Representation, Similarity, and the Chorus of Prototypes

Shimon Edelman
Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
Internet: edelman@wisdom.weizmann.ac.il

December 1993; revised June 1994

Abstract

It is proposed to conceive of representation as an emergent phenomenon that is supervenient on patterns of activity of coarsely tuned and highly redundant feature detectors. The computational underpinnings of the outlined concept of representation are (1) the properties of collections of overlapping graded receptive fields, as in the biological perceptual systems that exhibit hyperacuity-level performance, and (2) the sufficiency of a set of proximal distances between stimulus representations for the recovery of the corresponding distal contrasts between stimuli, as in multidimensional scaling. The present preliminary study appears to indicate that this concept of representation is computationally viable, and is compatible with psychological and neurobiological data. *Keywords:* vision, categorization, representation, similarity, receptive fields, multidimensional scaling, feature spaces.

1 Introduction

A perceptual system confronted with a stimulus must decide whether it belongs to an already encountered category, and, if the stimulus is sufficiently novel, create a new category and store it for future reference. It is widely agreed that the crucial issue in the recognition of familiar stimuli and in the generalization to novel ones is that of representation. This paper proposes that *similarity relative to a small but diverse set of prototypes* is a natural computationally feasible candidate for a generic representation scheme, and is consistent with neurophysiological and psychophysical data.

The paper is organized as follows. Section 2 reviews briefly some of the problems involved in defining similarity, as they arise in philosophical, psychological, and computational discussions of representation. Section 3 then proposes a way of making similarity work that is based on mechanisms of transduction and pattern matching peculiar to biological information processing systems. Finally, sections 4 and 5 discuss the proposed theory of representation in a wider context of understanding perception, categorization, and learning.

2 The problem: representation of similarity

Consider the familiar notion of generalization (see Shepard, 1987, for a discussion): it is easier to respond intelligently to a stimulus if one can recall previous satisfactory (e.g., rewarded) responses made under similar circumstances. Thus, a perceptual system does well insofar as it succeeds to represent internally the similarities between different stimuli.

The problem in understanding how biological perceptual systems represent similarity, and in building artificial systems that do so, is that the notion of perceptual similarity is notoriously difficult to formalize (Quine, 1969). It has been pointed out repeatedly, by C. S. Peirce and others, that definitions in terms of shared or contrastive properties only beg the question of property selection. Borrowing an example from Murphy and Medin (1985), the number of attributes shared by plums and lawn-mowers could be infinite: both weigh less than 1000 kilograms (and less than 1001 kilograms), both cannot hear well, both have a smell, etc. Any two entities can thus be arbitrarily similar or dissimilar, depending on what is to count as a relevant property.

The same pitfall associated with the concept of similarity is illustrated by the following theorem due to Watanabe (1985): "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."[1] Watanabe's conclusion (see also Tversky, 1977) is that different weights must be assigned to different predicates. This, however, merely shifts the focus of the problem to the choice of the appropriate weights. Furthermore, before one can choose weights (e.g., in a model designed to fit psychological data on perceptual similarity) the predicates or features to be weighed must be somehow determined. Clearly, any conceivable approach to the choice of features and of their weights will necessarily constitute a kind of bias, be it theoretical or experimental. In the present work, I have chosen to assume a *natural* bias that follows the observations of Quine (1969):

> A response to a red circle, if it is rewarded, will be elicited again by a pink ellipse more readily than by a blue triangle; the red circle resembles the pink ellipse more than the blue triangle. Without some such prior spacing of qualities, we could never acquire a habit; all stimuli would be equally alike and equally different. These spacings of qualities, on the part of men and other animals, can be explored and mapped in the laboratory by experiments in conditioning and extinction. Needed as they are for all learning, these distinctive spacings cannot themselves be all learned; some must be innate.

A basic characterization of innate and acquired features of similarity may be derived from constraints imposed (1) by the patterns of natural kinds prevailing in the world;[2] (2) by the manner in which, in principle, distal objective similarities and dissimilarities can be mirrored in the proximal representations, and (3) by the architecture of a given perceptual system. The rest of this paper is devoted

[1] This theorem, which he called the Theorem of the Ugly Duckling, holds if the set of predicates is finite and equally applicable to all objects, and if no two objects are identical with respect to this set.

[2] I assume flatly ontological realism, and, in particular, realism about natural kinds.

to bringing these three kinds of constraints — physical, computational, and implementational (cf. Marr and Poggio, 1977) — to bear on the issue of representation by similarity.

3 The outline of a solution: a Chorus of Prototypes

3.1 Motivation

In the previous section we have seen that the notion of similarity defies a formalization in absolute terms. In the absence of such a formalization, we must either give up the attempt to derive an intuitively appealing theory of representation based on similarity, or, in the spirit of Quine (1969), tailor the concept of similarity to the means and needs of biological perceptual systems. This second alternative offers the possibility of developing a new and powerful theory of representation.

It may be observed that a biological system can only base its inferences about the world on the firing of its neurons (Poincaré, 1963; Bialek et al., 1991). Thus, at any stage in the processing hierarchy, differences between stimuli only matter insofar as they can be represented by the activity patterns of the preceding stage. This means that, for better or for worse, already at the output of the retina the vague notion of similarity between two stimuli gives way to a concrete concept of distance between their representations in the space spanned by the activities of ganglion cells, which *must* serve as the foundation to any possible metric computed by the subsequent levels of processing.

In the recent years, much effort has been devoted to the study of the ability of receptive fields (RFs) in the visual information processing pathway to support fine spatial analysis of the stimulus despite their large size already at the level of retinal ganglion cells — a puzzle as old as the concept of a receptive field (Hartline, 1938). Some of the pieces of the puzzle, known also as the phenomenon of hyperacuity (Westheimer, 1981), may be found (to mention some of the more recent works) in the computational analysis of Snippe and Koenderink (1992), and in the models of Poggio et al. (1992) and Weiss et al. (1993). According to the latest integrated understanding of hyperacuity, a collection of graded and highly overlapping RFs forms a representation that can support discrimination of fine spatial detail of the input (not necessarily via the recovery of the exact distribution of the retinal stimulation).

Computer simulations indicate that the information contained in this kind of representation is sufficient for discriminating among highly complex stimuli such as images of human faces (Weiss and Edelman, 1993). Experience with a face recognition system (Edelman et al., 1992) showed, however, that raw patterns of RF activities are better not stored and compared directly, and that an acceptably low error rate in face discrimination can only be achieved by a two-stage scheme (see appendix A, and Figure 5). In the first stage, the base representation is fed into a bank of individual classifiers, each of which is trained to respond to the face of a particular person. As several classifiers typically respond (more or less strongly) to any given face, the first stage thus computes an intermediate representation that encodes the distances between the input and each of the stored prototypes (faces best recognized by the individual classifiers). In the second stage, this set of distances is used to classify the input with a much greater precision than what is possible without the ensemble representation provided by the first stage.

3.2 The proposed theory: a Chorus of prototypes

These and other computer experiments in face recognition suggest that images of faces occupy a space whose dimensionality is significantly lower than the number of pixels in each image, which is the default dimensionality that must be assumed in the absence of evidence to the contrary (Kirby and Sirovich, 1990; Turk and Pentland, 1991). Psychophysical findings on synthetic 3D object recognition (Cutzu and Edelman, 1992) and on face discrimination (Rhodes, 1988) in human subjects confirm the relevance of low-dimensional approaches. The dimensions or features used by human subjects tend, however, to defy an easy and general computational characterization. In face discrimination, for example, the physical variables best correlated with the principal dimensions identified by multidimensional scaling analysis of face similarity ratings are sex and age (Rhodes, 1988). Thus, one way to understand, on the computational level, how faces can be represented in a low-dimensional space is to find out how sex and age can be determined from a face image — a formidable task by itself.

