding in multiple-object environments that led to the nonaccidental prevalence of structure-oriented object recognition systems in computer vision (cf. Caelli & Bischof 1996; Flynn & Jain 1993). Moreover, such approaches are not necessarily limited to extreme reconstructionist positions as Edelman seems to imply. For example, evidence-based systems (EBS) originally proposed in the field of machine vision (Caelli & Dreier 1994; Jain & Hoffman 1988) have been successfully applied to reconstructing processes of pattern classification and generalization in humans (Jüttner et al. 1997). EBS do not argue for a fixed reservoir of shape primitives, nor do they adopt a definite position in the debate about whether objects are to be represented mentally as 2D or 3D models. Rather, they provide a method for transforming images into a rule-based representational format open to propositional reasoning.

This situation is reminiscent of the long-standing debate about "analogue" versus "propositional" representations. [See also Pylyshyn: "Computational Models and Empirical Constraints" BBS 1(1) 1978; and Kossyln: "On the Demystification of Mental Imagery" BBS 2(4) 1979.] It may be worthwhile to reconsider the argument of Anderson (1978). Given that all cognitive behaviour is the product of both *representation* and *process,* he argued for an indeterminacy concerning the representational format as long as the processes operating on them remain unspecified. Edelman's theoretical concern is restricted to representation as such, whereas its use is discussed only on the (secondary) level of implementation (i.e., the Chorus scheme). His approach is accordingly faced with a similar indeterminacy in its explanatory value.

The Chorus scheme: Representation or isomorphism, holistic or analytic?

Cyril Latimer

Department of Psychology, University of Sydney, Sydney NSW 2006, Australia. **cyril@psych.usyd.edu.au**
www.psych.usyd.edu.au/staff/cyril/

Abstract: The Chorus scheme could be an important step in the search for solutions to the symbol grounding problem (Harnad 1990), but Edelman does not address the potential difficulties inherent in downgrading differences in favor of similarities in a categorization device. Isomorphism rather than representation is a more coherent way of thinking about Chorus whose modules are probably analytic rather than holistic.

Representation of similarities or differences? Edelman proposes that representation is representation of similarities, but he nowhere addresses the problems associated with permitting the ascendancy of similarities over differences (Sutcliffe 1986). Cassirer (1966) notes that in such a scheme, those aspects that differentiate objects tend more and more to disappear and form only a shadowy background on which the constant features gain salience. Abstraction of sameness leaves behind all the particularities in such a way that they, and the transformations of which they are capable, become irrecoverable. On the other hand: "The genuine concept does not disregard the particularities which it holds under it, but seeks to show the *necessity* of the occurrence and connection of just those particularities. What it gives is a universal rule for the connection of the particulars themselves" (Cassirer 1966, p. 30). Edelman's scheme nonetheless has great appeal, offering as it does a potential mechanism for the iconic and categorical representations necessary for an attack on the symbol-grounding problem (Harnad 1990; 1992). But is *representation* the correct concept in this context?

Representation or isomorphism? Edelman's discussion vacillates uneasily between the notions of representation and isomor-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

phism. Representation is a ternary relation whereas isomorphism is a binary one. A representation has (a) the thing represented, (b) the thing representing (a), and (c) someone or something that knows that (a) represents (b). In theories of cognition, (c) could only refer to some homunculus or perhaps to some future super neurosurgeon able to observe a patient's brain states and note how they correlate perfectly with (represent) states of the world (See Maze [1983] and Michell [1988] for a comprehensive case against representative theories of cognition.) A more viable conception is *isomorphism,* where states of the world stand in a direct relationship with states of the brain, and Edelman's mechanism is ideally suited to modelling the processes that bring about this isomorphism in category learning and concept formation. Indeed, Harnad makes a similar point, "It is not that the mind *receives* the transducer/effector or analogue activity (or, for that matter, the symbolic activity) as data. If the mind is grounded in this way, then it just *is* the activity of those structures and processes" (1992, p. 80).

The world as its own representation? Given the above account of representation, the notion of the world acting as its own representation is an incoherent one. It would be much clearer to say (sect. 9.3.3) that Chorus simply responds or resonates to the environment; but this still leaves in doubt the ontological status of the modules in Chorus. Edelman cites John Locke's simple and complex ideas as precursors of feature detection theory, but surely his doctrine of abstract ideas is more apposite in the context of categorization. Locke struggled with the ontological status of his abstract ideas and was misinterpreted, not least by Berkeley, "What more easy for any one to look a little into his own thoughts and there try whether he has, or can attain to have, an idea that shall correspond to the description here given of the general idea of a triangle – which is neither oblique nor rectangle, equilateral, equicrural nor scalenon, but all and none of these at once?" (Berkeley 1710/1965, p. 52). Locke, however, foresaw the difficulty in abstract ideas as crude templates or prototypes with an existence in their own right, and asserts, "they frame an *idea* which they find those many particulars to partake in, and to that they give, with others, the name *man,* for example. And *thus they come to have a general name,* and a general *idea.* Wherein they make nothing new, but only leave out of the complex *idea* they had of *Peter* and *James, Mary* and *Jane,* that which is peculiar to each, and retain only that which is common to them all" (Locke 1690/1964, p. 17). Further on in the essay, he emphasizes the point, "But if we would rightly consider what is done in all these *genera* and *species* or sorts, we should find that there is no new thing made" (Locke 1690/1964, p. 62). Locke's doctrine is thus closer to that of Cassirer, who regards the concept, not as an entity in its own right, but as a rule that captures the relationships in which the particulars stand. The strength of Chorus is that not only does it too avoid reifying categories, but it could also provide an explicit, neurally plausible mechanism for responding directly and accurately to particulars and their interrelationships.

The holistic treatment of objects? In contrast to feature-detection theory, the Chorus modules are said to be holistic analyzers. There is not enough information in the the target article to verify this (there rarely is in papers that deal with wholes and parts that are relative and not absolute; Latimer & Stevens 1997). How is input presented to Chorus? If input is in pixels or even grey scale, then it is still being segmented into parts, albeit much smaller parts than in most feature-detection theories, but still parts. What role does the information contained in these parts (relative positions in the input array, etc.) play in later computations of similarity and difference of objects? If the so-called holistic properties of objects are being derived from properties of the parts, then in principle and in practice, the Chorus scheme is no more holistic than the mechanisms of feature-detection theory.

Boundary conditions and the need for multiple forms of representation

Arthur B. Markman and Takashi Yamauchi

Department of Psychology, Columbia University, New York, NY 10027.
markman@psych.columbia.edu www.columbia.edu/~abm16;
takashi@psych.columbia.edu

Abstract: Multidimensional space representations like those posited in Edelman's target article are not sufficient to capture all similarity phenomena. We discuss phenomena that are compatible with models of similarity that assume structured relational representations. An adequate model of similarity and perception will require multiple approaches to representation.

The representational system advocated in the target article is based on the use of multidimensional spaces, in which similarity is inversely proportional to distance in space. The model assumes that objects are represented by points in a space. The simple measurement of distance between points may be augmented with other processes to account for observed asymmetries and context effects in similarity judgments (Krumhansl 1978; Nosofsky 1986). Edelman suggests that multidimensional space representations might be used to account for judgments of similarity of complex visual scenes as well.

We describe phenomena that serve as boundary conditions on the proposal that similarities can be characterized as distances in a mental space. These phenomena do not rule out the use of multidimensional space representations in perception; rather, they suggest that many forms of representation must coexist in models of perception and similarity.

Boundary phenomena. A central boundary condition on similarity is that people have access to the commonalities and differences arising from comparison, even from comparisons of visual scenes (Markman & Gentner 1993b; 1996). For example, Markman and Gentner (1996) asked people to list the commonalities and differences of pairs of complex scenes. Three findings from these studies are important here. First, people easily listed the commonalities and differences of these pairs, suggesting that they could fix upon discrete aspects of the comparison. Second, there was a high correlation between the number of listed commonalities and the rated similarity for each pair, suggesting that the perception of similarity is determined by the properties arising from a comparison. Third, the commonalities and differences included correspondences between items that (a) were visually similar in the scenes (e.g., Christmas trees that looked similar), (b) sat in the same spatial relationship in pairs of scenes (e.g., a vase and an angel statuette on top of a mantle), or (c) sat in the same conceptual relationship in pairs of scenes (e.g., cars and robots that were both being repaired).

Other studies suggest that similarity is highly sensitive to identity of representational elements. In classic studies, Tversky and Gati (1982) found that exact matches along some dimension were weighted more heavily than were dimension values that were merely similar (the *coincidence effect*). Pairs of objects with identical values on one dimension and dissimilar values along the other dimension were rated as more similar than were pairs of objects with moderately similar values along each dimension. The reverse has been observed in studies of choice, where it has been shown that people will choose an item moderately similar to an ideal along two dimensions rather than an item identical on one dimension and dissimilar on the other (Kaplan & Medin 1997; Simonson & Tversky 1992).

The ability to access commonalities and differences of visual scenes suggests (1) that there are discrete representational elements in complex scenes that can be placed in correspondence and accessed and (2) that the basis of a correspondence can be similarities in perceptual properties, spatial relations, or conceptual relations. These abilities do not seem compatible with a multidimensional space representation because multidimensional

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

spaces only allow calculations of distances between points in which the dimensions of comparison are predetermined by the dimensions of the space.

The data from scene similarity suggest that there are many ways correspondences may be determined. Comparisons of individual perceptual objects are also influenced by the context of the comparison. For example, a melon cannot be distinguished from a basketball without information about color (and texture) because shape information is shared by these objects. Similarly, distinguishing a zebra from a horse or a cat from a tiger requires color information. At times, functional information also seems important, so distinguishing an orange from a tennis ball might require a combination of shape and color information as well as input from higher level knowledge about the uses of these objects. A model of perception needs to have a mechanism for integrating a variety of sources of information that come together.

Multiple representations. These boundary phenomena do not rule out the use of multidimensional space representations in object recognition. Instead, they suggest that no single representational system will successfully serve as the basis of cognitive models (Markman, in press; Markman & Dietrich, in preparation). It is likely that there are redundant representational systems underlying human cognitive abilities. The target article suggests that a multidimensional space representation makes sense for some aspects of object recognition. There is compelling evidence, however, that similarity comparisons require structured relational representations akin to those proposed in structural description theories of object recognition (e.g., Biederman 1987). Rather than seeking a winner-take-all battle, I urge a peaceful coexistence of representational systems in models of perception and similarity.

ACKNOWLEDGMENT

This work was supported by NSF grant SBR-95-10924.

How to combine interpolation with feedback?

Guenther Palm

Department of Neural Information Processing, University of Ulm, D-89069 Ulm, Germany. **palm@neuro.informatik.uni-ulm.de**

Abstract: The Chorus representation is a sparse, similarity-preserving representation achieved by a feedforward neural network. Hence it is probably better suited for interpolation than for categorization. This commentary raises the question of how to combine categorization with interpolation, whether feedforward networks can be reasonable models for parts of the cerebral cortex, and whether people can perform more than one interpolation at a time.

The essence of Edelman's target article is the introduction of a similarity-preserving representation (the Chorus representation) and a discussion of some of the virtues of similarity. I sympathize very much with this approach since we have concentrated much of our research effort on the creation of sparse similarity preserving representations for associative memory (Palm 1980; 1987a; 1987b; 1990; Palm & Palm 1991; Palm et al. 1997; Stellmann 1992). Their usefulness in associative memory, and, more generally in any kind of robust processing is an additional virtue of similarity-preserving codes or representations. In fact, the Chorus representation is not only similarity preserving but also sparse, that is, most of the representational units have zero (or near zero) activity. The word "sparse" perhaps characterizes this property even better than Edelman's "low dimensional." Some issues related to the Chorus representation and the computational use of similarity in general will be raised below:

(1) Feedforward neural networks like the Chorus scheme are in general good for interpolation. How can this be combined with the need for categorization and segmentation?

I have the impression that the proposed scheme works only on presegmented pieces of images; the segmentation itself probably has to be performed by a different network. Categorization, a typical feature of feedback associative memories, can help to perform segmentation in difficult cases, but this is not compatible with the interpolation properties of similarity-preserving representations in feedforward networks. An interesting question is how to combine feedback and feedforward networks to obtain an architecture that can perform both segmentation by categorization and interpolation. Pursuing this question may also lead to a more realistic model of inferotemporal (IT) cortex.

A related question amenable to experimental perceptual tests concerns whether humans can indeed solve problems involving a combination of segmentation and interpolation. For example, if subjects learn to recognize some novel shapes from particular views (as in Bülthoff & Edelman 1992; Logothetis et al. 1995) and first have to identify them (e.g., in forced choice experiments) hidden in a background of similar shapes, can they still identify these particular shapes when they are shown from intermediate but novel views?

(2) Are feedforward networks adequate explanations for information processing in the cerebral cortex, in view of the prominence of anatomical feedback within and between cortical areas? In particular, it is doubtful that interpolation is the sole or even the principal function of IT cortex.

(3) Another issue has to do with the problem of compositionality of representation mentioned only briefly in the target article. The Chorus scheme is a representation of a small segment of the visual scene (perhaps the focus of attention?) containing essentially one object.

What happens if there are two or three objects in this segment? Or one object composed of parts that can be addressed as objects in their own rights? If the system knows a train engine and a snake but not a train, would it treat the train as more similar to the engine or the snake?

Perhaps more interesting than amending Chorus with additional mechanisms to deal with composed or multiple objects is the corresponding experimental psychological question: Can humans interpolate two or three objects simultaneously? And are there perhaps different interpolation networks for different spots on the retina?

(4) The most fundamental issue related to similarity-preserving representations is the question of who or what defines the similarity.

I think the internal similarity cannot always be just the external sensory similarity, as supposed in the target article. There are other important similarities, for instance, a functional similarity: defining a chair as something to sit on, we can identify many objects as suitable chairs and regard them as similar in this respect without a simple visual similarity.

In understanding speech, for example, we can identify words as similar on the basis of contents that sound quite different; vice versa, similar sounding sentences can have quite different meanings (e.g., "let us recognize speech" vs. "let us wreck a nice beach").

As in speech recognition, this endowment of objects with a different nonsensory similarity normally comes after a stage of categorization (or categorical perception). As in the understanding of spoken sentences, there has to be a close interaction between similarity-based matching on the "lower" level and "higher" level of similarity. This very important and intricate problem is clearly beyond the scope of the target article. It may even be doubted whether the concept of similarity or of interpolation is still adequate on the higher level, and if it is, it may only be definable functionally, which leads to a certain circularity in the definition of higher level similarity. In any case, these questions are probably less amenable to current neuroscientific approaches and more important for the organization of complex technical systems (understanding of images or speech) in artificial intelligence.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

Attentional dynamics and a chorus of geons

Eric Postma, Jaap van den Herik, and Patrick Hudson

Computer Science Department, Maastricht University, 6200 MD Maastricht, The Netherlands. **postma@cs.unimaas.nl; herik@cs.unimaas.nl; hudson@cs.unimaas.nl www.cs.unimass.nl/~postma**

Abstract: This commentary discusses three main requirements for models of vision, namely, *translation and scale invariance, scalability,* and *hierarchy.* Edelman's Chorus model falls short of fulfilling these requirements because it ignores the highly dynamic nature of vision. Incorporating an attentional mechanism and assuming geon-like prototype representations may enhance Chorus's plausibility as a model of human object recognition.

Edelman presents an inspiring account of visual representations in the brain. The impressive recognition performance of the "Chorus model" is on a par with the best state-of-the-art algorithms for recognizing presegmented shapes. Chorus acknowledges the high dimensionality of the retinal image and does not assume the image-like input representation commonly used in other visual models. Since the million retinal signals received by the primary visual cortex are not spatially labeled (Koenderink 1984), the spatial order needs to be recovered from the signals themselves. In Chorus, this recovery proceeds by embedding the high-dimensional inputs in an appropriate low-dimensional shape space.

Notwithstanding the general appeal and originality of the proposed model, Chorus falls short of fulfilling three main requirements for models of visual recognition, namely, *translation and scale invariance, scalability,* and *hierarchy.* In the following, Chorus's failure on each of the requirements is discussed and related to a single underlying limitation.

First, the evidence for complete translation and scale invariance of human recognition (Biederman & Cooper 1991) imposes a severe structural constraint on models of human recognition. In section 3.2.2, Edelman acknowledges the need to compensate for (image-plane) translations and suggests a solution (i.e., covert attention) with which we fully agree. However, there is more to covert attention then just solving translation problems (see below). For scale invariance, Edelman suggests the solution proposed by Schwartz (1985). Unfortunately, this solution confounds spatial order with functional order by relying on the topography of primary visual cortex. Hence some other solution is needed.

Second, section 4.1 does not mention how many "landmarks" are required in Chorus to triangulate the shape space in a reliable way. Given the huge number of different shapes and their similarity relations, it is unlikely that only about a dozen reference shapes will suffice for distinguishing between each pair of all naturally occurring shapes. On the contrary, the number of prototypes required for reliable recognition will become very large. This increases the effective dimensionality of the shape-space representations when a visual object activates many prototypes. High-dimensional representations are profitable for their robustness (cf. Rao & Ballard 1995) but invalidate the generalization performance of Chorus when serving as a basis for classification.

Third, the Chorus model represents shapes in their entirety but cannot represent part-whole and part-part relations. Human observers, however, are able to recognize an object as a configuration of parts and features. This limitation may lead to false predictions about human similarity ratings (cf. Hummel & Stankiewicz 1996).

Dynamic vision: Attention and a chorus of geons. Chorus's failure on the three requirements can be characterized by a single shortcoming: Chorus ignores the highly dynamic nature of vision. Since objects and scenes are scanned sequentially through saccadic eye movements, they give rise to representations that are updated in an incremental way. In close connection, a gaze-independent attentional process, that is, covert attention, selects locations and scales appropriate for the task at hand (Postma et al. 1997). In this way, covert attention allows for *both* translation and scale-invariant recognition by dynamically varying the location and grain of its sampling grid. A solution to Chorus's scalability problem is to assume a limited set of basis shapes, not unlike Biederman's (1987) geons, which serve as prototypes for recognition. Such a scheme necessitates an incremental reconstruction of shape representations in which spatial attention plays a central part by selecting the parts and effectively preventing interference from other parts or objects. The ensuing representations form recursive structures, such as trees, which accommodate the need for part–whole representations of objects and scenes.

The sequential and hierarchical nature of visual processing in the brain as evidenced by biological and psychological findings reflects a multistage strategy consistent with a such a compositional process. Even during fixation, object recognition may proceed in a sequential fashion. In a backward-masking paradigm, presentation times as short as *100* milliseconds (which is too short to make eye movements) suffice for the recognition of well-known objects and scenes. Within *100* milliseconds, two to three attentional snapshots can be taken (Saarinen & Julesz 1991). Interestingly, the extraction of two to three parts together with their invariant relations is sufficient for view-invariant entry level classification (Fiser et al. 1996). Hence object recognition, even without eye movements, can still be considered an incremental process that proceeds by the mechanism underlying the shifts of covert attention.

Conclusion. We feel that the representation-by-similarities approach offers a viable theory of representing shapes and their similarities, but not of representing objects and scenes. In combination with an active selection process and a means for representing part–whole and part–part relations, however, the approach may lead to a plausible model of human object recognition.

Vector code differences and similarities

E. N. Sokolov

Department of Psychophysiology, Moscow State Lomonosov University, Moscow 103009, Russia. **sokolov@cogsci.msu.su**

Abstract: Edelman suggests that any shape is encoded by an excitation vector with components corresponding to excitations of corresponding neuronal modules. This results in discrimination of stimuli in a shape space of low dimensionality. Similar vector encoding is present in color vision. Red-green, blue-yellow, bright and dark neurons are modules that represent a number of different color stimuli in color space of low dimensionality. Vector encoding allows effective computation of color differences and color similarities. Such a neuronal vector-encoding approach has also been applied to the perception of visual movement, line orientation, and stereopsis.

Edelman's theory is a unified approach to visual representation. It suggests that a shape is represented by the activation of a limited number of neuron modules each broadly selective for a set of shapes. Thus, any shape is encoded by an excitation vector with components equal to excitations of the corresponding neuronal modules. This strategy results in the discrimination of stimuli in a shape space of a low dimensionality.

Similar vector encoding occurs in color vision. The multidimensional scaling of a matrix of subjective color differences derived from paired presentation of color stimuli yielded a four-dimensional space. Each color is characterized by a selective color detector tuned to the excitation of four types of color encoding neurons: red-green, blue-yellow, bright, and dark. The lengths of the excitation vectors are equal, so colors are represented on a hypersphere in four-dimensional space. Three angles of the hypersphere closely match hue, lightness, and saturation, respectively (Izmailov & Sokolov 1991).

Subjective differences correlate highly with the Euclidean distances between corresponding color points. This correspondence

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

between subjective differences and Euclidean distances suggests that absolute values of vectorial differences are the basis for the perception of color differences.

This vector coding of color stimuli also occurs in fish and monkeys as measured with instrumentally conditioned responses. Factor analysis of confusion matrices with probabilities and differential stimuli revealed a four-dimensional color space closely resembling color space in humans (Sokolov 1994).

Each color stimulus is characterized by a specific excitation vector, all vectors being equal in length. Using the coordinates of the vectors, one can compute their inner products. It has been shown that the matrix of the inner products of these color vectors corresponds with the probability matrix obtained by instrumental conditioning (Latanov et al. 1997).

The correspondence of response probabilities to inner products of color vectors implies that during conditioning an output command neuron's inner products are computing to get similarity measures between conditional and differential color stimuli.

Thus red-green, blue-yellow, bright, and dark neurons are the modules that represent a number of different color stimuli in a color space of low dimensionality. Vector encoding allows effective computation of color similarities and differences. A similar vector code is also likely in other modalities. This neuronal vector-encoding approach has also been applied to the perception of visual movement, line orientation, and stereopsis (Fomin et al. 1979).

Visual tasks require manipulable representations[1]

Bradley V. Stuart

Center for Automation Research, University of Maryland, College Park, MD 20742. **brad@cfar.umd.edu www.cfar.umd.edu/~brad**

Abstract: Representation of similarities is not sufficient for most visual tasks. The proposed framework collapses useful dimensions such as position and pose for the sake of naming the object. Collapsing these dimensions leaves no representation of the object itself, but only an internal name that cannot be meaningfully manipulated.

In the proposed representational system, a set of classifiers is trained to recognize a number of nonrigid objects in different configurations and orientations. New objects are then classified (or defined) by their similarity to the nearest few of these training examples. The approach is to reduce an extremely high-dimensional input space (the retina) to a medium-dimensional measurement space, and then to define object classes as fuzzy regions in this space using radial basis function classifiers. Images are then described by comparison with a number of those previously built measurement vectors that most closely match the measurement vector of the input.

Any system of representation will reflect similarities in the distal environment, but that alone is not sufficient. Modern computer hardware even allows systems previously deemed unworkable or too expensive to successfully classify objects among a set of training images. Edelman's system and the appearance-based systems reduce the dimensionality of the input by comparison with a representation that combines all the training examples into averages and distributions. Edelman's approach precedes this comparison with a dimensionality-reducing measurement step, whereas this is not required in the appearance-based approach.

A robotic or biological system with vision that needs to manipulate objects (including its own body) also needs to manipulate representations of those objects. Interactions such as tool or part grasping and local path planning require not only identifying the target object but also representing the object's location and pose. These aspects of the object are explicitly eliminated in the proposed description, as the training process collapses dimensions unrelated to the object class.

On the other hand, a manipulable representation encodes these aspects explicitly, and allows the agent to change them. By manipulating the representation, the agent can assess the likely effects of an action before it is performed. While the measurement-space level may be manipulable, this is not the basic representation proposed by Edelman. The neural network classifiers and multi-dimensional scaling (MDS) collapse dimensions of the measurement space not germane to the object's name – precisely those dimensions on which objects are manipulated.

The world does not consist of objects in isolation, so even this representation will not be rich enough to capture the relationships between objects. Individual object representations need to be put together in relation to one another as the agent builds a representation of the entire scene. It isn't necessary that these relationships be in metric correspondence with the environment, just that they encode the relations pertinent to the agent's goals.

A representation suitable for an active agent also needs to represent actions or the influence of one object on another. Moving agents need to reason about how their motion changes the relationship of the agent (itself an object) to its surroundings. Potential collisions need to be discovered and their possible consequences determined. What prestored object could be similar to a collision?

Systems with vision need to navigate in their environment, build up maps of various scenes, link these scene maps into global maps, and represent possible changes in the external world. Prediction is a valuable mode of reasoning, and it requires a manipulable representation of objects and places. These representations are modified as changes in the environment are noticed; they are improved as more details are perceived. Links are added and changed as new facts are discovered. Representations need to be predictive; we should be able to determine whether a tool fits a part or whether a collision is imminent without always trying it first. Navigating systems need more than simple reactive interaction with the visible environment. It is not enough to simply identify a place; the agent must localize itself in the place, recall its structure, and decide how to act.

There are computer vision problems where representations of space or the relationships between objects are not needed. The approach could probably be successful in industrial applications where the number of objects is limited and the segmentation problem is not an issue. In problems such as image database indexing, where images need to be searched for dictionary terms, this representation scheme will probably work well. These are essentially naming tasks; the problem is to select a set of categories based on the image data.

Conclusion. By focusing on what is learnable by radial basis function classifier networks, Edelman has selected a representation that supports only bottom-up processing of visual data, and that solves a problem particularly well suited to classifier networks: the placement of an object among a set of previously learned examples. The representation explicitly removes aspects of the object related to its manipulation or its environmental context. These aspects, much more than an object's identity or name, are important parts of the visual world and its representation in an active agent.

NOTE

1. Dr. Stuart's commentary was received too late to be responded to by the author in the first round; Dr. Edelman will reply to it in a Continuing Commentary section of a forthcoming issue.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

A neural basis for the Chorus model?

M. J. Tovée

Psychology Department, Newcastle University, Newcastle Upon Tyne, NE1 7RU, United Kingdom. **m.j.tovee@ncl.ac.uk**
www.psychology.ncl.ac.uk/www/psychol.html

Abstract: The neural basis of the Chorus model has been cast in terms of the visual alphabet theory, but the neural evidence can also be interpreted as supporting a theory of higher level representation in which neurons are responsive to complex 3D stimuli. These neurons, functioning as a population, could also form the basis of a representation such as envisaged by the Chorus model.

In his discussion of a neural basis for the Chorus model, Edelman seems to accept the visual alphabet hypothesis and cuts his model to fit this concept of higher processing. However, the nature of representation in inferotemporal cortex (IT) is far from clear-cut. Based on the responses of Tanaka's elaborate cells (Tanaka et al. 1991), a representation could be derived in at least two ways. First there could be a traditional hierarchy in which the cells responsive to simple shapes would feed into a higher cell layer, whose cells respond preferentially to complex stimuli. The output of these cells would then signal the presence of a complex object to higher areas, such as the prefrontal cortex. Alternatively, there may be no upper layer. The pattern of responses across the various columns of elaborate cells may directly signal the presence of a complex object to a higher area without having to converge on a cell in IT sensitive to complex stimuli. The latter view is favoured by disciples of the visual alphabet hypothesis.

The visual alphabet theory assumes that an IT cell will reliably signal the presence of the particular simple shape that excites it regardless of whatever else is present in the visual field. However, this may not always be the case. In a recent study, the responses of single neurons in the anterior IT of awake-behaving macaques were recorded when the monkey performed a visual fixation task (Rolls & Tovée 1995). The responses of individual neurons to visual stimuli presented individually at one of two positions in the cell's receptive field were then determined. The images were then shown in pairs and the cell's response was compared with its response to the images presented separately in the corresponding area of the receptive field. There was a significant interaction effect between the stimuli when shown as pairs. A similar effect has been reported for neurons in more posterior IT (Sato 1989). Additionally, some of Tanaka's own data may show this effect. In a cell shown in his 1991 paper, Tanaka's simplification procedure converged on an inverted T shape as the preferred simple shape for this cell (Tanaka et al. 1991). A more complex object that contains this simple shape should evoke a strong response from the cell, as the cell should signal the presence of the shape. However, the cell did not respond well to a + shape, in which the preferred simple shape is still present, in concert with a bar below its centre (Young 1995). Thus, the presence of other visual features may disrupt the response of a cell to its preferred shape, a result that is the opposite of what is assumed in the visual alphabet conception of IT. If other cells behave in a similar way, the characterisation by this method of the simple shapes preferred by IT cells cannot be sufficient to account for the performance of the cells in the recognition of even slightly more complex objects.