The success of the two-stage ensemble-based scheme for face recognition suggests an alternative approach to the low-dimensional representation of visual objects. The proposed approach, at the core of which there is a *Chorus* of prototypes, employs vectors of first-stage distances to a small number of reference objects to span the second-stage representation space.[3] In Chorus, the representation space for objects is built over a higher-dimensional space of primitive features (in the face recognition system discussed above, the primitive features are the activities of the simple receptive fields placed over the input image). Recurring stable patterns of primitive features, which are expected to correspond to persistent objects,[4] are represented explicitly, and constitute the prototypes that span the object space. Each persistent prototype may be represented by a set of detectors, implemented by receptive field-like mechanisms tuned to a number of the object's views (Poggio and Edelman, 1990), and may be constructed in a self-organizing fashion following mere exposure to the object (Edelman and Weinshall, 1991). In distinction to the persistent entities, rare or ephemeral patterns of primitive features are represented implicitly, by the distributed activity they induce in the prototype detectors (see Figure 1).

The power of ephemeral implicit representations stems from the same principle that makes multidimensional scaling (MDS) work: in a metric space, fixing the relative distances of a set of

[3]Webster's Dictionary has: cho.rus \'ko-r-*s, 'ko.r-\ n [L, ring dance, chorus, fr. Gk choros] 1a: a company of singers and dancers in Athenian drama participating in or commenting on the action. In a chorus, unlike in Selfridge's (1959) Pandemonium, the contributions of the individual actors are in harmony with each other. Lee Brooks (1987, p.165) uses the expression "chorus of instances" in his discussion of Medin and Schaffer's (1978) theory of representation.

[4]But which may also correspond to entities devoid of a real objecthood, just as canonical views of 3D objects sometimes cannot be represented as actual projections of the corresponding 3D shapes (Cutzu and Edelman, 1992).

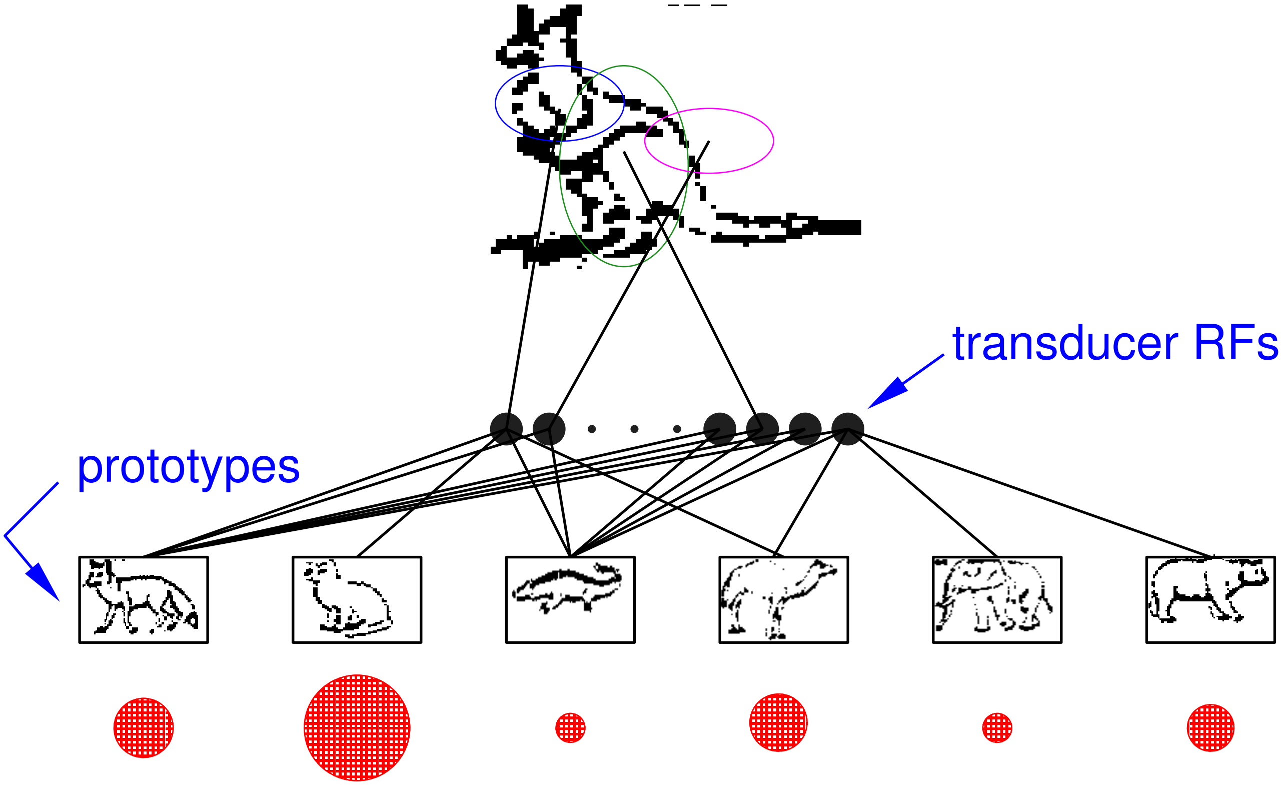

Figure 1: Presenting a novel stimulus to a system familiar with foxes, cats, alligators, camels, elephants, and bears. Each of the detectors for the familiar animals responds at a fraction of the maximal activity (the strength of the response is illustrated symbolically by the size of the disk beneath the detector box). Only part of the connections between layers are shown. Unlike in Selfridge's (1959) Pandemonium, all the responses and not merely the strongest one matter here.

points effectively determines their coordinates, up to a translation and rotation of axes (Shepard, 1980). If several basic requirements, listed in section 4.1 below, are satisfied, and if the input stimuli do in fact possess an (unknown) low-dimensional structure, the combined persistent/ephemeral feature method based on MDS is assured to recover a faithful replica of that structure, solely from qualitative (rank order) similarity measurements made in the space of primitive features (see Figure 2).

3.3 Persistent and ephemeral representations in Chorus

The postulated difference between persistent and ephemeral representations stems from the constraints imposed by limited resources in a real perceptual system: all objects cannot possibly be assigned individual representations. Fortunately, as we have seen above, MDS considerations indicate that only a relatively small number of objects need be represented in a persistent fashion. How can one decide whether a given stimulus has a persistent representation in a biological visual system? A criterion that seems proper is based on the notion *priming:* if repeated exposure modi-