Perhaps the strongest evidence for neurons responsive to complex 3D objects are the face-selective neurons, which seem resistant to stimulus simplification protocols (see Tovée 1995). This finding is supported by Tanaka's own research, in which a combined electrophysiological and optical imaging study found face-selective neurons arranged in columns in anterior IT (Wang et al. 1996). Tanaka (1996) has argued that faces are a special case, and that all other complex stimuli are represented by a distributed code across cells responsive to simple 2D shapes. Only faces have a specific class of neurons tuned to them. However, it could be argued that it is unlikely that only faces, out of all the complex stimuli that the brain needs to represent, should have a specific class of cells to represent them. There have been persistent reports that in addition to cells responsive to faces, there are cells responsive to other complex biological stimuli, such as hands. Faces are important, especially for social animals such as primates, but so are other visual stimuli, such as food sources and predators. It seems more likely that face-selective neurons are an example of a class of neurons responsive to complex stimuli and that other classes of neurons are responsive to other complex stimuli (Tovée 1995). This is consistent with clinical data from brain-damaged subjects where some patients have been reported with selective impairments of the ability to recognise and classify a particular class of complex stimuli such as faces, coins, cars, or domestic animals (e.g., Young 1992).

In conclusion, we can say the available neural data can be interpreted as supporting a theory of representation in which neurons are responsive to complex 3D stimuli. Populations of these view-invariant cells, which can modify and change their responses on the basis of experience (Tovée et al. 1996), are able to represent vast numbers of stimuli (Abbott et al. 1996). It is these neurons that may provide the neural basis of the Chorus model.

A multiculture of veridicalities

J. van Brakel

Institute of Philosophy, University of Leuven, 3000 Leuven, Belgium.
pop00127@cc5.kuleuven.ac.be

Abstract: Edelman's target article purports to be about veridical representations. I argue that it would be a mistake to think it has much to do with veridicality as normally understood.

If representation is *something,* it is certainly an improvement to talk about "representation of similarities" instead of representation *simpliciter.* It would be even better to talk about "representation of similarity in one respect or another." What tying "representation" to "similarity" leaves open, as Edelman notes in passing is: (1) that the power of representation would grow, were the system (i.e., the brain) plastic enough to attune itself to novel object classes and (2) that two perceptual systems implementing different mappings could have incompatible (or even conflicting) pictures of the world.

I propose that such observations should be at the centre of research on vision, which should focus on the plasticity of categories and the absence of "objective" similarities (van Brakel 1991; 1993).

Notwithstanding the interesting advantages Edelman's approach may have, the technical details are couched in a language that hides suppositions possibly not conducive to research on perception. I will concentrate on one term, "veridical"; other expressions innocently used, include: salience, natural kind, structure of natural kinds, structure of the world, the order and the connection of things, nearest-neighbour structures, widely disparate categories, basic-level categorisation, prototypes, familiar categories.

Appealing to colloquial usage, Edelman suggests, that "veridical" refers to a relation between something (a sentence, a representation, a Chorus of classifiers) and The World. He claims, for example, that perceptual systems are faced with the problem of securing a veridical relationship between the world and its internal representation (Abstract and sect. 1.1), and he stresses (sect. 9.3.3) that the burden of representation belongs in the world, implying that his distal space shape is the world (the same conflation is made in Appendix B).

But one would be mistaken to think that Edelman's proposals have anything to do with this notion of veridicality. How could it? Should we envisage that here to the left, we have the world, and, here to the right, we have a representation, and, here in front of

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

us, on the table, we have a machine that measures veridicality? We enter world and representation into the machine, and, lo and behold, its digital output says: "Your representation is 82% veridical. Congratulations!" If this is the wrong picture, then *how* should veridicality (and representation) be envisaged?

That there is something fishy about his veridicality is indirectly acknowledged by Edelman: in the concluding remarks he talks about *formal* veridicality. What is actually meant by "veridicality" in the target article is isomorphy of two similarity spaces: the space of distal shape similarities and the proximal space containing the output of a Chorus of classifiers. The implicit suggestion is that the distal space is already a veridical representation of the world, but little is said as to why that might be so. There is talk of "common parametrisation" and "multidimensional scaling" as if those techniques guarantee that those spaces mirror what is real or otherwise veridical. The nearest Edelman comes to "justifying" his approach is when, in Appendix B he claims that common parametrisation is the basis for the definition of distal similarity.

The techniques on which Edelman's veridicality relies are highly disputable. Here is an example (commenting in passing on Edelman's incidental references to colour and qualia and the work of Clark 1993). In a recent exchange on the issue Jameson (1997) says: "There is a body of similarity scaling research . . . indicating . . . H[ue], S[aturation] & B[rightness] . . . are real psychological constructs." The "body of similarity scaling research" however does *not* support Jameson's belief in hue, saturation and brightness being real psychological constructs but, if anything, the opposite (Saunders & van Brakel 1997).

There is another, more fundamental problem in talking innocently about the veridicality of any distal space. Assume we are talking Tsistsistas (Cheyenne), two centuries ago. We are standing in the middle of a desert, keen on perceiving salient objects and events. "*Vovetas,*" your companion says. Of course, you know a *vovetas* when you see one (van Brakel 1991). It is either a black vulture, or a common nighthawk, or a swarm of dragonflies, or red skimmers, or a tornado (skipping some details). So, what do you see? Well, you are not expected to see either this or that or that or that; what you are expected to see is *vovetas,* an ordinary manifest, observable. How would the Chorus of classifiers deal with *vovetas?* It shouldn't come up with a set composed of five categories (hawks, tornado, . . .), if only because not all hawks are *vovetas;* it should just learn the salient natural kind *vovetas.* And of course that is what the Chorus will do, if that is what its conductor wants it to do. Just as it will respond correctly to cats and dogs, if that is what it has been trained to do.

Edelman's proposals may be useful in making Choirs of classifiers better learners – at least relative to the tasks set by certain types of conductors; but any veridicality in that story is a veridicality of the sort that makes talk of "salience," "natural kinds," "structure of the world," "familiar categories," and so on, equally applicable to vultures and *vovetas,* that is, a multiculture of veridicalities.

Regular spaces versus computing with chaos

Cees van Leeuwen

Faculty of Psychology, Department of Psychonomics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands.
ceesvl@uvapsy.psy.uva.nl

Abstract: The attempt to provide a faithful mapping from distal shape space to proximal state space in terms of a higher order relationship defined over proximal similarity space stumbles on the context sensitivity of higher order relationships. Proportional analogy problems using quadruples of figures illustrate that for a number of interesting perceptual problems, the number of relevant dimensions cannot be reduced.

Edelman invokes the old concept of isomorphism in an inspiring effort to obtain a basis for a unified approach to visual representation. The mathematical properties of isomorphism involve a mapping between two domains. This mapping is one to one, and a relationship R should faithfully be preserved by it. This means that whenever there is a relation R between items in one domain, a corresponding relation R′ will occur between their images in the other. Countless isomorphisms can be defined between two manifolds, including many arbitrary ones. So isomorphisms have to be kept principled and meaningful. Such conditions were satisfied in the way the isomorphism concept was originally applied to the psychology of perception. In Köhler's (1929) psychoneural isomorphism, proximal objects were assumed to be isomorphic to electrostatic force fields in the brain. The relation preserved in the projection is principled. It is constrained by physical properties of the force field. Its meaning is determined by the holistic phenomenal characteristics of the proximal object.

Vagueness is often a better guarantee of immortality than specificity, as the latter risks refutation. Köhler's psychoneural isomorphism was sufficiently specific to allow testable predictions. Subsequent testing unequivocally disconfirmed the cerebral integration implied by his isomorphism thesis (Lashley et al. 1951; Pribram 1984). A neurally less constrained isomorphism (Henle 1984) may be taken to imply that a proximal representation of a figure in state space would faithfully represent the position of the figure in distal shape space. Let the relationship to be preserved be first order; for example, the distance between figures in shape space. Meaningful distances will express similarity between the figures in proximal state space.

Edelman assumes such a proximal similarity space but argues against identifying the corresponding R with first-order distance. On the basis of the phenomena, he may be quite justified in doing so. Subjects' impressions of figures shapes and of which figures in proximal space they resemble depend on context. For instance, in Figure 1, the L-shaped distal component will give rise to quite different impressions depending on whether it contacts the square. In Figure 1A it will still resemble an L, but in Figure 1B, it will resemble a square. Because minimal changes in context can have large, and highly specific effects on similarity, similarity cannot be the R that is to be preserved in a principled way. Contact-sensitivity, in other words, poses an obstacle to the application of the isomorphism principle (van Leeuwen 1990).

Edelman proposes an alternative R. Instead of first-order similarity, he proposes a higher order relation to be preserved across distal and proximal space. The R to be preserved is defined over a subset of the similarities of distal shapes. This should lead to an image R′ which faithfully represents, for example, the rank order between the distances in the shape space. As a consequence, sub-

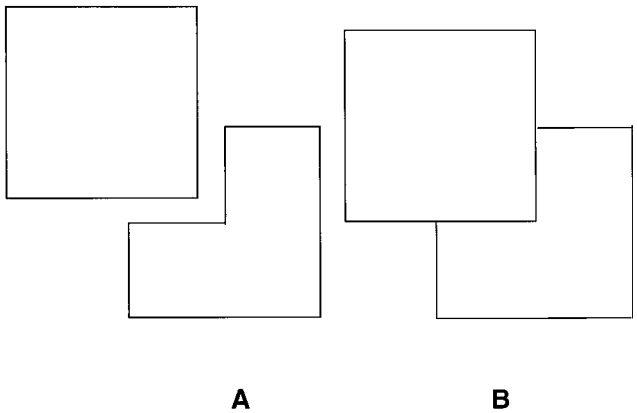


Figure 1 (van Leeuwen). The phenomenon of occlusion illustrating the principle of small changes, large consequences in first-order similarity.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

jects' rank ordering in similarity judgments across pairs of distal shapes should be stable and reliable.

Let us discuss the issue whether this isomorphism is principled or suffers from the same context-dependency problems as first-order similarity. The principled character of higher order isomorphism is contradicted by the context dependency of higher order similarity judgments. Such context dependency has been demonstrated, in proportional analogy problems (Indurkhya 1992). A proportional analogy is usually represented as A : B = C : D. The problem A : B = C : ? is to find a shape D that stands in the same similarity relation to C as a B does to A. Figure D1 would be a solution for problem 1, and Figure 2 for problem 2. Figure D2 would be a bad solution for problem 1 and D1 would be a bad solution for Problem 2. This implies that Figure C is similar to D1 and dissimilar to D2 in the context of problem 1, but the reverse is true in the context of problem 2; in other words, there is a reversal in the similarity space ordering for figures A, C, D1, and D2, depending on the presence of Figure B1 or B2.

Such changes in similarity ordering are not easily understood as a consequence of some global parametrization of similarity space. A solution should be in terms of a difference given to alternative dimensions of similarity space in problems 1 and 2. Shape C resembles D1 more than D2 in a dimension relevant to problem 1, and it resembles D2 more than D1 in another relevant dimension to problem 2. The problem is that that these dimensions are determined by the specific contrast between the respective B figures used in problems 1 and 2. Since there can be an infinite number of such contrasts in the set of all possible proportional analogy problems, this solution can only lead to an explosion of the number of relevant dimensions and hence to ad hocness.

In a rebuttal, one might say that these proportional analogy problems are not quite representative for perceptual tasks; they belong rather to the domain of cognition. This rebuttal, however, ignores a central capacity of the visual system. The proportional analogy problems require context-specific restructuring of visual similarity space. A structure for a figure is discovered in the problem that would not have been perceived in other circumstances. It requires a perceiver to extract information from the distal configuration beyond its predominant perceptual organization. This capacity is of significance for a variety of visual perception tasks.

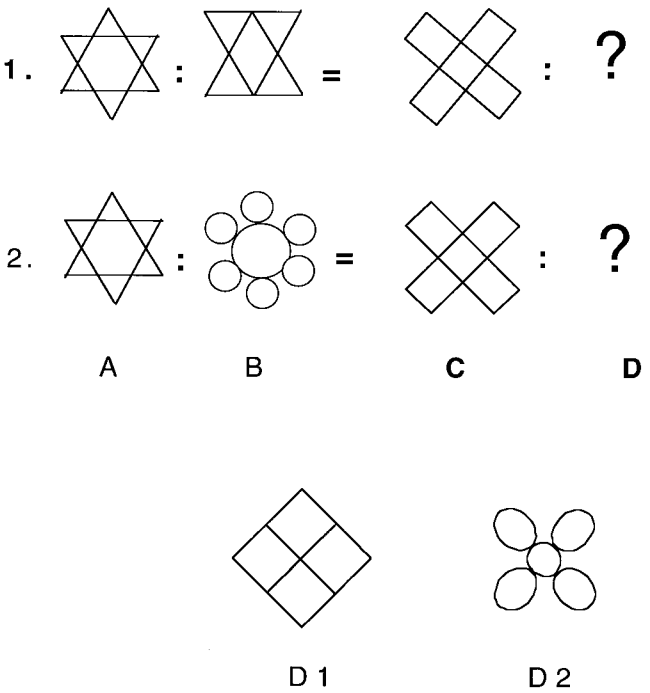


Figure 2 (van Leeuwen). Proportional analogy problems illustrating context dependency of higher order relations in similarity spaces.

Surplus structural information appears to be relevant in the aesthetic apprehension of a work of abstract art (Boselie 1983; Boselie & Leeuwenberg 1985). Creative design processes are shown also to rely on this capacity. Architects and industrial designers, for instance are known to produce external displays such as idea sketches. From these sketches, they extract surplus structural information, which contributes positively to the quality of their design product (Verstijnen et al. 1997). As a consequence, the creativity rating of a design product is larger if designers are allowed to sketch during the process of invention.

When a creative invention is made without sketching, subjects have to do so by means of visual imagery alone. The contrast of sketching and imagery-alone conditions illustrates the important role of the extraction of surplus structure in visual perception. Whereas it is easy to extract surplus structural information under sketching conditions, hardly any can be found with imagery alone. Hence, despite the similarities between visual perception and visual imagery (Kosslyn 1980) and the fact that creative inventions can be made by figural combination with imagery alone (Finke 1990), the extraction of surplus structure is one type of process that contrasts perception and imagery (Verstijnen et al. 1997). This contrast may explain, among others, why certain figural reversals that occur spontanously in the perception of certain ambiguous figures do not occur in imagery (Chambers & Reisberg 1985). These facts suggest that Edelman's approach has difficulties with what might be a very central function of perceptual activity. If the extraction of surplus information cannot be explained in Edelman's framework then he cannot deal with how children or adults learn to discriminate objects or how they come to appreciate the aesthetics of a certain work of abstract art or the signature of a new fashion style.

Edelman's view is quite traditional. It has more in common with the structural description approach than is apparent at first glance. He shares, for example, the view that for a set of components, a primary, context-free specification is possible. The difference is that, in the constructivist approach, these components are elementary features, from which a hierarchical, structural description is constructed. In Edelman's approach, these primary shapes are complex prototypes and no hierarchical representation is constructed. A (novel) object is encoded by graded similarity to a restricted set of prototypes. But this approach is not sufficient to distinguish the prototypes from (elementary) features, from which object representations are construed.

What we need is an approach that does justice to the strong context dependency of object representations in similarity space. The context sensitivity would imply that small changes in contextual circumstances could have large effects on similarity space. Similarity space would become extremely warped, and systems operating in this space would be structurally unstable. This would undermine the usefulness of approaches based on a similarity metric, whether they are first or higher order.

To do justice to the warped character of the activation landscapes of perceptual system spaces, we need a different set of models and different techniques for their evaluation than those based on distance metrics. The state space for figural representations is likely to be generated by a highly nonlinear function. I have advocated the approach of chaotic computation for perceptual shapes (van Leeuwen 1997a; van Leeuwen et al. 1997b). These models operate in highly warped state spaces, and show nonstable limit behaviours. To evaluate these systems, we look at transients. Techniques that could be used for their evaluation against electrophysiological data are, for instance, the dynamical measurement of stochastic coherence (Schack & Krause 1995). That any (higher or first-order) isomorphism will ever contribute to solving the puzzle of perception is highly unlikely from this perspective.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

How is representation learned?

James R. Williamson

Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215. **jrw@cns.bu.edu cns-web.bu.edu/pub/jrw/www/jrw.html**

Abstract: Edelman's memory-based approach to visual representation is preferable to parts-based alternatives. However, the existing algorithms for learning the shape prototypes are biologically implausible because they are nonlocal and nonconstructive. There is an alternative learning algorithm that constructs a mixture model of prototypes on-line, using only local information, and is more biologically plausible and may perform sufficiently well.

At first glance, the memory-based approach to visual representation proposed by Edelman might seem inadequate for recognizing real-world objects across all 3D orientations, given the variety and complexity of their shapes. Instead, the combinatorial nature of parts-based approaches makes them seem more appropriate for this task. However, experience tells us that it is extremely difficult to recover generic object parts from real-world images. Furthermore, as Edelman points out, there is no need to represent an object's shape or structure explicitly, only to represent enough information about its appearance so that we can preferentially respond to views of that object over those of other objects.

Hence the key question is: "How much information is enough?" Edelman argues that relatively few prototypes are enough, because of coarse coding. If the prototypes' response functions are sufficiently smooth, they can support interpolation, within the high-dimensional measurement space, of low-dimensional "proximal shape space" manifolds that correspond to parameters of the distal shape space. I find this argument convincing.

Since this is an approach for visual representation in biological systems, my next question is: How can biological systems learn these representations? Biological plausibility requires that we satisfy the following three conditions, among others.

1. **On-line.** The network should learn to recognize objects as they are encountered in real time. It is also possible that off-line optimization, utilizing more global statistics, takes place during memory consolidation.

2. **Local.** The network parameters should be updated using only local information.

3. **Constructive.** For any of the subregions of the shape space, how does the network know, a priori, how many prototypes should be allocated? If one region of the shape space has special environmental relevance, it may require a disproportionate number of prototypes.

Edelman states that, among connectionist classifiers, the radial basis function (RBF) networks of Poggio and Girosi (1989) are preferable on the grounds of biological plausibility (sect. 1.3, para. 5). Edelman and Poggio (1992) used RBF networks to successfully interpolate the view space of objects they were trained on. The architecture of RBF networks does indeed map straightforwardly to the neurobiology of the brain; however, the algorithms typically used to train these networks, such as those used by Edelman and Poggio, do not. While RBF networks can learn on-line, their gradient-descent learning equations require that error signals computed at each of the parameters in the output layer feed back to each of the parameters in the hidden layer. This is not very local due to the large number of backprojections that are required. Moreover, the learning equations in the hidden layer are considered to be biologically implausible (Poggio & Girosi 1989, p. 48). RBF networks also typically lack a constructive mechanism that builds, or self-organizes, a representation of appropriate size for a given problem domain, or that allocates resources to regions of the shape space according to their environmental relevance.

One class of networks that meets these three conditions is Adaptive Resonance Theory (ART) networks (Carpenter & Grossberg 1987). Of particular relevance here is a recently developed ART network, called Gaussian ARTMAP (GAM), which is a type of RBF network (Williamson 1996; 1997). Each GAM internal category node (which corresponds to a prototype) has a Gaussian-defined receptive field in the input space, as well as a mapping to an output prediction. GAM learns a Gaussian/multinomial mixture model of the joint input/output space using, essentially, an on-line approximation of the well-known Expectation-Maximization (EM) algorithm, an iterative technique for maximizing a mixture model's likelihood. Unlike RBF networks trained with gradient descent, GAM does not require a massive number of backprojections to transmit error signals. Rather, GAM receives only a single error signal if it makes an incorrect prediction. This raises a global vigilance level, which has the effect, via learning, of either sharpening relevant existing receptive fields or adding new prototypes, in order to improve discrimination in the appropriate region of the input space.

GAM has been combined with a biologically motivated model of early vision into a system for visual representation consistent with the approach described by Edelman (Grossberg & Williamson 1997). This system has been applied to image segmentation and classification based on textural and brightness attributes, where it outperformed alternative approaches that use rule-based, multilayer perceptron, and K-nearest-neighbor classifiers. In addition, those errors that the system did make were correlated with the pairwise distances between textures in MDS coordinates based on psychophysical measurements (Rao & Lohse 1996).

One potential advantage of mixture models is their flexibility. Because mixture models represent the joint density of the input/output space, they support mappings in multiple directions. For example, if an object category is primed, it can generate expectations of the input features. However, a note of caution is that this flexibility comes at a cost. Because mixture models represent the density across all the input dimensions, they may not support the level of smoothness in the output space that a memory-based approach requires, and that RBFs trained with gradient descent can obtain.

Author's Response

Shape representation by *S*econd-order *I*somorphism and the *C*horus model: *SIC*

Shimon Edelman

School of Cognitive and Computing Sciences, University of Sussex at Brighton, Falmer BN1 9QH, United Kingdom. **shimone@cogs.susx.ac.uk www.cogs.susx.ac.uk/users/shimone**

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 to refer selectively to retinotopically defined object fragments should lead to a comprehensive theory of shape processing.

R1. An overview of the commentaries

The relationships among the stances taken by the commentators on the various issues surrounding 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 with its author's last name.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

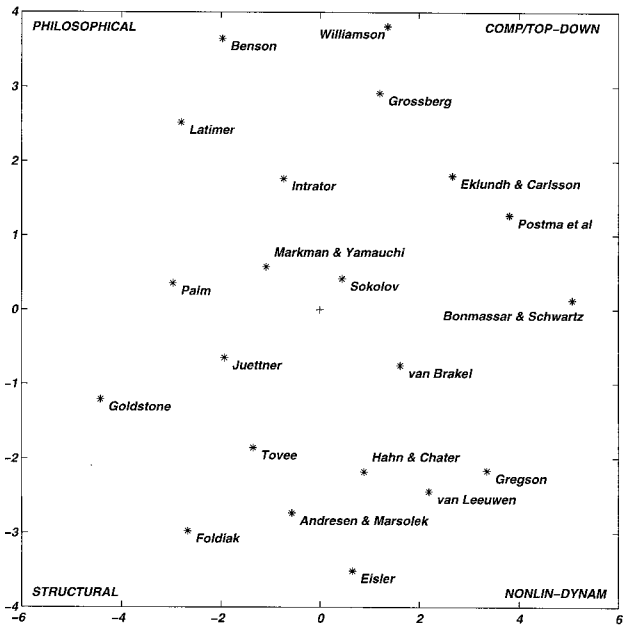


Figure R1. A 2-D rendition of an 11-dimensional "commentary space" derived from the 21 commentaries. [Stuart's commentary is not included; response will appear in second round in a forthcoming issue.] Each commentary was first described by 11 binary predicates, chosen so as to cover the major issues 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 2-D space by metric multidimensional scaling (MDS). The coefficient of congruence (Borg & Lingoes 1987) between the distances in the MDS-derived 2-D 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.

The center of the plot is occupied by commentaries that touch on 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 contains the proponents of combined metric and structural representations. All these issues, along with some of the specific concerns raised by the commentators, are discussed in this response.

R2. Veridicality

The strongest concerns in connection with veridicality are voiced by **Hahn & 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 (1972b) and by Watanabe (1985a), 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.

A typical argument against the notion of objective similarity is made by Murphy and Medin (1985), who note that the number of attributes shared by plums and lawn mowers could be infinite: both weigh less than 1,000 kilograms (and less than 1,001 kilograms), neither can hear well, both have a smell, and so on. Watanabe (1985a) 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).

Although they are 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, 1991, 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 1985a) or "prior spacing of qualities" (Quine 1969) removes the ambiguity by introducing a bias (Goldstone 1994). Indeed, in a precursor to the target article (Edelman 1995b), 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 & 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 article (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 means, however, 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 & Chater** answer in the affirmative, drawing an analogy between the discredited correspondence theory of truth and the second-order isomorphic representation of

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

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 (Bookstein 1996; Carne 1990; Kendall 1984; Le 1991; Le & Kendall 1993).

Unfortunately, all the commentators who had problems with my notion of veridicality ignored the proposal mentioned above, despite its appearance in the target article. 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 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, a swarm of dragonflies, or, for all a nonspeaker 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]

R3. Compositionality and the representation of structure

Földiák, Goldstone, Intrator, Markman & Yamauchi, and **Postma et al.** 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, **Földiák** is the only one who rejects representation by similarities to prototypes altogether. The arguments raised by Földiák 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, however, is 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 Földiák himself, as well as by **Postma et al.** (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 toward 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 & 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 with Földiák'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 (Nosofsky 1991; Shepard 1964), 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 (Schyns et al. 1998), only makes sense if the underlying representation reflects as many stimulus dimensions as possible, because different subsets of these dimensions will be relevant in different situations. Such a *sparse* code, advocated by Barlow (1959) and by others (including **Földiák**), can be achieved in two ways: by a combination of abstract features (such as "red," an example suggested by Földiák) 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 preferable a priori; 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 Quine (1960) 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 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 & Gauthier 1997). Only after receiving a different label for an already encountered object do they associate it with the object's color, material, or local features.

Holistic representation (Fig. 2) is hence a sensible opening strategy, which can serve as the basis for the development of more sophisticated analytical approaches. The need to augment a holistic similarity-based model with some capability for structure manipulation is stressed by **Goldstone** and **Markman & Yamauchi,** who list experimental findings concerning the perception and categoriza-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

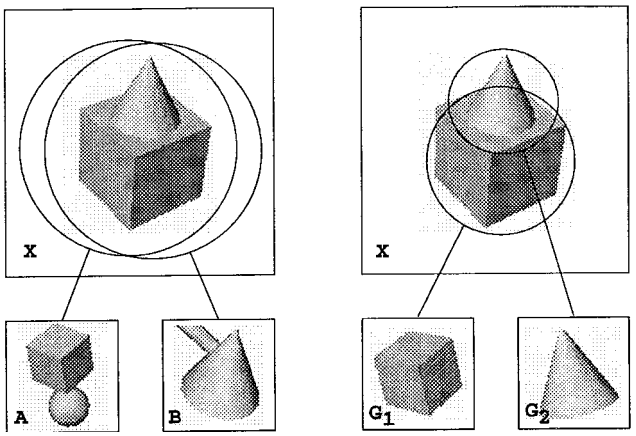


Figure R2. *Left:* Chorus of holistic prototypes; the new object X is represented by its similarities to objects A and B. This representation scheme, which I described in the target article, can support various recognition-related tasks, working from gray-level images of real objects (Duvdevani-Bar et al. 1998; Edelman & Duvdevani-Bar 1997a, 1997c). According to **Latimer,** "it could also provide an explicit, neurally plausible mechanism for responding directly and accurately to [objects] and their interrelationships"; **Jüttner** notes that it transforms "images into a rule-based representational format open to propositional reasoning" (cf. Barsalou's 1997 notion of perceptual symbol systems). However, as pointed out by **Földiák, Goldstone, Intrator, Markman & 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 3-D parts and their spatial relationships) has ever been 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.

tion 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 & Carlsson**) that the coexistence of multidimensional feature space and structural models is desirable. Such coexistence should not become an end in itself, however, 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 et al.** 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 need not be a problem for a large-scale Chorus-like model, however. Such a model can have at its disposal hundreds of prototype modules, of which only a small subset becomes active in any given discrimination task.[4] In comparison, Postma et al.'s proposal 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 (Jolicoeur & Humphrey, in press; Kurbat 1994; Tarr et al. 1997).

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 this 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 an 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 size of the retinal receptive field of each module (which can be

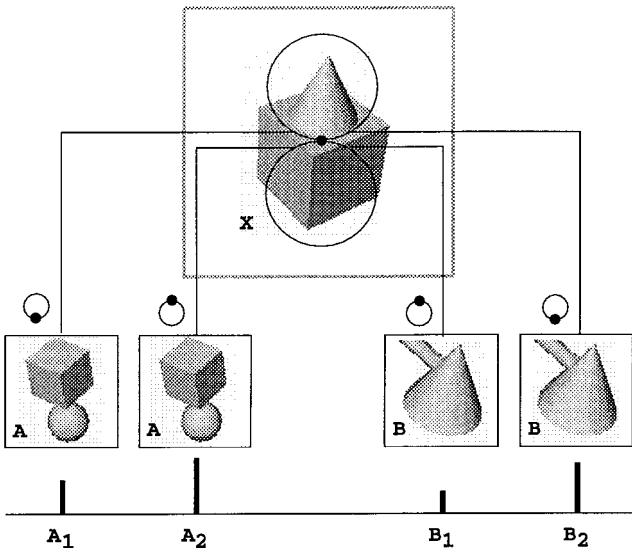


Figure R3. Chorus of prototypical fragments. In this proposed scheme, each object-specific module comes in several varieties, distinguished by the location of the module's receptive field (indicated schematically by the open circle) relative to the fixation point (indicated by the thick dot). For example, 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; the activities of A_1 and B_2 together determine the shape of its upper fragment – without recourse to either generic parts or to any kind of binding mechanism (beyond coactivation and retinotopy). This scheme is even closer to Barsalou's (1997) perceptual symbol system.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

done in a hard-wired fashion, as depicted in Fig. 3). In other words, the Chorus of prototypes can be turned into a Chorus of fragments, when necessary. For now, however, this is 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.