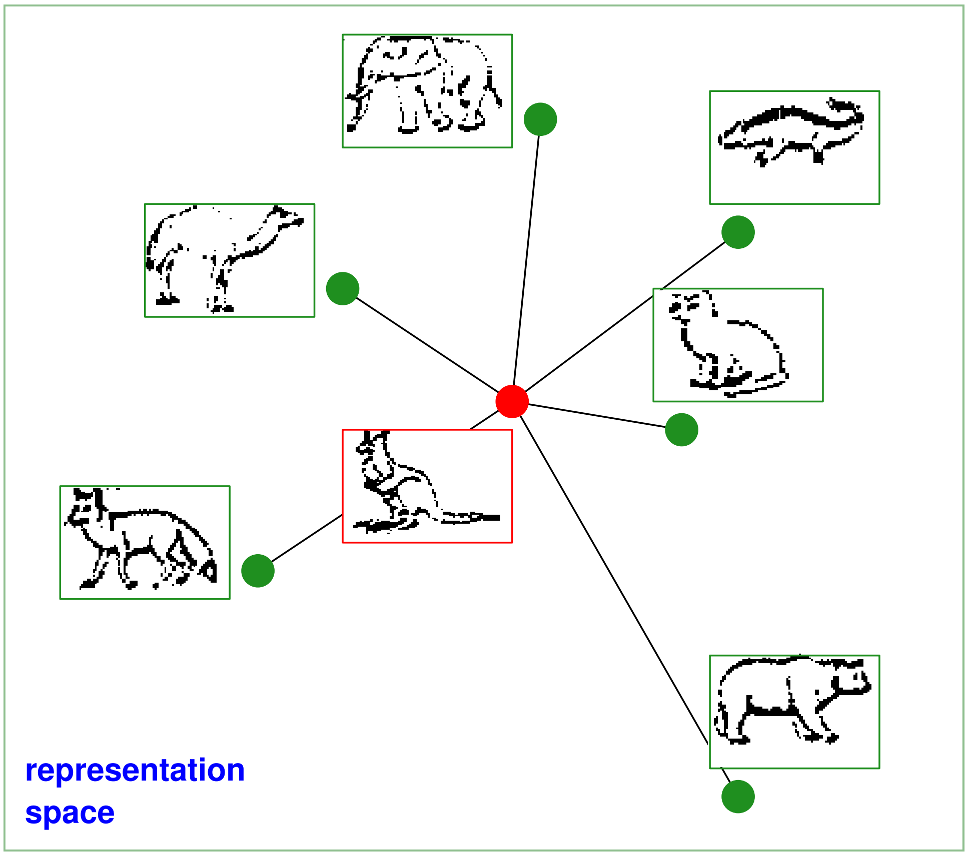

Figure 2: A different look at the situation depicted in Figure 1. The set of graded responses of the (persistent) detectors, each tuned to a familiar animal shape, can be processed by a technique related to multidimensional scaling (MDS) to yield the location of the (ephemeral) representation of the novel stimulus in the space "spanned" by the familiar shapes.

fies the response of the system to a stimulus, there is a good chance that this stimulus possesses a physically localized representation (e.g., a unit or a tightly coupled clique of units) which is affected (fatigued or excited) by the stimulation and which retains a memory of it over a certain period of time. This line of reasoning is developed, e.g., in (Biederman and Cooper, 1991), where it is argued that recognition priming at the level of object parts, or geons, constitutes evidence in favor of physically explicit geon-level representations.

There are two remarks to be made at this point. The first one has to do with the condition under which priming can be considered an indication as to the persistent nature of an internal representation. The condition is that of stimulus specificity, which figures prominently in the recent resurgence of work on perceptual learning (Sagi and Tanne, 1994). If the processing savings due to prior exposure do not transfer from one stimulus to another (despite these being merely rotated versions of the same pattern (Poggio et al., 1992)), then it is highly likely that a well-defined representation of the first stimulus is involved.

The second remark has to do with the amenability of the physical nature of the persistent representation mechanism to neurophysiological study. In principle, priming need not be physically local: units that participate in a "coalition" representing a given concept could also be primed, provided that the coalition persists at least for the duration of the experiment. In that case, the distinction between persistent and ephemeral representations would be blurred, and, moreover, would be difficult to demonstrate using currently available experimental methods in neurophysiology. On the other hand, the discovery of a class of (presumably ephemerally represented) objects immune to visual, as opposed to the ever-present "semantic," priming would strongly support such distinction.

The nature of what I have chosen to call ephemeral representations is well summarized by the following passage from (Barsalou, 1987):

> *Concepts as constructs.* Instead of viewing long-term memory as being divided into invariant concepts, it may make more sense to view long-term memory as containing large amounts of highly interrelated and "continuous" knowledge that is used to construct concepts in working memory.

Assuming that some basic or background abilities are necessary even in such a flexible representational framework, Barsalou's idea of "continuous knowledge" would translate into the following observation: persistent representations can only support the formation of useful ephemeral ones if they have (1) highly overlapping receptive fields, and (2) a graded structure. These two requirements correspond exactly to the conclusions of (Snippe and Koenderink, 1992) regarding the properties of receptive fields necessary for achieving hyperacuity-level performance (see section 3.1).

4 Discussion

4.1 Basic computational requirements of Chorus

Consider a collection of objects in the world that is to be represented in a perceptual system, and suppose that each object has a true physical description in terms common to all objects. While the object's 3D shape certainly constitutes an example of such a description, it has consistently proved extremely difficult to recover in a reliable fashion.

Luckily, for the purpose of classification only the comparisons or contrasts between the objects are interesting: if the world consisted of just one object, it would not really matter how that object were represented. Thus, recognition of a familiar shape and intelligent categorization of any shape do not require that the system recover those shapes in full 3D detail: it suffices to represent and use the low-dimensional information inherent in the structure of the natural kinds. Because neither this information nor the dimensions of the shape space are directly accessible to the perceptual system, the comparison between true distal descriptions of object shapes depends on a comparison of proximal descriptions related to the distal ones by multidimensional scaling (see Figure 3). If this comparison is to be faithful to the physical reality, several conditions must be met.

1. *Smoothness.* At the relevant level of description, the natural kinds are assumed to be related to each other by gradual change.[5]

2. *Monotonic covariation.* Differences between proximal descriptions must covary monotonically with differences between distal descriptions.[6] This is a basic requirement for the applicability of MDS (Shepard, 1980).

3. *Dynamic range.* For the proximal distances to be estimated reliably, the transduction mechanism must be allowed to operate within its dynamic range. Differences that are too small or too large would render the data submitted to MDS degenerate.

4. *Linearity.* This is a recommendation rather than a strict requirement. If the proximal measurements vary linearly (and not merely monotonically) with distal data, then metric MDS may be applicable. In that case, the system would require fewer persistent (reference) representations to achieve the same accuracy of discrimination.

The first point listed above, smoothness, is an assumption about the world, rather than a constraint on the structure of the perceptual system. Generalization would be impossible were it not for this property of the world, and, indeed, smoothness plays a central role in computational formulations of learning from examples (see section 4.4).

The second requirement — that of monotonic covariation — is of crucial importance in Chorus. For objects to be properly represented, the monotonically increasing difference between two objects must precipitate a concomitant increase in the difference between the patterns they evoke in a space spanned by a set of feature detectors. In vision, one may distinguish between the idealized case of objects consisting of points in 3D, in which monotonicity would have to follow from the geometrical optics involved in point transformation and projection, and the more realistic case of objects endowed with surfaces, in which monotonicity would have to depend on the imaging geometry, on surface photometry, and, eventually, on the transduction properties of the receptive fields in the processing pathway. A recent computational investigation indicates that monotonicity indeed obtains in both those cases (see Duvdevani-Bar and Edelman, 1994, for details).

As to the last two points in the above list, they may be seen to parallel the requirements of high degree of overlap or redundancy, and of graded profile of RFs. The redundancy requirement is fulfilled by biological perceptual systems at every level of representation. Moreover, the sigmoidal response characteristics of units throughout the visual system (and, in fact, the sigmoidal psychometric curves produced by a behaving organism as a whole) can frequently be linearized around a given operating point, making linearity a plausible assumption in many cases.

[5]Cf. John Locke: "...in all the visible corporeal World, we see no Chasms, or Gaps. All quite down from us, the descent is by easy steps, and a continued series of Things, that in each remove, differ very little one from the other." (Kornblith, 1993, p.20).

[6]Locke's much criticized concept of representation by covariation (Cummins, 1989) may after all deserve a reconsideration.

4.2 Relationship to multidimensional scaling

The role of multidimensional scaling in the proposed representational framework must be further clarified. MDS has been originally developed as a method for the recovery of a metric structure of a set of points from measurements of quantities monotonically related to pairwise distances between those points. When MDS is applied as a tool for the study of internal representations, care must be taken to ensure that its basic assumptions are satisfied (e.g., that it makes sense to assume that the representation space is metric, etc; (Beals et al., 1968)). Unlike in the application of MDS in psychophysics, where the inference is from the derived overt measurements to the primary structure of the hidden inner space, in perception the purported inference is from the derived inner (proximal) similarities to the primary distal ones. This means that the metric properties are attributed first and foremost to the distal (real-world) entities. Locke's observation regarding the "continuity" of real-world objects, as well as the more recent exercises in computer graphics in which 3D objects are made to deform smoothly into each other, should convince us that this attribution is not entirely unfounded. As to the nature of the inner representation space, the monotonic transduction process enables this space to reflect the distal metrics.

Faithful proximal recovery of the metrics of the distal space does not preclude additional factors (such as top-down influences) from introducing occasional violations of the metric axioms; it merely provides a principled basis for the representation of the smooth order of the natural kinds, which can be subsequently warped, as in the phenomenon of categorical perception (Harnad, 1987). When such warping occurs, associations acquired through experience or instruction may affect the structure of the representational space by tying together some of its points that normally are far apart. As a simple example, one may think of the association between the ringing of a bell and the smell of food in Pavlov's dogs.[7]

4.3 The hierarchy of features and dimensionality reduction

An implementation of Chorus along the lines suggested above would include a layer of primitive feature detectors, an intermediate layer of persistent features, and an output layer of ephemeral feature detectors. Two questions that may be raised regarding this structure are (1) whether the persistent representation layer can be left out altogether, and (2) whether more than one such layer would impute additional computational power to the system. Suppose that the primitive features are oriented patterns similar to the receptive field profiles of simple cells in the primary visual cortex. An activity of a collection of such feature detectors can signal reliably the presence of a complex object such as a face. If a new face is shown to the system, the primitive feature pattern will change according to the dissimilarity between the new face and the old one. However, this change will reflect distance in the primitive feature space, rather than in "face space" (see Figure 4). Consequently, a

[7] In visual neurophysiology, cortical representation of random associations between image pairs has been demonstrated in the monkey (Sakai and Miyashita, 1992); a possible computational role of such associations is discussed in (Edelman and Weinshall, 1991).

system without a persistent and dedicated representation of faces would be subject to an atomistic bias of the kind found, e.g., in pigeons (Cerella, 1987), but not in humans.

The above intuition can be made more precise by invoking the notion of task-dependent dimensionality reduction. Note that whereas the pattern of activity at the primitive feature level constitutes a fine-grained representation of the stimulus, it is not particularly suitable for classification, because it involves, at the same time, similarity to all possible parts of all objects known to the system. The move to the persistent representation layer is more useful in this respect, because it corresponds to similarity in a considerably lower-dimensional shape space. Moreover, dimensionality reduction performed by measuring distances between the stimulus and the persistent complex features has the desirable mathematical property of approximate isometry (that is, it is likely to preserve the metric structure of the input space; see Duvdevani-Bar and Edelman, 1994). Thus, the involvement of persistent features permits dimensionality reduction (an essential step in any system that is to learn from examples (Poggio and Girosi, 1989)) to be done in a principled manner, preserving the metric information inherent in the primitive-feature representation.

Would additional "hidden" layers of (persistent) feature detectors improve the learning ability and the generalization performance of Chorus? If learning is treated as function approximation (Poggio, 1990), one hidden layer suffices under a broad range of conditions on the inputs and on the primitive features (Cybenko, 1989; Girosi and Poggio, 1990; Hartman et al., 1990). It should be noted, however, that if the metrics of the top-level representation space are to differ qualitatively from the metrics at the base representation level, more layers of function approximation modules may become necessary. Consider, for example, the case of a 3D object undergoing transformation in space. As pointed out by Shepard (1987), generalization to a novel pose of such an object may be nonmonotonic, depending on the symmetry properties of the object. Similarly, Biederman and Cooper (1991) showed that subjects generalize perfectly across a mirror reflection of 3D objects. To account for such nonmonotonic (at the level of the primitive features) generalization, an additional level of processing in the visual system may have to be postulated, which would explicitly detect and represent (or cancel out) the appropriate transformations. Alternatively, the requisite "distortions" in the global metrics of the representation space can be introduced by long-range associations, as hinted at the end of section 4.2.

4.4 Relationship to the Hyper Basis Function theory of the brain

The Chorus scheme bears a close relationship to the theory of the brain recently proposed by Poggio (1990). This theory postulates that the main computational challenge which the brain must meet is learning from examples, or, more specifically, smooth approximation of the function that maps a stimulus to its desired response (or, in the case of perception, to the desired inner representation). Given the values of the function at a number of sample points, the value at a new point is found essentially by taking a linear combination of appropriately scaled basis functions centered on the sample points (Poggio and Girosi, 1990).

A clear analogy exists between the basis functions of the HBF model and the prototypes of Chorus: the response profile of the prototype unit in Chorus can be considered as a basis function in a network implementing HBF approximation. This analogy, as well as some fundamental conceptual differences between the two schemes, are discussed below.

Smoothness vs. monotonic covariation. In HBF, the shape of the basis functions is determined by the kind of smoothness assumption on the class of target functions that are to be approximated (Poggio and Girosi, 1990). Similarly, in Chorus it is required that the physical description of an object change smoothly as it becomes less and less like the prototype of its class. However, Chorus imposes the additional requirement of monotonic covariation, over and above smoothness, and when the monotonicity fails, the fidelity of the representation will suffer.

Consider, for example, objects composed of clouds of points in 3D. When two such objects rotate in space, the image-plane distance between their chosen views (defined, e.g., as the sum of distances between corresponding points) changes smoothly with rotation, but does not change monotonically.[8] Interestingly, there are psychophysical indications (Edelman and Bülthoff, 1990) that subjects find it more difficult to generalize over rotation in depth than over other transformations that are not even rigid, such as 3D shear (note that shear causes a monotonic increase in the 2D pointwise distance to a reference or prototype view).

Classification vs. recognition The above example of the failure of monotonicity suggests the following division of labor between HBF approximation and Chorus (which, as a matter of fact, can be implemented by a two-layer system of HBF modules; see appendix A). On the one hand, in the recognition of different views of the same object, where the smoothness assumption is warranted over the entire range of possible rotations (Ullman and Basri, 1991), but the monotonicity only holds for a relatively small range of views around a given reference view, a straightforward application of HBF appears to be a useful strategy (Poggio and Edelman, 1990). On the other hand, in the classification of potentially unfamiliar objects, where in general there is no theoretical guarantee of the applicability of HBF approximation, the reliance on monotonic transduction and a Chorus-like scheme may be a more acceptable approach.