R4. Specific vs. abstract similarity

Andresen & 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, whereas 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, however, is not valid: 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 with totally disparate shapes such as "rage" and "RAGE" (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 with 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, although a separate "abstract" representation system like the one suggested by Andresen & Marsolek may exist, its existence remains a conjecture yet to be supported by data.

R5. 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 & 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 that 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 on, without being visually similar). **Postma et al.** draw attention to the need for invariance with respect to transformations such as translation and scaling, which leave shape unchanged, yet strongly affect 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 singled out by Hahn & 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 the comments of **Benson,** who (as far as I can gather) criticizes the lack of representation of these kinds 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 nonobvious relations (either abstract or concrete), proper representation of similarity under common transformations such as translation and scaling is a real concern, as is indeed noted in the target article. This issue, however, is more complicated than **Postma et al.** would have it, if only because human recognition is *not* completely invariant either to translation or to scaling (*pace* Biederman & 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 & Edelman 1997). Thus, a "blanket" approach to invariance via a global transformation (even a space-variant one, as proposed by **Bonmassar & Schwartz**) does not seem to be appropriate in modeling human performance. A more credible approach is suggested by neurophysiological findings (Ito et al. 1995; Tovee et al. 1994), where cellular responses, even if invariant under a certain amount of translation and scaling, pertain only to particular stimuli, hence excluding the possibility that invariance arises out of some global and universal mechanism.

R6. Similarity in context

As noted by **Eisler, van Leeuwen,** and **Tovée,** similarities depend strongly on the context of the comparison (what Eisler calls "the pertinent universe" and what **Jüttner** refers to as the choice of the map, which is prior to the choice of landmarks, apropos my analogy between categorization and navigation in a shape space). A similar point has been made by Mumford (1991a) (as discussed in the target article), by Tversky (1977), 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 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 how well 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.

R7. Top-down effects

Several of the commentaries question the rationale of choosing a basically feedforward architecture, such as that of Cho-

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

rus, 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 the implementation of a model should follow rather than precede the development of the theory. In contrast, connectionist modelers 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). The 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 similar models with respect to abstract computation-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 mixtures of both." There are certain obstacles to overcome, however, 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 the interaction between iteration and veridicality is. Second, in the context of representing (not yet categorizing) novel stimuli, an ART-based approach such as that of the system described by Bradski and Grossberg (1995) is actually detrimental, because it requires assigning the current stimulus to one of the familiar categories (or creating a new category), although it may be preferable to represent it within the existing framework (e.g., in terms of similarities to existing categories, as 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 a score of about 85% correct over a database of 50 such objects (Edelman & Duvdevani-Bar 1997d).

The power of interpolation among stored examples obviously depends on 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 (Duvdevani-Bar et al., in press; Edelman & Duvdevani-Bar 1997c). 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]

R8. What Chorus really does

Of the commentators who raise computational issues, **Bonmassar & Schwartz** are the only ones who appear to misunderstand the target article completely. The first of their misunderstandings has to do with multidimensional scaling (MDS), which is not "a particular form of cluster analysis" (Kruskal 1977), but rather a kind of distance-preserving dimensionality reduction. Their second misreading 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 2-D space; the model itself need not do that. If there are 1,000–2,000 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 by 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 1,000,000 (as it is in the primary visual cortex, or V1) rather than 1,000 (as in the inferotemporal, or IT, cortex). Thus, Bonmassar & 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, sect. 3.2).

The third misunderstanding by **Bonmassar & 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 & Schwartz conflate two issues here; that of multiple-view interpolation by the prototype modules, and that of translation invariance. The former is certainly not a linear phenomenon (Bülthoff & Edelman 1992; Poggio & Edelman 1990). 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

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

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 & 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 visual field. This invalidates Bonmassar & 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 I was writing the target article, I believed that a space-variant mapping proposed by Schwartz and Cavanagh and developed further by **Bonmassar & Schwartz** might 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 & Grossberg 1995). However, a review of the neurobiological literature (see Ch. 6 in Edelman, forthcoming), 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 (Ito et al. 1995; Tovée et al. 1994). 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 & 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 – contradicting the neurobiological and the psychophysical data. The upshot of this discussion is that Bonmassar & Schwartz's commentary is rather tangential to the issue at hand, and that the problem of size/translation invariance must still be considered open.

R9. Complexity and scalability

The commentary by **Eklundh & 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 30,000 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 to (1) represent 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 & Duvdevani-Bar 1997a; 1997d) 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.

R10. 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 & 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, however, the situation is not as simple. First, a mixture model such as Williamson's Gaussian ARTMAP inherits from ART the predisposition toward single-cause explanations of the input, at the expense of impartial representation (which would allow the input to belong to neither category); I 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 & 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 on the other is beyond the scope of this article.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

R11. 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 **Földiák**'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 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 Földiák, 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 and others (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 Földiák) are also relevant in qualifying Földiák'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 & Schwartz** argue (*contra* Földiák) 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 1,000,000 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 & 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 & 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 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 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 1993b)? In his commentary, **Tovée** 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., sect. 9.3.2), with the purpose of finding out whether this route, which is much more convenient computationally than the compositional one, can lead to sufficiently powerful representations. My conclusion, supported by computational experiments (Edelman & Duvdevani-Bar 1997a; 1997c), is that the holistic approach to representation advocated by Tovée is feasible. Additional considerations, such as the need for an explicit representation of structure in some tasks (discussed in sect. 3, 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 Fig. 3) is computationally feasible and biologically relevant.

R12. Methodological and metatheoretical issues

A combination of theoretical considerations with the 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 concerns 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 to specify 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). Jüttner's second remark refers to Anderson's (1978) plea for "indeterminacy concerning the representational format 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, even if they do 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 metatheoretical questions he poses along the way seem to me to obscure rather than clarify things.

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

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. [See also Millikan: "A Common Structure for Concepts of Individuals, Stuffs, and Real Kinds" BBS 21(1) 1998.]

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 & Stevens 1997) as a quibble because it leaves the most important thing unsaid: exactly *how* 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.

R13. And now, something completely different

The two remaining commentaries 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 that are metric only in a local neighborhood, but have no global properties implying constraints on monotone distance-separation relations, can be defined" (para. 4) and that "element-wise matchings between corresponding partitioned subsets of stimulus attributes . . . can sometimes be locally reconciled with metric space mappings" (para. 5). Chorus, which aims at representing the local metric structure of distal similarities (see Appendix B of the target article), fits 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 3-D shapes.[9] Consequently, I believe that both its possible deficiencies in modeling the perception of "geometric 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.

R14. Conclusions

In summary, 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 article 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 capacity 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 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 & 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), but were never shown to work on more than a dozen hand-labeled line drawings of stylized two-part shapes (Hummel & 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 likewise 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 that some headway is possible even here, however, 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 article. 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 (in computer slang, a quick and dirty fix for a bug.) for the problem of distal similarity, namely, the imposition of an observer bias.

NOTES

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.

2. Borges quotes, in the essay "The Analytical Language of John Wilkins" (1981) 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, (b) 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, pp. 23ff), which starts: "Let an object be

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

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."

4. This corresponds to combining Barlow's (1959) idea of a sparse code with Tanaka's (1996) estimate of 1,300–2,000 object-tuned modules in the inferotemporal cortex of the monkey.

5. Cf. the argument I made in section 2 in favor of objective shape spaces.

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.

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) that lists cortical areas activated in a mental rotation task – to the issues he raises elsewhere in his commentary.

References

Letters 'a' and 'r' appearing before author's initials refer to target article and response respectively.

Abbott, L. F., Rolls, E. T. & Tovée, M. J. (1996) Representational capacity of face coding in monkeys. *Cerebral Cortex* 6:498–505. [MJT]

Abeles, M. (1981) *Corticonics.* Cambridge University Press. [NI]

Albright, T. D. (1991) Motion perception and the mind-body problem. *Current Biology* 1:391–93. [aSE]

Aloimonos, J. Y. (1990) Purposive and qualitative vision. In: *Proceedings of the AAAI-90 Workshop on Qualitative Vision.* Morgan Kaufmann. [aSE]

Anderson, C. H. & Van Essen, D. C. (1987) Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences* 84:6297–301. [aSE]

Anderson, J. R. (1978) Arguments concerning representations for mental imagery. *Psychological Review* 85:249–77. [rSE, MJ]

Arrington, K. F. (1994) The temporal dynamics of brightness filling-in. *Vision Research* 34:3371–87. [SG]

Bajcsy, R. (1988) Active perception. *Proceedings of IEEE* 76(8):996–1005. [aSE]

Barlow, H. B. (1959) Sensory mechanisms, the reduction of redundancy, and intelligence. In: *The mechanisation of thought processes,*. Her Majesty's Stationery Office. [rSE, PF]

(1972) Single units and sensation. *Perception* 1:371–94. [PF]

(1979) The past, present and future of feature detectors. In: *Recognition of pattern and form,* ed. D. Albrecht. *Lecture Notes in Biomathematics,* vol.44. Springer. [aSE]

(1989) Unsupervised learning. *Neural Computation* 1:295–311. [PF]

(1990) Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research* 30:1561–71. [aSE]

(1994) What is the computational goal of the neocortex? In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis. MIT Press. [aSE]

Barsalou, L. W. (1997) *Perceptual symbol systems.* (submitted). [rSE]

Bartlett, F. C. (1932) *Remembering: An experimental and social study.* Cambridge University Press. [aSE]

Baxter, J. (1995) The canonical metric for vector quantization. *NeuroCOLT NC-TR-95- 047.* University of London. [aSE]

Bell, A. J. (1996) Learning the higher-order structure of a natural sound. *Network: Computation in Neural Systems* 7:889–904. [HE]

Berkeley, G. (1710/1965) The principles of human knowledge. In: *Berkeley's philosophical writings,* ed. D. Armstrong. Collier-Macmillan. [CL]

(1710/1996) *A treatise concerning the principles of human knowledge.* Oxford University Press. [aSE]

Beymer, D. & Poggio, T. (1996) Image representations for visual learning. *Science* 272:1905–1909. [aSE]

Biederman, I. (1987) Recognition by components: A theory of human image understanding. *Psychological Review* 94:115–47. [arSE, ABM, EP]

Biederman, I. & Cooper, E. E. (1991) Evidence for complete translational and reflectional invariance in visual object priming. *Perception* 20:585–93. [rSE, EP]

Biederman, I., Mezzanotte, R. J. & Rabinowitz, J. C. (1982) Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology* 14:143–77. [aSE]

Biederman, I., Rabinowitz, J. C., Glass, A. L. & Stacy, E. W. (1974) On the information extracted from a glance at a scene. *Journal of Experimental Psychology* 103:597- 600. [aSE]

Bienenstock, E. & Geman, S. (1995) Compositionality in neural systems. In: *The handbook of brain theory and neural networks,* ed. M. A. Arbib. MIT Press. [aSE]

Blackmore, S. J., Brelstaff, G., Nelson, K. & Troscianko, T. (1995) Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception* 24:1075–81. [aSE]

Bookstein, F. L. (1991) *Morphometric tools for landmark data: Geometry and biology.* Cambridge University Press. [aSE]

(1996) Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology* 58:313–65. [rSE]

Bonmasser, G. & Schwartz, E. (1997a) Lie groups, space-variant Fourier analysis and the exponential chirp transform. In: *Computer Vision and Pattern Recognition* 96, vol. 3. [GB, rSE]

(1997b) Space-variant Fourier analysis: The exponential chirp transform. *IEEE Transactions on Pattern Analysis and Machine Vision* 19:1080–89. [GB]

Borg, I. & Lingoes, J. (1987) *Multidimensional similarity structure analysis.* Springer. [arSE]

Borges, J. L. (1981) The analytical language of John Wilkins. In: *Borges: A reader,* ed. E. R. Monegal & A. Reid. Dutton. [rSE]

Boselie, F. (1983) Ambiguity, beauty, and interestingness of line drawings. *Canadian Journal of Psychology* 37:287–92. [CvL]

Boselie, F. & Leeuwenberg, E. (1985) Birkoff revisited: Beauty as a function of effect and means. *American Journal of Psychology* 98:1–39. [CvL]

Bourgain, J. (1985) On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics* 52:46–52. [aSE]

Bowers, J. S. (1996) Different percpetual codes support priming for words and pseudowords: Was Morton right all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:1336–53. [DRA]

Bradski, G. & Grossberg, S. (1995) Fast-learning VIEWNET architectures for recognizing three dimensional objects from multiple two-dimensional views. *Neural Networks 8*:1053–80. [rSE, SG]

Brigham, J. C. (1986) The influence of race on face recognition. In: *Aspects of face processing,* ed. H. D. Ellis, M. A. Jeeves & F. Newcombe. Martinus Nijhoff. [aSE]

Brooks, R. A. (1991) Intelligence without representation. *Artificial Intelligence* 47:139- 60. [rSE]

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

Caelli, T. & Bischof, W. (1996) Machine learning paradigms for pattern recognition and image understanding. *Spatial Vision* 10:87–103. [MJ]

Caelli, T. & Dreier, A. (1994) Variations on the evidence-based object recognition theme. *Pattern Recognition* 27:185–204. [MJ]

Carne, T. K. (1990) The geometry of shape spaces. *Proceedings of the London Mathematical Society* 61:407 -32. [arSE]

Carpenter, G. A. (1996) Distributed activation, search, and learning by ART and ARTMAP neural networks. *Proceedings of the International Conference on Neural Networks: Plenary, Panel and Special Sessions, IEEE.* [SG]

Carpenter, G. A. & Grossberg, S. (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing* 37:54–115. [JRW]

(1991) *Pattern recognition by self-organizing neural networks.* MIT Press. [SG]

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H. & Rosen, D. B. (1992) Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks* 3:698–713. [aSE]