The required number of prototypes. In HBF, the minimum number of training stimuli necessary for achieving a certain probabilistically guaranteed level of generalization performance can be determined using the tools of computational learning theory (Haussler, 1992; Edelman, 1993). If additional stimuli for which the generalization proves to be poor become available, the architecture of the HBF module can be modified to accommodate the new data, without increasing the number of layers (Platt, 1991). Similar considerations apply to Chorus, where the prototypes must cover

[8] Recall how strange did the "backtracking" of the apparent trajectories of the planets seem to the ancient astronomers.

the input space so that there is high likelihood that enough prototype units respond above floor and below saturation, for any possible input.

Relationship to coarse coding. Both HBF approximation and the interaction of prototypes in Chorus may be considered as computational implementations of the well-established notion of coarse coding (Feldman and Ballard, 1982). Importantly, these two theories also contribute to the explanatory value of coarse coding, by tying it to the notion of hyperacuity: what makes it work is the computational advantage conferred by the use of overlapping graded receptive fields. On a more general level, HBF and Chorus relate to complementary aspects of coarse coding. HBF and the universal approximation theorems on which HBF relies reveal the power of coarse coding as a device for mapping stimuli into representations. In comparison, Chorus stresses the importance of identifying the conditions under which the resulting representations can be provably relevant to the real world.

4.5 Biological considerations

4.5.1 Psychophysics

Subjects in a wide variety of generalization studies in different perceptual modalities have been found to behave as if they represent the stimuli in a low-dimensional psychological space (Shepard, 1987), as would be expected from an approach based on comparison of the stimulus with a small number of prototypes. In a majority of these studies, however, the stimuli were structurally simple, indicating that, at best, similarity relationships with respect to prototypes hold at the lower levels of the relevant perceptual subsystems. For complex 3D object discrimination, results indicating possible involvement of low-dimensional Chorus-type representations are becoming available (Cutzu and Edelman, 1992; Edelman, 1994; see also appendix B). More research is needed to substantiate these findings, and to define the relationships between Chorus and psychological theories of categorization and recognition based on multivariate approaches (for a recent review see Ashby, 1992).

4.5.2 Physiology

The shallow representational hierarchy posited by Chorus is compatible with the current notions of the function of the shape processing stream in primate vision. In the primate visual system, the role of primitive features can be assigned tentatively to the orientation-selective simple and complex cells in the primary visual cortex. Persistent representations would then correspond to the inferotemporal (IT) cortex cells selective for faces and face parts (Gross et al., 1972; Perrett et al., 1982; Perrett et al., 1989); these representations are expected, to a certain extent, to be modifiable by practice (Rolls et al., 1989). Other objects, for which no specially selective cells have been found, may be represented ephemerally.

The cells in V4 and IT selective for well-defined and rather complex shapes, described by Tanaka and his collaborators (Tanaka, 1992), are another example of possible hardwired persistent representations. (Fujita et al., 1992) recently reported that the shape-selective cells may in fact be arranged in a columnar format, ordered by shape preference, along the surface of the cortex. This finding is particularly relevant because orderly columnar arrangement is frequently explained by appeal to the need for an analog representation of similarity by physical distance on the cortex. Such an analog mechanism would greatly assist the implementation of Chorus in neural hardware.

It is interesting to note that the receptive field, which is, according to the original definition the part of the visual space to which a given unit is sensitive, can actually be defined with respect to three different spaces. The first two of these are defined by the transformations under which the retinal projection of the stimulus preserves its rigid structure. Thus, the translation space yields the classical notion of retinotopic receptive field, while the rotation about the optical axis of the system exposed to a bar stimulus yields what is usually termed the orientation selectivity curves of, say, the simple cells in V1. Finally, there is the shape space, whose dimensions are defined by the possible deformations of the stimulus. In this space, the receptive field means simply the shape selectivity profile of the unit in question.

The present work concentrates on the properties of receptive fields in the shape space, where the main computational problem is making sense of objects not previously seen before. The problems of dealing with objects translated or rotated in depth (and the associated notions of receptive fields in transformation spaces) are considered to be of secondary importance. The reasons for this decision are derived from the availability of biologically motivated models that can support invariance to rigid transformations and to scaling (Schwartz, 1985; Anderson and Essen, 1987), or tolerate rotation in depth by learning to compensate for it from examples (Poggio and Edelman, 1990). The possibility of discounting such transformations prior to dealing with shape is actually beneficial for the present model. A more intriguing idea is that of mutability of the receptive fields in the shape space itself. For example, it has been demonstrated that the selectivity profile of V4 receptive fields can be manipulated by the parameters of the task (Spitzer et al., 1988). The computational considerations stated in section 4.1 suggest that the shape selectivity profile of IT cells should be sufficiently broad, as in (Desimone et al., 1984): "most of the stimulus-selective cells gave at least a small response to virtually every stimulus tested, especially complex stimuli." A subsequent increase in the response specificity (leading to the relatively sharply tuned responses such as those found by Tanaka's group) can be precipitated by familiarity with the stimuli (Li et al., 1993).[9] Findings of such plasticity of receptive fields in shape space should eventually determine the degree of permanence of the persistent representations postulated by Chorus.

[9]Notably, a massive increase in the incidence of IT cells selective to a set of shapes was found after prolonged exposure to the chosen shapes (Tanaka, 1993).

4.6 Philosophical considerations

The central philosophical statement of Chorus, being largely an allusion to Locke's notion of representation by covariation, has to do with the problem of distal knowledge. This classical epistemological problem resurfaced in experimental psychology because of the need to provide a foundation for multidimensional scaling, conceived as a psychologist's tool in understanding perception. Because the present paper aims to introduce Chorus as a computational foundation for representation, I will avoid philosophical technicalities such as the formal content of representation by prototypes (*pace* (Fodor, 1981); these issues will be discussed elsewhere), and will remark instead on the psychological roots of Chorus, having discussed its computational characteristics in the preceding sections.

Much of the original motivation for the development of multidimensional scaling was provided by considerations not unlike the notion of multiple simultaneous measurements underlying Chorus. For example, in an early work, Thurstone (1927) proposed an empirical law relating the psychological distance between two stimuli to the dispersion of their difference as judged by an observer over a series of trials. Thurstone also suggested (*ibid.*, p.278) that an equivalent law may hold in the case involving *many observers* each of whom makes a single comparative judgment (the analogy here is between a collection of observers making a simultaneous judgment and a system of persistent feature detectors in Chorus). This multiple-"observer" approach appears to provide a realistic basis for the modeling of statistical aspects of perception. Real perceptual systems rarely have the chance to experience a given stimulus repeatedly (Bialek et al., 1991), hence the value of a simultaneous analysis of the input via a number of parallel channels. Related ideas may be found in Brunswik's (1956) "lens model" of perception and response, and in Feigl's (1958) notion of "triangulation." Campbell (1985) formulates it as follows: "From several widely separated proximal points, there is triangulation upon the distal object, "fixing" it and its distance in a way quite impossible from a single proximal point. Binocular vision can be seen quite literally as such a triangulation."[10]

5 Conclusions

5.1 Summary

The central tenet of Chorus is that a perceptual object can be effectively represented by computing its similarity to a collection of prototypes of related object classes. A *set* of similarity values, one per prototype, is needed; if these values are summed, as in (Nosofsky, 1991), important information is lost. Chorus is motivated by empirical observations of the performance of a computer scheme for face recognition (Edelman et al., 1992), and by recent neurobiological findings regarding the functional

[10] Of course, triangulation is in fact the least difficult problem in stereopsis, the most difficult one being the recovery of the correspondence between scene features in the two images (Marr and Poggio, 1979). Similarly, in an application of MDS to perception, care must be taken to ensure that the difficult part of the procedure (e.g., the choice of transducer RF profiles and of the prototypes to be stored) is computationally feasible.

architecture of the inferotemporal cortex in the monkey (Tanaka, 1992). Computationally, Chorus relies on two phenomena:

- *Hyperacuity:* the possibility of achieving hyperacuity-level performance in a system of overlapping graded receptive fields (Altes, 1988; Snippe and Koenderink, 1992);
- *Multidimensional scaling:* the sufficiency of a set of perceptual (proximal) distances between stimuli representations for the recovery of the corresponding distal contrasts between stimuli (Shepard, 1980).

The proposed concept of representation appears to be computationally viable, and is compatible with a range of findings from psychophysics and neurobiology of vision.

5.2 Implications

The "binding problem." By offering an alternative to the structural approaches to representation, the present framework may obviate the computational need for binding, because it represents structure effectively by responding multiply and selectively to structure present in the input.[11] The so-called binding problem arises because of the natural propensity of the structural approaches first to take the represented objects apart (by describing them in terms of generic primitives), only to face later the problem of putting them together again, usually in the form of a computational structure known as an attributed graph. Modelers nowadays tend to "solve" the binding problem by postulating mechanisms for biological implementation of abstract graphs, such as coupled oscillations that maintain phase lock across significant distances in the cortex (Hummel and Biederman, 1992). No evidence for such mechanisms has, however, been found in primates (Young et al., 1992). Thus, a scheme that circumvents binding altogether (see Figure 4) enjoys the advantage of biological plausibility over the standard structural approaches.

Invariances. Because Chorus relies on a proximal recovery of the metrics of the distal shape space, any factor that acts along the pathway that leads from the true shape via the imaging process to the internal representation may interfere with the veridicality of that representation. For example, Adini et al. (1993) showed that the pixel-based distance between two images of the same face taken under two different illuminant directions may be greater than the distance between the images of two different persons under similar illumination. In other words, in this case the metrics prevailing in the image space as a result of illumination changes would not allow Chorus to operate properly. It appears, however, that this problem largely disappears already in the space of center-surround receptive fields resembling those of the ganglion cells at the output of the retina, because of the spatial frequency selectivity of such RFs (Weiss and Edelman, 1993). It remains to be seen whether the principle of "amending the metrics" can be applied to the understanding of other biological information processing subsystems.

[11] This may be compared to the idea of representations being in the world, rather than in the head (Putnam, 1988).

5.3 Prospects

The applicability of MDS to the understanding of biological information processing in areas other than vision appears to be worth exploring. Olfaction seems to be a promising candidate for modeling in terms of representation by similarity (see Granger and Lynch 1991, p.211; neurophysiological findings that agree with a recent hierarchical clustering model of the piriform cortex have been reported in (McCollum et al., 1991)). Another area in which similarity-based models are likely to emerge is speech perception (Miller and Eimas, 1979). In psychophysical problems that arise in all domains of perception, MDS may provide the guiding principle whereby a perceptual system can estimate physical (distal) qualities from the psychological (proximal) measurements it performs on the world.

As a more far-fetched issue, one may consider the relevance of MDS-like models to the understanding of how language bridges the gap separating the mental representation spaces of the communicants. Across this gap, the similarity structure of the individual representation spaces may still be recoverable via MDS, from the patterns of activities evoked internally by exposure to the linguistic stimuli. In this connection, it is interesting to note that both epistemological considerations (Quine, 1960; Quine, 1969) and psycholinguistic evidence (Markman, 1989) point towards the advantage of holistic treatment of stimuli in the process of concept acquisition by ostension (that is, in learning by hearing the word applied to samples of the concept, without prior knowledge of the relevant features of the concept; see Quine, 1969). This bias towards an initially holistic approach in concept learning may correspond to the need to acquire a rudimentary basis of persistent prototypes, which then serve as yardsticks used by the perceptual system to measure the world (Stich, 1990; Margolis, 1991). Thus, it may turn out that at least part of the philosophy of representation *is* a footnote — not to Plato, but to Protagoras.

Acknowledgments

Part of this work has been presented at the Workshop on Mechanisms of Visual Object Recognition, organized by Charles Stevens and Michael Stryker at the Santa Fe Institute in August 1993. Thanks to Shabtai Barash and Tommy Poggio for timely encouragement, to Florin Cutzu, Sharon Duvdevani-Bar, Daniel Glaser, Nathan Intrator, Yair Weiss, Stevan Harnad, and an anonymous reviewer for useful discussions and suggestions, to the subjects of the psychophysical experiments for their time and patience, and to all the students in the seminar on Computational Neuroscience of Representation held at Weizmann during the Spring 1993 semester, for wonderful time. This report describes work in progress, supported by grants from the Basic Research Foundation, administered by the Israel Academy of Arts and Sciences, and from the Grodetzky Center for the Study of Higher Brain Function at the Weizmann Institute of Science. SE is an incumbent of the Sir Charles Clore Career Development Chair at the Weizmann Institute of Science.

References

Adini, Y., Moses, Y., and Ullman, S. (1993). Face recognition: the problem of compensating for changes in illumination direction. CS-TR 21, Weizmann Institute of Science.

Altes, R. A. (1988). Ubiquity of hyperacuity. *J. Acoust. Soc. Am.*, 85:943–952.

Anderson, C. H. and Essen, D. C. V. (1987). Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Science*, 84:6297–6301.

Ashby, F. G., editor (1992). *Multidimensional models of perception and cognition.* Erlbaum, Hillsdale, NJ.

Barsalou, L. W. (1987). The instability of graded structure: implications for the nature of concepts. In Neisser, U., editor, *Concepts and conceptual development*, pages 101–140. Cambridge Univ. Press.

Beals, R., Krantz, D. H., and Tversky, A. (1968). The foundations of multidimensional scaling. *Psychological Review*, 75:127–142.

Bialek, W., Rieke, F., de Ruyter Van Steveninck, R. R., and Warland, D. (1991). Reading a neural code. *Science*, 252:1854–1857.

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

Brooks, L. R. (1987). Decentralized control of categorization: the role of prior processing episodes. In Neisser, U., editor, *Concepts and conceptual development*, pages 141–174. Cambridge Univ. Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* U. of California Press, Berkeley, CA.

Campbell, D. T. (1985). Pattern matching as an essential in distal knowing. In Kornblith, H., editor, *Naturalizing epistemology*, pages 49–70. MIT Press.

Cerella, J. (1987). Pigeons and perceptrons. *Pattern Recognition*, 19:431–438.

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

Cutzu, F. and Edelman, S. (1992). Viewpoint-dependence of response time in object recognition. CS-TR 10, Weizmann Institute of Science.

Cybenko, G. (1989). Approximations by superpositions of sigmoidal functions. *Math. Control, Signals, Systems*, 2:303–314.

Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. J. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.*, 4:2051–2062.

Duvdevani-Bar, S. and Edelman, S. (1994). Representation by Chorus of Prototypes. in preparation.

Edelman, S. (1993). On learning to recognize 3D objects from examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:833–837.

Edelman, S. (1994). Representation of similarity in 3D object discrimination. CS-TR 94-02, Weizmann Institute of Science.

Edelman, S. and Bülthoff, H. H. (1990). Generalization of object recognition in human vision across stimulus transformations and deformations. In Feldman, Y. and Bruckstein, A., editors, *Proc. 7th Israeli AICV Conference*, pages 479–487. Elsevier.