Carpenter, G. A., Grossberg, S. & Rosen, D. B. (1991) Fuzzy ART: An adaptive resonance algorithm for rapid stable classification of analog patterns. In: *Proceedings of the International Joint Conference on Neural Networks* 411–16. [aSE]

Carpenter, G. A. & Ross, W. D. (1995) ART-EMAP: A neural network architecture of object recognition by evidence accumulation. *IEEE Transactions on Neural Networks* 6:805–18. [SG]

Cassirer, E. (1966) On the theory of the formation of concepts. In: *Pattern*

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

recognition: Theory, experiment, computer simulations and dynamic models of form perception and discovery, ed. L. Uhr. Wiley. [CL]
Cavanagh, P. (1995) Vision is getting easier every day. *Perception* 24:1227–32. [aSE]
Chambers, D. & Reisberg, D. (1985) Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance* 11:317–28. [CvL]
Chey, J., Grossberg, S. & Mingolla, E. (1997) Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America A* 14:2570–94. [SG]
Clark, A. (1993) *Sensory qualities.* Clarendon Press. [aSE, JvB]
Cohen, M. A. & Grossberg, S. (1986) Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short term memory. *Human Neurobiology* 5:1–22. [SG]
Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., Bookheimer, S. Y., Rosen, B. R. & Belliveau, J. W. (1996) Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain* 19:89–100. [rSE, RAMG]
Cohn, H. (1967) *Conformal mappings on Riemann surfaces.* McGraw-Hill. [aSE]
Cortes, C. & Vapnik, V. (1995) Support-vector networks. *Machine Learning* 20:273–97. [aSE]
Cortese, J. M. & Dyre, B. P. (1996) Perceptual similarity of shapes generated from Fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance* 22:133–43. [aSE]
Cummins, R. (1989) *Meaning and mental representation.* MIT Press. [arSE]
(1996) *Representations, targets, and attitudes.* MIT Press. [aSE]
Cutzu, F. & Edelman, S. (1996) Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Sciences* 93:12046–50. [arSE, MJ]
(1998) Representation of object similarity in human vision: Psychophysics and a computational model. *Vision Research.* (in press). [aSE]
Dayan, P., Hinton, G. E. & Neal, R. M. (1995) The Helmholtz machine. *Neural Computation* 7:889–904. [aSE, PF]
Dennett, D. C. (1991) *Consciousness explained.* Little, Brown & Company. [aSE]
Dill, M. & Edelman, S. (1997) Translation invariance in object recognition, and its relation to other visual transformations. *Artificial Intelligence, Memo No. 1610.* Artificial Intelligence Laboratory, Massachusetts Institute of Technology. [rSE]
Dow, B., Vautin, R. G. & Bauer, R. (1985) The mapping of visual space onto foveal striate cortex in the macaque monkey. *Journal of Neuroscience* 5:890–902. [GB]
Dretske, F. (1981) *Knowledge and the flow of information.* MIT Press. [aSE]
Duvdevani-Bar, S., Edelman, S., Howell, A. J. & Buxton, H. (1998) A similarity-based method for the generalization of face recognition over pose and expression. In: *Proceedings of the 3rd International Symposium on Face and Gesture Recognition (FG98), Washington, D. C.,* ed. S. Akamatsu & K. Mase. IEEE. (in press). [rSE]
Edelman, S. (1994) Biological constraints and the representation of structure in vision and language. *Psycoloquy* 5(57). [arSE]
(1995a) Representation of similarity in 3D object discrimination. *Neural Computation* 7:407–22. [aSE]
(1995b) Representation, similarity, and the Chorus of prototypes. *Minds and Machines* 5:45–68. [arSE]
(1997) Computational theories of object recognition. *Trends in Cognitive Science* 1:296–304. [rSE]
(1998) Vision reanimated. In: *Proceedings of the 7th Rosenön Workshop on Computer Vision,* ed. Y. Aloimonos, S. Carlsson & J.-O. Eklundh. L. Erlbaum (in press). [aSE]
(forthcoming) *Representation and recognition in vision.* MIT Press. [rSE]
Edelman, S., Bülthoff, H. H. & Bülthoff, I. (1996) Features of the representation space for 3D objects. MPIK-TR 40. *Max Planck Institute for Biological Cybernetics.* [aSE]
Edelman, S. & Duvdevani-Bar, S. (1997a) A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London* (B) 352(1358):1191–202. [arSE]
(1997b) Similarity, connectionism, and the problem of representation in vision. *Neural Computation* 9:701–20. [aSE]
(1997c) Similarity-based viewspace interpolation and the categorization of 3D objects. In: *Proceedings of the Similarity and Categorization Workshop, 75–81.* Department of Artificial Intelligence, University of Edinburgh. [rSE]
(1997d) Visual recognition and categorization on the basis of similarities to multiple class prototypes. *Artificial Intelligence Memo No. 1615,* Artificial Intelligence Laboratory, Massachusetts Institute of Technology. [rSE]
Edelman, S. & Intrator, N. (1997) Learning as extraction of low-dimensional representations. In: *Mechanisms of perceptual learning,* ed. D. Medin, R. Goldstone & P. Schyns. Academic Press (in press). [aSE]
Edelman, S. & Poggio, T. (1992) Bringing the grandmother back into the picture: A memory-based view of object recognition. *International Journal of Pattern Recognition and Artificial Intelligence* 6:37–61. [JRW]
Edelman, S. & Weinshall, D. (1991) A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics* 64:209–19. [arSE]
(1998) Computational approaches to shape constancy. In: *Perceptual constancies: Why things look as they do,* ed. V. Walsh & J. Kulikowski. Cambridge University Press (in press). [aSE]
Eisler, H. (1960) Similarity in the continuum of heaviness with some methodological and theoretical considerations. *Scandinavians Journal of Psychology* 1:69–81. [HE]
(1982) On the nature of subjective scales. *Scandinavian Journal of Psychology* 23:161–71. [HE]
Eisler, H. & Edberg, G. (1982) The visual perception of texture: A psychological investigation of an architectural problem. In: *Social attitudes and psychological measurement,* ed. B. Wegener. Erlbaum. [HE]
Eisler, H. & Lindman, R. (1990) Representations of dimensional models of similarity. In: *Psychophysical explorations of mental structures,* ed. H.-G. Geissler. Hogrefe & Huber. [HE]
Eisler, H. & Roskam, E. E. (1977a) Multidimensional similarity: An experimental and theoretical comparison of vector, distance, and set theoretical models. I. Models and internal consistency of data. *Acta Psychologica* 41:1–46. [HE]
(1977b) Multidimensional similarity: An experimental and theoretical comparison of vector, distance, and set theoretical models. II. Multidimensional analyses: The subjective space. *Acta Psychologica* 41:335–63. [HE]
Efron, B. & Tibshirani, R. (1993) *An introduction to the bootstrap.* Chapman and Hall. [aSE]
Ekman, G. & Lindman, R. (1961) Multidimensional ratio scaling and multidimensional similarity. *Reports from the Psychological Laboratories, University of Stockholm* 103. [aSE]
Farah, M. J. (1994) Specialization within visual object recognition: Clues from prosopagnosia and alexia. In: *The neuropsychology of high-level vision,* ed. M. J. Farah & G. Ratcliff. Erlbaum. [J-OE]
Field, D. J. (1994) What is the goal of sensory coding? *Neural Computation* 6:559–601. [PF]
Finke, R. A. (1990) *Creative imagery: Discoveries and inventions in visualization.* Erlbaum. [CvL]
Fiser, J., Biederman, I. & Cooper, E. E. (1996) To what extent can matching algorithms based on direct outputs of spatial filters account for human shape recognition? *Spatial Vision* 10:237–71. [EP]
Flynn, P. & Jain, A. K. (1993) Three-dimensional object recognition. In: *Handbook of pattern recognition and image processing, vol. 2: Computer vision,* ed. T. Young. Academic Press. [MJ]
Fodor, J. A. (1981) *RePresentations.* MIT Press. [aSE]
(1987) *Psychosemantics.* MIT Press. [aSE]
Földiak, P. (1990) Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics* 64:165–70. [PF]
Foldiak, P. & Young, M. P. (1995) Sparse coding in the primate cortex. In: *The handbook of brain theory and neural networks,* ed. M. A. Arbib. MIT Press. [PF]
Fomin, S. V., Sokolov, E. N. & Vaitkyavious, G. G. (1979) *Iskusstvennie organi chuvstv.* [*Artificial sense organs*]. (In Russian). Nauka. [ENS]
Francis, G. & Grossberg, S. (1996) Cortical dynamics of form and motion integration: Persistence, apparent motion, and illusory contours. *Vision Research* 36:149–73. [SG]
Friedman, J. H. & Tukey, J. W. (1974) A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers* 23:881–90. [PJB]
Fujita, I., Tanaka, K., Ito, M. & Cheng, K. (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360:343–46. [arSE]
Galin, E. & Akkouche, S. (1996) Métamorphose d'objets tridimensionnels: Quelques méthodes d'accélération. *Revue Techniques et Sciences Informatique* 15:329–50. [aSE]
Gallistel, C. R. (1990) *The organization of learning.* MIT Press. [aSE]
Garbin, C. P. (1990) Visual-touch perceptual equivalence for shape information in children and adults. *Perception and Psychophysics* 48:271–79. [aSE]
Garner, W. R. (1974) *The processing of information and structure.* Wiley. [PF]
Gawne, T. J. & Richmond, B. J. (1993) How independent are the messages carried by adjacent inferior temporal cortical-neurons. *Journal of Neuroscience* 13:2758–71. [PF]
Geman, D., Amit, Y. & Wilder, K. (forthcoming) Joint induction of shape features and tree classifiers. IEEE PAMI. [NI]
Gibson, J. J. (1966) *The senses considered as perceptual systems.* Houghton Mifflin. [aSE]
Goldstone, R. L. (1994) The role of similarity in categorization: providing a groundwork. *Cognition* 52:125–57. [arSE]
Goodman, N. (1972) *Problems and projects. Seven strictures on similarity.* Bobbs Merrill. [UH]

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

(1977) *The structure of appearance.* Reidel. [aSE]
Gove, A., Grossberg, S. & Mingolla, E. (1995) Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience* 12:1027–52. [SG]
Gregory, R. L. (1978) Illusions and hallucinations. In: *Handbook of perception, vol. IX,* ed. E. C. Carterette & M. P. Friedman. Academic Press. [aSE]
Gregson, R. A. M. (1975) *Psychometrics of similarity.* Academic Press. [aSE, RAMG]
(1976) A comparative evaluation of seven similarity models. *British Journal of Mathematical and Statistical Psychology* 29:139–56. [RAMG]
(1979) Content and distance similarity models: A correction to Sjöberg. *Scandinavian Journal of Psychology* 20:110–11. [RAMG]
(1980) Model evaluation via stochastic parameter convergence as on-line system identification. *British Journal of Mathematical and Statistical Psychology* 33:17- 35. [RAMG]
(1984) Similarities between odor mixtures with known components. *Perception and Psychophysics* 35:33–40. [RAMG]
(1985) Vergleich einiger mengentheoretischer und Distanz-Repräsentationen der Ähnlichkeiten von Broderson (1968). *Zeitschrift für experimentelle und angewandte Psychologie* 32:573–87. [RAMG]
(1988) *Nonlinear psychophysical dynamics.* Erlbaum. [aSE, RAMG]
(1992) *n-Dimensional nonlinear psychophysics: Theory and case studies.* Erlbaum. [RAMG]
(1993) The form of isosimilarity contours in nonlinear psychophysics. *Proceedings of the ISP Conference, Palma, Mallorca,* 51–56. [RAMG]
(1994) Similarities derived from 3-D nonlinear psychophysics: Variance distributions. *Psychometrika* 59:97–110. [RAMG]
(1995) Cascades and fields in perceptual psychophysics. *World Scientific.* [RAMG]
Gregson, R. A. M. & Britton, L. A. (1990) The size-weight illusion in 2D nonlinear psychophysics. *Perception and Psychophysics* 48:343–56. [aSE]
Gregson, R. A. M. & Harvey, J. P. (1992) Similarities of low-dimensional auditory chaotic sequences to quasirandom noise. *Perception and Psychophysics* 51:267– 78. [RAMG]
Grimes, J. (1995) On the failure to detect changes in scenes across saccades. In: *Perception. Vancouver studies in cognitive science, vol. 5,* ed. K. Akins. Oxford University Press. [aSE]
Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. (1972) Visual properties of cells in inferotemporal cortex of the macaque. *Journal of Neurophysiology* 35:96–111. [aSE]
Grossberg, S. (1980) How does a brain build a cognitive code? *Psychological Review* 87:1–51. [SG]
(1987) *The adaptive brain, vol. I and II.* Elsevier/North-Holland. [SG]
(1994) 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics* 55:48–120. [SG]
(1995) The attentive brain. *American Scientist* 83:438–49. [SG]
(1997) Cortical dynamics of 3-D figure-ground perception of 2-D pictures. *Psychological Review* 104:618–58. [SG]
Grossberg, S., Boardman, I. & Cohen, M. A. (1997a) Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 23:481–503. [SG]
Grossberg, S. & Merrill, J. W. L. (1996) The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. *Journal of Cognitive Neuroscience* 8:257–77. [SG]
Grossberg, S., Mingolla, E. & Ross, W. D. (1997b) Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences* 20:106–11. [SG]
Grossberg, S. & Todorovic, D. (1988) Neural dynamics of 1-D and 2-D brightness perception. *Perception and Psychophysics* 43:723–42. [SG]
Grossberg, S. & Williamson, J. R. (1997) A self-organizing neural system for learning to recognize textured scenes. *Boston University Technical Report,* CAS/CNS-TR-97- 001. [JRW]
Hampton, J. (1993) Prototype models of concept representation. In: *Categories and concepts,* ed. I. V. Mechelen, J. Hampton, R. S. Michalski & P. Theuns. Academic Press. [PF]
Hanson, S. J. & Gluck, M. A. (1993) Spherical units as dynamic consequential regions: Implications for attention, competition and categorization. In: *Advances in neural information processing systems 5,* ed. S. J. Hanson, J. D. Cowan & C. L. Giles. Morgan Kaufmann. [aSE]
Harnad, S., ed. (1987) *Categorical perception: The groundwork of cognition.* Cambridge University Press. [aSE]
Harnad, S. (1990) The symbol grounding problem. *Physica D* 42:335–46. [aSE, CL]
(1992) Connecting object to symbol in modelling cognition. In: *Connectionism in context,* ed. A. Clark & R. Lutz. Springer-Verlag. [CL]
Hasselmo, M. E. (1995) Neuromodulation and cortical function: Modeling the physiological basis of behavior. *Behavioral Brain Research* 67:1–27. [rSE]
Hebb, D. O. (1949) *The organization of behavior.* Wiley. [aSE]
Henle, M. (1984) Isomorphism: Setting the record straight. *Psychological Research* 46:317–27. [CvL]
Hinton, G. (1992) How neural networks learn from experience. *Scientific American* 267:145–51. [PF]
Hinton, G. E., Dayan, P., Frey, B. J. & Neal, R. (1995) The wake-sleep algorithm for unsupervised neural networks. *Science* 268:1158–61. [aSE]
Holland, J. H., Holyoak, K. J., Nisbett, R. E. & Thagard, P. R. (1986) *Induction: Processes of inference, learning, and discovery.* MIT Press. [aSE]
Hubel, D. H. & Wiesel, T. N. (1959) Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology* 148:574–91. [aSE]
(1968) Receptive fields and functional architecture of monkey striate cortex. *Journal of Neurophysiology* 195:215–43. [RLG]
Hume, D. (1748) *An enquiry concerning human understanding.* The Internet. Available electronically at URL http://coomba.anu.edu.au/Depts/RSSS/ Philosophy/ Texts/ EnquiryTOC.html. [rSE]
Hummel, J. E. & Biederman, I. (1992) Dynamic binding in a neural network for shape recognition. *Psychological Review* 99:480–517. [rSE]
Hummel, J. E. & Stankiewicz, B. J. (1996) Categorical relations in shape perception. *Spatial Vision* 10:201–36. [EP]
Indurkhya, B. (1992) *Metaphor and cognition. An interactionist approach.* Kluwer. [CvL]
Intrator, N. (1993) Combining exploratory projection pursuit and projection pursuit regression. *Neural Computation* 5:443–55. [aSE]
Intrator, N. & Cooper, L. N. (1992) Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks* 5:3–17. [aSE]
Intrator, N. & Edelman, S. (1997) Learning low dimensional representations of visual objects with extensive use of prior knowledge. *Network* 8:259–81. [rSE]
Ito, M., Tamura, H., Fujita, I. & Tanaka, K. (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology* 73:218–26. [rSE]
Izmailov, C. A. & Sokolov, E. N. (1991) Spherical model of color and brightness discrimination. *Psychological Science* 2:249–59. [ENS]
Jain, A. K. & Hoffmann, D. (1988) Evidence-based recognition of objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10:783–802. [MJ]
Jameson, K. A. (1997) What Saunders and van Brakel chose to ignore in color and cognition research. *Behavioral and Brain Sciences* 20(2):195–96. [JvB]
Jolicoeur, P., Gluck, M. A. & Kosslyn, S. M. (1984) Pictures and names: Making the connection. *Cognitive Psychology* 16:243–75. [DRA]
Jolicoeur, P. & Humphrey, G. K. (1998) Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In: *Perceptual constancies,* ed. V. Walsh & J. Kulikowski. Cambridge University Press (in press). [arSE]
Jüttner, M., Caelli, T. & Rentschler, I. (1997) Evidence-based pattern classification: A structural approach to human perceptual learning and generalization. *Journal of Mathematical Psychology* 41:244–58. [MJ]
Jüttner, M. & Rentschler, I. (1996) Reduced perceptual dimensionality in extrafoveal vision. *Vision Research* 36:1007–21. [MJ]
Kaplan, A. S. & Medin, D. L. (1997) The coincidence effect in similarity and choice. *Memory and Cognition* 25(4):570–76. [ABM]
Kendall, D. G. (1984) Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society* 16:81–121. [arSE]
(1989) A survey of the statistical theory of shape. *Statistical Science* 4:87–120. [aSE]
Kobatake, E. & Tanaka, K. (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology* 71:2269–80. [aSE]
Kobatake, E., Tanaka, K. & Tamori, Y. (1992) Long-term learning changes the stimulus selectivity of cells in the inferotemporal cortex of adult monkeys. *Neuroscience Research S*17:237. [aSE]
Koch, C. & Ullman, S. (1985) Selecting one among the many: A simple network implementing shifts in selective visual attention. *Human Neurobiology* 4:219–27. [aSE]
Koenderink, J. J. (1984) Simultaneous order in nervous nets from a functional standpoint. *Biological Cybernetics* 50:35–41. [EP]
Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L. (1996) Pictorial surface attitude and local depth comparisons. *Perception and Psychophysics* 58:163–73. [aSE]
Köhler, W. (1929) *Gestalt psychology.* Liveright. [CvL]
Koriat, A. & Goldsmith, M. (1995) Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences* 19:167–228. [aSE]
Kosslyn, S. M. (1980) *Image and mind.* Harvard University Press. [CvL]
Krumhansl, C. L. (1978) Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review* 85:445–63. [aSE, ABM]

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

Krushkal, S. L. (1979) *Quasiconformal mappings and Riemann surfaces.* Wiley. [aSE]
Kruskal, J. B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29(1):1–27. [aSE]
(1977) The relationship between multidimensional scaling and clustering. In: *Classification and clustering,* ed. J. V. Ryzin. Academic Press. [rSE]
Kruskal, J. B. & Wish, M. (1978) *Multidimensional scaling.* Sage Publications. [aSE]
Kurbat, M. A. (1994) Is RBC/JIM a general-purpose theory of human entry-level object recognition? *Perception* 23:1339–68. [rSE]
Landahl, H. D. (1945) Neural mechanisms for the concepts of difference and similarity. *Bulletin of Mathematical Biophysics* 7:83–88. [HE]
Landau, B., Smith, L. B. & Jones, S. (1988) The importance of shape in early lexical learning. *Cognitive Development* 3:299–321. [aSE]
Lando, M. & Edelman, S. (1995) Receptive field spaces and class-based generalization from a single view in face recognition. *Network* 6:551–76. [aSE]
Lashley, K. S., Chow, K. L & Semmes, J. (1951) An examination of the electrical field theory of cerebral integration. *Psychological Review* 58:123–36.
Latanov, A. V., Leonova, A. Y., Svtikhin, D. V. & Sokolov, E. N. (1997) Sravnitelnaya neurobiologia tsvetovogo zrenya cheloveka i zhivotnykh [Comparative neurobiology of color vision in humans and animals]. *Zhurnal Vysshei Nervnoi Deyatel'nosti* 47(2):308–19. [ENS]
Latimer, C. & Stevens, C. (1997) Some remarks on wholes, parts and their perception. *Psycoloquy* 8:13. [rSE, CL]
Le, H. (1991) On geodesics in Euclidean shape spaces. *Journal of the London Mathematical Society* 44:360–72. [rSE]
Le, H. & Kendall, D. G. (1993) The Riemannian structure of Euclidean shape spaces: A novel environment for statistics. *The Annals of Statistics* 21:1221–71. [arSE]
Lettvin, J. Y., Maturana, H. R., McCulloch, W. S. & Pitts, W. H. (1959) What the frog's eye tells the frog's brain. *Proceedings of the IRE* 47:1940–59. [aSE]
Lindsay, P. H. & Norman, D. A. (1977) *Human information processing: An introduction to psychology.* Academic Press. [aSE]
Linial, N., London, E. & Rabinovich, Y. (1994) The geometry of graphs and some of its algorithmic applications. *Foundations of Computer Science* 35:577–91. [aSE]
Littlestone, N. (1988) Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning* 2:285–318. [aSE]
Locke, J. (1690/1994) *An essay concerning human understanding.* Modern Library/Dent. [aSE, CL]
Logothetis, N. K., Pauls, J. & Poggio, T. (1995) Shape recognition in the inferior temporal cortex of monkeys. *Current Biology* 5:552–63. [aSE, GP]
Maffei, L. (1978) Spatial frequency channels: Neural mechanisms. In: *Handbook of sensory physiology: Perception,* ed. R. Held, H. W. Leibowitz & H.-L. Teuber. Springer-Verlag. [aSE]
Markman, A. B. (in press) *Knowledge representation.* Erlbaum. [ABM]
Markman, A. B. & Dietrich, E. (in preparation) In defense of representation. [ABM]
Markman, A. B. & Gentner, D. (1993a) Structural alignment during similarity comparisons. *Cognitive Psychology* 25:431–67. [aSE]
(1993b) Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language* 32(4):517–35. [ABM]
(1996) Commonalities and differences in similarity comparisons. *Memory and Cognition* 24(2):235–49. [ABM]
Markman, E. (1989) *Categorization and naming in children.* MIT Press. [rSE]
Marr, D. (1970) A theory for cerebral neocortex. *Proceedings of the Royal Society of London B* 176:161–234. [aSE]
(1976) Early processing of visual information. *Philosophical Transactions of the Royal Society of London B* 275:483–524. [aSE, J-OE]
(1982) *Vision.* W. H. Freeman. [arSE, RLG]
Marr, D. & Nishihara, H. K. (1978) Representation and recognition of the spatial organization of three-dimensional structure. *Proceedings of the Royal Society of London B* 200:269–94. [arSE]
Marsolek, C. J. (1995) Abstract visual-form representations in the left cerebral hemisphere. *Journal of Experimental Psychology: Human Perception and Performance* 2:375–86. [DRA]
(1997) Dissociable neural subsystems underlie abstract and specific object recognition. (submitted). [DRA]
Marsolek, C. J. & Burgund, E. D. (1997) Computational analyses and hemispheric asymmetries in visual-form recognition. In: *Cerebral asymmetries in sensory and perceptual processing,* ed. S. Christman. Elsevier. [DRA]
Marsolek, C. J., Schachter, D. L. & Nicholas, C. D. (1996) Form-specific visual priming for new associations in the right cerebral hemisphere. *Memory and Cognition* 24:539–56. [DRA]
Maze, J. R. (1983) *The meaning of behaviour.* Allen & Unwin. [CL]
McCulloch, W. S. (1965) *Embodiments of mind.* MIT Press. [rSE]
Medin, D. L., Goldstone, R. L. & Gentner, D. (1993) Respects for similarity. *Psychological Review* 100:254–78. [aSE]
Mel, B. (1997) SEEMORE: Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition. *Neural Computation* 9:777–804. [aSE]
Michell, J. (1988) Maze's direct realism and the character of cognition. *Australian Journal of Psychology* 40:227–49. [CL]
Millikan, R. (1984) *Language, thought, and other biological categories.* MIT Press. [rSE]
(1995) *White Queen psychology and other essays for Alice.* MIT Press. [aSE]
Moses, Y., Ullman, S. & Edelman, S. (1996) Generalization to novel images in upright and inverted faces. *Perception* 25:443–62. [aSE]
Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. (1985) The analysis of moving visual patterns. In: *Pattern recognition mechanisms,* ed. C. Chagas, R. Gattas & C. G. Gross. Vatican Press. [aSE]
Mumford, D. (1991a) Mathematical theories of shape: Do they model perception? In: *Geometric methods in computer vision.* SPIE. [arSE]
(1991b) On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological Cybernetics* 65:135–45. [aSE]
(1992) On the computational architecture of the neocortex. II. The role of the cortico-cortical loops. *Biological Cybernetics* 66:241–51. [aSE]
(1994) Neuronal architectures for pattern-theoretic problems. In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis. MIT Press. [arSE]
Murase, H. & Nayar, S. (1995) Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision* 14:5–24. [aSE]
Murphy, G. L. & Medin, D. L. (1985) The role of theories in conceptual coherence. *Psychological Review* 92:289–316. [rSE]
Newsome, W. T. & Pare, E. B. (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience* 8:2201–11. [aSE]
Nosofsky, R. M. (1986) Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115(1):39–57. [ABM]
(1988) Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition* 14:700–08. [aSE]
(1991) Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology* 23:94–140. [arSE]
(1992) Similarity scaling and cognitive process models. *Annual Review of Psychology* 43:25–53. [aSE]
Olshausen, B. A. & Field, D. J. (1994) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–09. [PF]
O'Regan, J. K. (1992) Solving the real mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46:461–488. [aSE]
Palm, G. (1980) On associative memory. *Biological Cybernetics* 36:19–31. [GP]
(1987a) Computing with neural networks. *Science* 235:1227–28. [GP]
(1987b) On associative memories. In: *Physics of cognitive processes,* ed. E. R. Caianello. World Scientific Publishing. [GP]
(1990) Local learning rules and sparse coding in neural networks. In: *Advanced neural computers,* ed. R. Eckmiller. North-Holland. [GP]
Palm, G. & Palm, M. (1991) Parallel associative networks: The Pan-system and the Bacchus-chip. In: *Proceedings of the second international conference on microelectronics for neural networks,* ed. U. Ramacher, U. Rückert & J. A. Nossek. Kyrill & Method Verlag. [GP]
Palm, G., Schwenker, F., Sommer, F. T. & Strey, A. (1997) Neural associative memory. In: *Associative processing and processors,* ed. A. Krikelis & C. C. Weems. IEEE Computer Society Press. [GP]
Palmer, S. E. (1978) Fundamental aspects of cognitive representation. In: *Cognition and categorization,* ed. E. Rosch & B. B. Lloyd. Erlbaum. [aSE]
Pentland, A. & Sclaroff, S. (1991) Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13:715–29. [aSE]
Perrett, D. I., Mistlin, A. J. & Chitty, A. J. (1989) Visual neurones responsive to faces. *Trends in Neurosciences* 10:358–64. [aSE]
Perrett, D. I., Rolls, E. T. & Caan, W. (1982) Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research* 47:329–42. [aSE]
Phillips, F. & Todd, J. T. (1996) Perception of local three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance* 22:230–44. [aSE]
Poggio, T. (1990) A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology* 55:899–910. [aSE]
Poggio, T. & Edelman, S. (1990) A network that learns to recognize three-dimensional objects. *Nature* 343:263–66. [arSE]
Poggio, T., Fahle, M. & Edelman, S. (1992) Fast perceptual learning in visual hyperacuity. *Science* 256:1018–21. [aSE]