Edelman, S., Reisfeld, D., and Yeshurun, Y. (1992). Learning to recognize faces from examples. In Sandini, G., editor, *Proc. 2nd European Conf. on Computer Vision, Lecture Notes in Computer Science*, volume 588, pages 787–791. Springer Verlag.

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

Feigl, H. (1958). The 'Mental' and the 'Physical'. In Feigl, H., Scriven, M., and Maxwell, G., editors, *Concepts, theories, and the mind-body problem.* U. of Minnesota Press, Minneapolis. MN.

Feldman, J. A. and Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, 6:205–254.

Fodor, J. A. (1981). *RePresentations.* MIT Press, Cambridge, MA.

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

Girosi, F. and Poggio, T. (1990). Networks and the best approximation property. A.I. Memo 1164, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Granger, R. and Lynch, G. (1991). Higher olfactory processes: perceptual learning and memory. *Current Opinion in Neurobiology*, 1:209–214.

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of cells in inferotemporal cortex of the macaque. *J. Neurophysiol.*, 35:96–111.

Harnad, S., editor (1987). *Categorical Perception: The Groundwork of Cognition.* Cambridge University Press, New York.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am. J. Physiol.*, 121:400–415.

Hartman, E. J., Keeler, J. D., and Kowalski, J. M. (1990). Layered neural networks with Gaussian hidden units as universal approximations. *Neural Computation*, 2:210–215.

Haussler, D. (1992). Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100:78–150.

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

Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loève procedure for characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108.

Kornblith, H. (1993). *Inductive inference and its natural ground.* MIT Press, Cambridge, MA.

Li, L., Miller, E. K., and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *J. of Neurophysiology*, 69:1918–1929.

Margolis, J. (1991). *The truth about relativism.* Basil Blackwell, Oxford, UK.

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

Marr, D. and Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, 15:470–488.

Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, 204:301–328.

McCollum, J., Larson, J., Otto, T., Schottler, F., Granger, R., and Lynch, G. (1991). Short-latency single-unit processing in olfactory cortex. *Journal of Cognitive Neuroscience*, 3:293–299.

Miller, J. and Eimas, P. (1979). Feature detectors and speech perception: a critical evaluation. In Albrecht, D., editor, *Recognition of Pattern and Form (Lecture Notes in Biomathematics)*, volume 44, pages 111–145. Springer, Berlin.

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

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, 17:3–27.

Perrett, D. I., Mistlin, A. J., and Chitty, A. J. (1989). Visual neurones responsive to faces. *Trends in Neurosciences*, 10:358–364.

Perrett, D. I., Rolls, E. T., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.*, 47:329–342.

Platt, J. (1991). A resource-allocating network for function interpolation. *Neural Computation*, 3:213–225.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:899–910.

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

Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, 256:1018–1021.

Poggio, T. and Girosi, F. (1989). A theory of networks for approximation and learning. A.I. Memo No. 1140, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Poggio, T. and Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982.

Poincaré, H. (1913/1963). *Mathematics and Science: Last Essays*. Dover, New York. translated by J. W. Bolduc.

Putnam, H. (1988). *Representation and reality*. MIT Press, Cambridge, MA.

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

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

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception*, 17:43–63.

Rolls, E. T., Baylis, G. C., Hasselmo, M. E., and 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. *Exp. Brain Res.*, 76:153–164.

Sagi, D. and Tanne, D. (1994). Perceptual learning: learning to see. *Current opinion in neurobiology*, 4:195–199.

Sakai, K. and Miyashita, Y. (1992). Neural origanization for the long-term memory of paired associates. *Nature*, 354:152–155.

Schwartz, E. L. (1985). Local and global functional architecture in primate striate cortex: outline of a spatial mapping doctrine for perception. In Rose, D. and Dobson, V. G., editors, *Models of the visual cortex*, pages 146–157. Wiley, New York, NY.

Selfridge, O. G. (1959). Pandemonium: a paradigm for learning. In *The mechanisation of thought processes*. H.M.S.O., London.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210:390–397.

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

Snippe, H. P. and Koenderink, J. J. (1992). Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, 66:543–551.

Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240:338–340.

Stich, S. (1990). *The fragmentation of reason*. MIT Press, Cambridge, MA.

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

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

Thurstone, L. L. (1927). The law of comparative judgement. *Psychological Review*, 34:273–286.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 3:71–86.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84:327–352.

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

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

Weiss, Y. and Edelman, S. (1993). Representation with receptive fields: gearing up for recognition. CS-TR 93-09, Weizmann Institute of Science.

Weiss, Y., Edelman, S., and Fahle, M. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Computation*, 5:695–718.

Westheimer, G. (1981). Visual hyperacuity. *Prog. Sensory Physiol.*, 1:1–37.

Young, M., Tanaka, K., and Yamane, S. (1992). On oscillating neuronal responses in the visual cortex of the monkey. *J. of Neurophysiology*, 67:1464–1474.

Appendices

A Computer recognition of faces

This appendix is a short account of the experiments with a two-stage computer program for face recognition reported in (Edelman et al., 1992). In the first stage (see Figure 5), the base representation (the activities of seventy-odd receptive fields spread over the input image) was fed to a bank of 16 individual classifiers, each implemented as RBF networks (Poggio and Girosi, 1990; Poggio and Edelman, 1990) and trained to respond to the face of a particular person. The second stage RBF module was trained on the vectors of responses of the individual classifiers and was required to produce a vector of length 16 with just one dominant component (corresponding to the recognized individual).

The recognition program was tested on a subset of the MIT Media Lab database of face images made available by Turk and Pentland (1991), which contained 27 face images of each of 16 different persons. The images were taken under varying illumination and camera location. Of the 27 images available for each person, 17 randomly chosen ones served for training the program, and the remaining 10 were used for testing. A different recognizer was created for each person, and was trained to output 1 for the images in the training set.

The performance of the individual recognizers was assessed by computing a 16×16 confusion table, in which the entries along the diagonal signified mean miss rates and the off-diagonal entries — mean false alarm rates (see Figure 6).[12] An examination of the confusion table reveals that some of the individuals tended to be confused with almost any other person in the database. To take advantage of this "ensemble phenomenon," another RBF module was trained to accept vectors of individual recognizer activities and to produce vectors of the same length in which the value corresponding to the activity of the correct recognizer was 1, and all other values were 0 (see Figure 5). The training set for the second-stage RBF module was obtained by pooling the training sets of all 16 first-stage recognizers. The outcome of the recognition of a test image was determined by finding the coordinate in the output vector whose value was the closest to 1. The performance of the two-stage scheme was considerably better than that of the individual recognizer stage alone (9% error rate, compared to 22%), demonstrating the importance of ensemble knowledge for recognition.

[12] The table was computed row by row, as follows. First, recognizer for the person whose name appears at the head of the row was trained. Second, the recognition threshold was set to the mean output of the recognizer over the training set less two standard deviations. Third, the performance of the recognizer on the test images of the same person was computed and the miss rate entered on the diagonal of the table. The above choice of threshold resulted in a mean miss rate of about 10%. Finally, the false alarm rates for the recognizer on the images of the other 15 persons were computed and entered under the appropriate columns of the table.

B Similarity to prototypes in 3D shape discrimination

Subjects in a wide variety of generalization tasks in a number of perceptual modalities behave as if they represent the stimuli in a low-dimensional psychological space (Shepard, 1987, for a review). The four experiments, described fully in (Edelman, 1994) and summarized here, were designed (1) to find out whether similarity in a low-dimensional feature space is a good predictor of performance in 3D shape discrimination, and (2) to characterize such a feature space in objective terms.