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

Poggio, T. & Girosi, F. (1989) A theory of networks for approximation and learning. *A. I. Memo No. 1140,* MIT, 1989. [JRW]
(1990) Regularization algorithms for learning that are equivalent to multilayer networks. *Science* 247:978–82. [aSE]
Pollatsek, A., Rayner, K. & Collins, W. E. (1984) Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General* 113:426–42. [aSE]
Postma, E. O., Herik, H. J., van den & Hudson, P. T. W. (1997) SCAN: A scalable model of attentional selection. *Neural Networks* 10:993–1015. [EP]
Pribram, K. (1984) What is iso and what is morphic in isomorphism? *Psychological Research* 46:329–32. [CvL]
Putnam, H. (1988) *Representation and reality.* MIT Press. [aSE]
Quine, W. V. O. (1960) *Word and object.* MIT Press. [rSE]
(1969) Natural kinds. In: *Ontological relativity and other essays.* Columbia University Press. [arSE]
Rao, A. R. & Lohse, G. (1996) Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research* 36:1649–69. [JRW]
Rao, R. P. N. & Ballard, D. H. (1995) An active vision architecture based on iconic representations. *Artificial Intelligence* 78:461–505. [EP]
Rensink, R., O'Regan, K. & Clark, J. J. (1995) Image flicker is as good as saccades in making large scene changes invisible. *Perception* 24(suppl.):26–27. [aSE]
Rentschler, I., Jüttner, M. & Caelli, T. (1994) Probabilistic analysis of human supervised learning and classification. *Vision Research* 34:669–87. [MJ]
Reshetnyak, Y. G. (1989) *Space mappings with bounded distortion. Translations of mathematical monographs, vol. 73.* American Mathematical Society. [aSE]
Riesenhuber, M. & Dayan, P. (1997) Neural models for the part-whole hierarchies. In: *Advances in neural information processing 9,* ed. M. Jordan. MIT Press. (in press). [aSE]
Rolls, E. T., Baylis, G. C., Hasselmo, M. E. & Nalwa, V. (1989) The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research* 76:153–64. [aSE]
Rolls, E. T. & Tovée, M. J. (1995) The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Experimental Brain Research* 103:409–20. [MJT]
Rojet, A. S. & Schwartz, E. L. (1990) Design considerations for a space-variant visual sensor with complex-logarithmic geometry. *Tenth International Conference on Pattern Recognition* 2:278–85. [GB]
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology* 8:382–439. [DRA, aSE]
Rumelhart, D. E. (1980) Schemata: The building blocks of cognition. In: *Theoretical issues in reading and comprehension,* ed. R. J. Spiro, B. Bruce & W. F. Brewer. Erlbaum. [aSE]
Runeson, S. (1977) On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology* 18:172–79. [HE]
Saarinen, J. & Julesz, B. (1991) The speed of attentional shifts in the visual field. *Proceedings of the National Academy of Sciences USA* 88:1812–14. [EP]
Sakai, K., Naya, Y. & Miyashita, Y. (1994) Neuronal tuning and associative mechanisms in form representation. *Learning and Memory* 1:83–105. [aSE]
Salzman, C. D., Britten, K. H. & Newsome, W. T. (1990) Cortical microstimulation influences perceptual judgements of motion direction. *Nature* 346:174–77. [arSE]
Sato, T. (1989) Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research* 77:23–30. [MJT]
Saund, E. (1995) A multiple cause mixture model for unsupervised learning. *Neural Computation* 7:51–71. [PF]
Saunders, B. A. C. & van Brakel, J. (1997) Author's reply: Colour - an exosomatic organ? *Behavioral and Brain Sciences* 20:217–32. [JvB]
Schack, B. & Krause, W. (1995) Dynamic power and coherence analysis of ultra short-term cognitive processes - a methodical study. *Brain Topography* 8:127–36. [CvL]
Schiele, B. & Crowley, J. L. (1996) Object recognition using multidimensional receptive field histograms. In: *Proceedings of ECCV '96, lecture notes in computer science, vol. 1,* ed. B. Buxton & R. Cipolla. Springer. [aSE]
Schmidhuber, J. (1992) Learning factorial codes by predictability minimization. *Neural Computation* 4:863–79. [PF]
Schwartz, E. L. (1985) Local and global functional architecture in primate striate cortex: Outline of a spatial mapping doctrine for perception. In: *Models of the visual cortex,* ed. D. Rose & V. G. Dobson. Wiley. [aSE, EP]
(1977) Spatial mapping in primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics* 25:181–94. [GB]
(1994) Computational studies of the spatial architecture of primate visual cortex: Columns, maps, and protomaps. In: *Primary visual cortex in primates, vol. 10: Cerebral cortex,* ed. A. Peters & K. Rocklund. Plenum Press. [GB]
Schwartz, E. L., Desimone, R., Albright, T. & Gross, C. G. (1983) Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences* 80:5776–78. [GB, rSE]
Schwartz, E. L., Greve, D. & Bonmassar, G. (1995) Space-variant active vision: Definition, overview and examples. *Neural Networks* 8(7/8):1297–1308. [GB]
Schyns, P. G., Goldstone, R. L. & Thibaut, J.-P. (1998) The development of features in object concepts. *Behavioral and Brain Sciences* 21(1):1–53. [rSE]
Selfridge, O. G. (1959) Pandemonium: A paradigm for learning. In: *The mechanisation of thought processes.* Her Majesty's Stationery Office. [aSE]
Shapely, R., Caelli, T., Grossberg, S., Morgan, M. & Rentschler, I. (1990) Computational theories of visual perception. In: *Visual perception: The neurophysiological foundations,* ed. L. Spillman & J. B. Werner. Academic Press. [MJ]
Shepard, R. N. (1962) The analysis of proximities: Multidimensional scaling with unknown distance function. part I. *Psychometrika* 27(2):125–40. [aSE]
(1964) Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology* 1:54–87. [HE, rSE]
(1968) Cognitive psychology: A review of the book by U. Neisser. *American Journal of Psychology* 81:285–89. [aSE, MJ]
(1980) Multidimensional scaling, tree-fitting, and clustering. *Science* 210:390–97. [aSE]
(1984) Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review* 91:417–47. [aSE]
(1987) Toward a universal law of generalization for psychological science. *Science* 237:1317–23. [arSE, MJ]
Shepard, R. N. & Arabie, P. (1979) Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review* 86:87–123. [aSE]
Shepard, R. N. & Cermak, G. W. (1973) Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology* 4:351–77. [arSE]
Shepard, R. N. & Chipman, S. (1970) Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology* 1:1–17. [HE, aSE]
Shepard, R. N. & Kannappan, S. (1993) Connectionist implementation of a theory of generalization. In: *Advances in neural information processing systems 5,* ed. S. J. Hanson, J. D. Cowan & C. L. Giles. Morgan Kaufmann. [aSE]
Simard, P., Victorri, B., LeCun, Y. & Denker, J. (1992) Tangent prop - a formalism for specifying selected invariances in an adaptive network. In: *Neural information processing systems, vol. 4,* ed. J. Moody, R. Lippman & S. J. Hanson. Morgan Kaufmann. [aSE]
Simonson, I. & Tversky, A. (1992) Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* 29:281–95. [ABM]
Sjöberg, L. & Thorslund, C. (1979) A classificatory theory of similarity. *Psychological Research* 40:223–47. [HE]
Smith, L. B., Gasser, M. & Sandhofer, C. M. (1997) Learning to talk about the properties of objects: A network model of the development of dimensions. In: *Mechanisms of perceptual learning,* ed. D. Medin, R. Goldstone & P. Schyns. Academic Press. [rSE]
Snippe, H. P. & Koenderink, J. J. (1992) Discrimination thresholds for channel-coded systems. *Biological Cybernetics* 66:543–51. [aSE]
Sokolov, E. N. (1994) Vector coding in neuronal nets: Color vision. In: *Origins: Brain and self organization,* ed. K. H. Pribram. Erlbaum. [ENS]
Spinoza, B. (1677/1981) *The ethics.* J. Simon, Publisher.. [aSE]
Stellman, U. (1992) *Ähnlichkeitserhaltende Codierung.* Ph.D. dissertation, University of Ulm. [GP]
Sugihara, T., Edelman, S. & Tanaka, K. (1998) Representation of objective similarity in the monkey. *Biological Cybernetics* 78:1–7. [aSE]
Sundararaman, D. (1980) *Moduli, deformations and classifications of compact complex manifolds.* Pitman. [aSE]
Suppes, P., Pavel, M. & Falmagne, J. (1994) Representations and models in psychology. *Annual Review of Psychology* 45:517–44. [aSE]
Sutcliffe, J. P. (1986) Differential ordering of objects and attributes. *Psychometrika* 51:209–40. [CL]
Tanaka, J. & Gauthier, I. (1997) Expertise in object and face recognition. In: *Mechanisms of perceptual learning,* ed. D. Medin, R. Goldstone & P. Schyns. Academic Press. [rSE]
Tanaka, K. (1992) Inferotemporal cortex and higher visual functions. *Current Opinion in Neurobiology* 2:502–505. [aSE]
(1993a) Neuronal mechanisms of object recognition. *Science* 262:685–88. [aSE]
(1993b) Column structure of inferotemporal cortex: "Visual alphabet" or "differential amplifiers"? In: *Proceedings of International Joint Conference on Neural Networks-93,* Nagoya. [rSE]
(1996) Inferotemporal cortex and object vision. *Annual Review of Neuroscience* 19:109–39. [rSE] Also in: *Vision and movement mechanisms in the cerebral cortex,* ed. R. Caminiti, K. P. Hoffman, F. Lacquaniti & J. Altman. HFSP. [MJT]

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press

Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology* 66:170–89. [aSE, MJT]

Tarr, M. J., Bülthoff, H. H., Zabinski, M. & Blanz, V. (197) To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science* 8:282–89. [rSE]

Thurber, J. (1943) *Men, women and dogs.* Harcourt, Brace. [HE]

Tootell, R. B., Silverman, M. S., Switkes, E. & deValois, R. (1985) Deoxyglucose, retinotopic mapping and the complex log model in striate cortex. *Science* 227:1066. [GB]

Tovée, M. J. (1995) Face recognition: What are faces for? *Current Biology* 5:480–82. [MJT]

Tovée, M. J., Rolls, E. T. & Azzopardi, P. (1994) Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert monkey. *Journal of Neurophysiology* 72:1049–60. [rSE]

Tovée, M. J., Rolls, E. T. & Ramachandran, V. S. (1996) Rapid visual learning in the neurons of the primate temporal visual cortex. *Neuroreport* 7:2757–60. [MJT]

Treisman, A. M. & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology 12*:97–136. [RLG]

Tversky, A. (1977) Features of similarity. *Psychological Review* 84:327–52. [arSE]

Tversky, A. & Gati, D. (1978) Studies of similarity. In: *Cognition and categorization,* ed. E. Rosch & B. Lloyd. Erlbaum. [aSE]

(1982) Similarity, separability, and the triangle inequality. *Psychological Review* 89(2):123–54. [RAMG, ABM]

Ullman, S. (1980) Against direct perception. *Behavioral and Brain Sciences* 3:373–416. [aSE]

(1989) Aligning pictorial descriptions: An approach to object recognition. *Cognition* 32:193–254. [aSE]

(1995) Sequence-seeking and counter-streams: A model for information flow in the cortex. *Cerebral Cortex* 5:1–11. [aSE]

Unzicker, A., Jüttner, M. & Rentschler, I. (in press) Similarity models of human visual recognition. *Vision Research.* [rSE, MJ]

Väisälä, J. (1971) *Lectures on n-dimensional quasiconformal mappings. Lecture notes in mathematics, 229.* Springer-Verlag.

(1992) Domains and maps. In: *Quasiconformal space mappings. Lecture notes in mathematics, 1508,* ed. M. Vuorinen. Springer-Verlag. [aSE]

van Brakel, J. (1991) Meaning, prototypes and the future of cognitive science. *Minds and Machine* 1:233–57. [JvB]

(1993) The plasticity of categories: The case of colour. *British Journal for the Philosophy of Science* 44:103–35. [JvB]

van Essen, D. C., Newsome, W. T. & Maunsell, J. H. R. (1984) The visual representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. *Vision Research* 24:429–48. [GB]

van Leeuwen, C. (1997) Dynamical models of perceptual grouping. In: *Systems theories and a priori aspects of perception,* ed. J. S. Jordan. Elsevier. [CvL]

(1990) Indeterminacy of the isomorphism heuristic. *Psychological Research* 52:1- 4. [CvL]

van Leeuwen, C., Steijvers, M. & Nooter, M. (1997) Stability and intermittency in large-scale coupled oscillator models for perceptual segmentation. *Journal of Mathematical Psychology* 41:319–44. [CvL]

Verstijnen, I. M., van Leeuwen, C., Goldschmidt, G., Hamel, R. & Hennessey, J. M. (1997) Sketching and creative discovery. *Design Studies.* (in press). [CvL]

von Grünau, M., Dubé, S. & Galera, C. (1994) Local and global factors of similarity in visual search. *Perception and Psychophysics* 55:575–92. [HE]

von Helmholtz, H. (1856/1964) Unconscious conclusions. In: *Visual perception: The nineteenth century,* ed. W. N. Dember. Wiley. [aSE]

Wang, G., Tanaka, K. & Tanifuji, M. (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272:1665–67. [MJT]

Watanabe, S. (1985a) *Pattern recognition: Human and mechanical.* Wiley. [rSE]

(1985b) Pattern-recognition as value-oriented ponderation. In: *Pattern recognition: Human and mechanical.* Wiley. [rSE, UH]

Westheimer, G. (1981) Visual hyperacuity. *Progress in Sensory Physiology* 1:1–37. [aSE]

Williamson, J. R. (1996) Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps. *Neural Networks* 9:881–97. [SG, JRW]

(1997) A constructive, incremental-learning network for mixture modeling and classification. *Neural Computation* 9:1555–81. [JRW]

Willshaw, D. J., Buneman, O. P. & Longuet-Higgins, H. C. (1969) Non-holographic associative memory. *Nature* 222:960–62. [rSE]

Young, A. W. (1992) Face recognition impairments. *Philosophical Transactions of the Royal Society of London B* 335:47–54. [MJT]

Young, M. P. (1995) Open questions about the neural mechanisms of visual pattern recognition. In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. MIT Press. [MJT]

Young, M. P. & Yamane, S. (1992) Sparse population coding of faces in the inferotemporal cortex. *Science* 256:1327–31. [aSE]

Zorich, V. A. (1992). The global homeomorphism theorem for space quasiconformal mappings. In: *Quasiconformal space mappings. Lecture notes in mathematics, 1508,* ed. M. Vuorinen. Springer-Verlag. [aSE]

https://doi.org/10.1017/S0140525X98421253 Published online by Cambridge University Press