Fourteen subjects performed a delayed match to sample task, with each of the two stimuli belonging to a set of 16 images (2 object classes $\times$ 2 exemplars $\times$ 4 orientations). The object classes in the four experiments were, respectively, animal shapes, scrambled animal shapes, wires made of distinctive 3D segments (geons), and wires made of plain cylinders. Stimuli were rendered as shaded matte metal and displayed on a computer. Response time (RT) data from each subject were entered in a 16×16 confusion table arranged by stimulus identities (RTs of correct "yes" responses unmodified; RTs of correct "no" responses as $\max_{subj} - \mathrm{RT}$; erroneous responses yielded missing values).

The RT data were assumed to be monotonically related to distances among relevant representations of the stimuli in the hypothesized feature space. The confusion tables were submitted to a nonmetric multidimensional scaling analysis. For animal shapes, but not for wires, the resulting 2D configuration revealed an astonishingly faithful replica of the low-dimensional structure of the parameter space used to generate the stimuli, namely, the distinction between the two object classes, and the orthogonal within-class variation. A similar configuration was obtained with a prototype-based model that operated on the same images seen by the human subjects (a simpler model based on dissimilarities between vectors of receptive field activities did not perform as well). This finding is compatible with the notion that the feature space involved in the classification decision made by the subjects is spanned by distances to the class prototypes, as called for by the Chorus scheme.

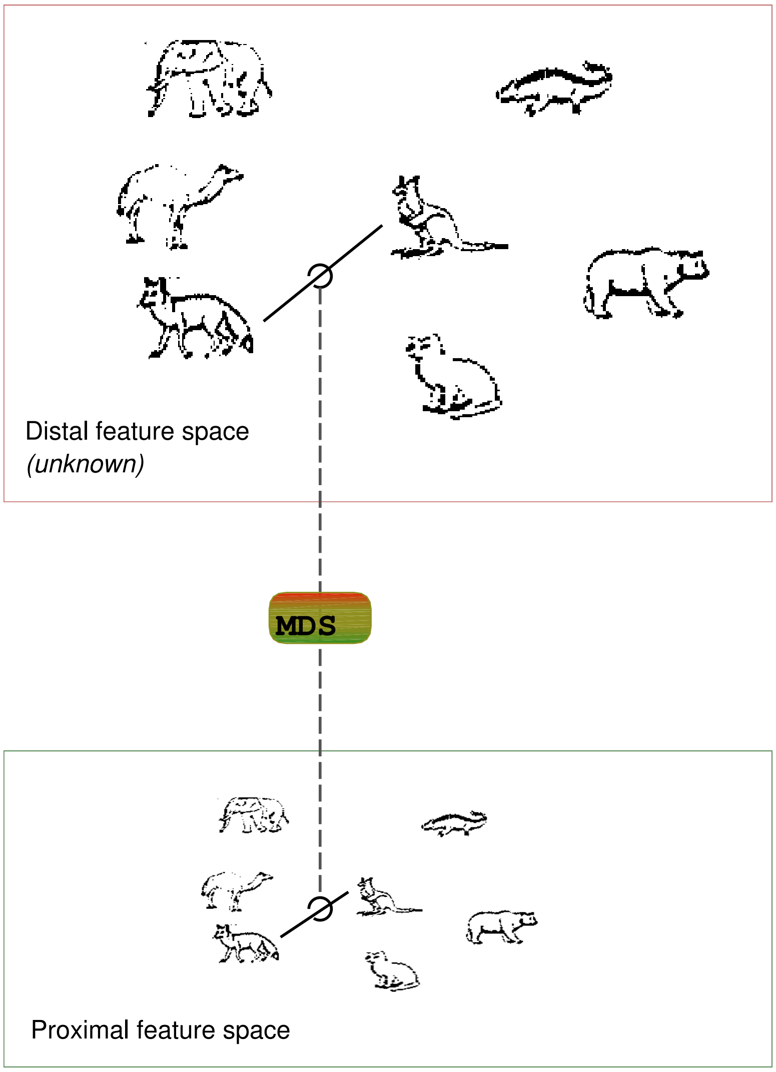

Figure 3: Multidimensional scaling (MDS) and the problem of distal knowledge (in biological systems, the proximal representations are really patterns of responses of spatial filters, and not little pictures of the represented objects; see section 3.2). A discussion of this problem and of possible approaches to its solution can be found in (Campbell, 1985).

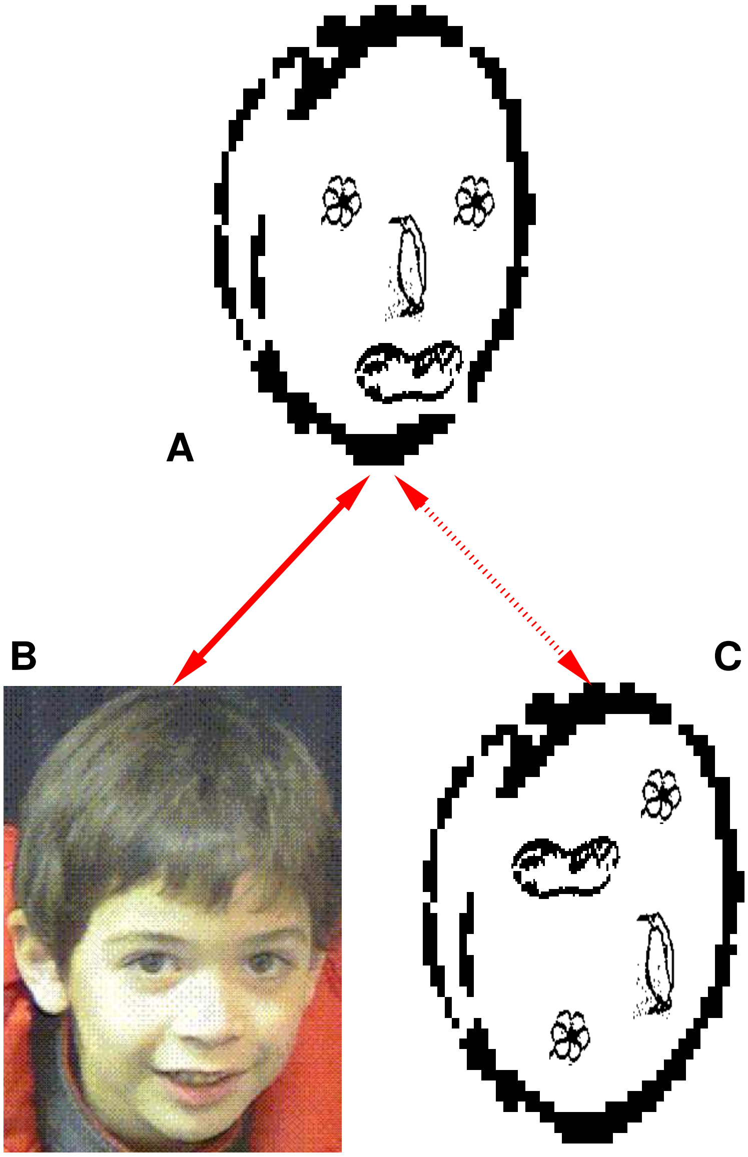

Figure 4: The requirement that the representation of the face in image A be closer to that in image B than to C constrains the level at which faces should be represented. Specifically, a representation at the level of mere presence of individual features appears inadequate; the spatial relationships among the features must be encoded as well. Within the Chorus framework, these relationships are encoded at the level of persistent representations.

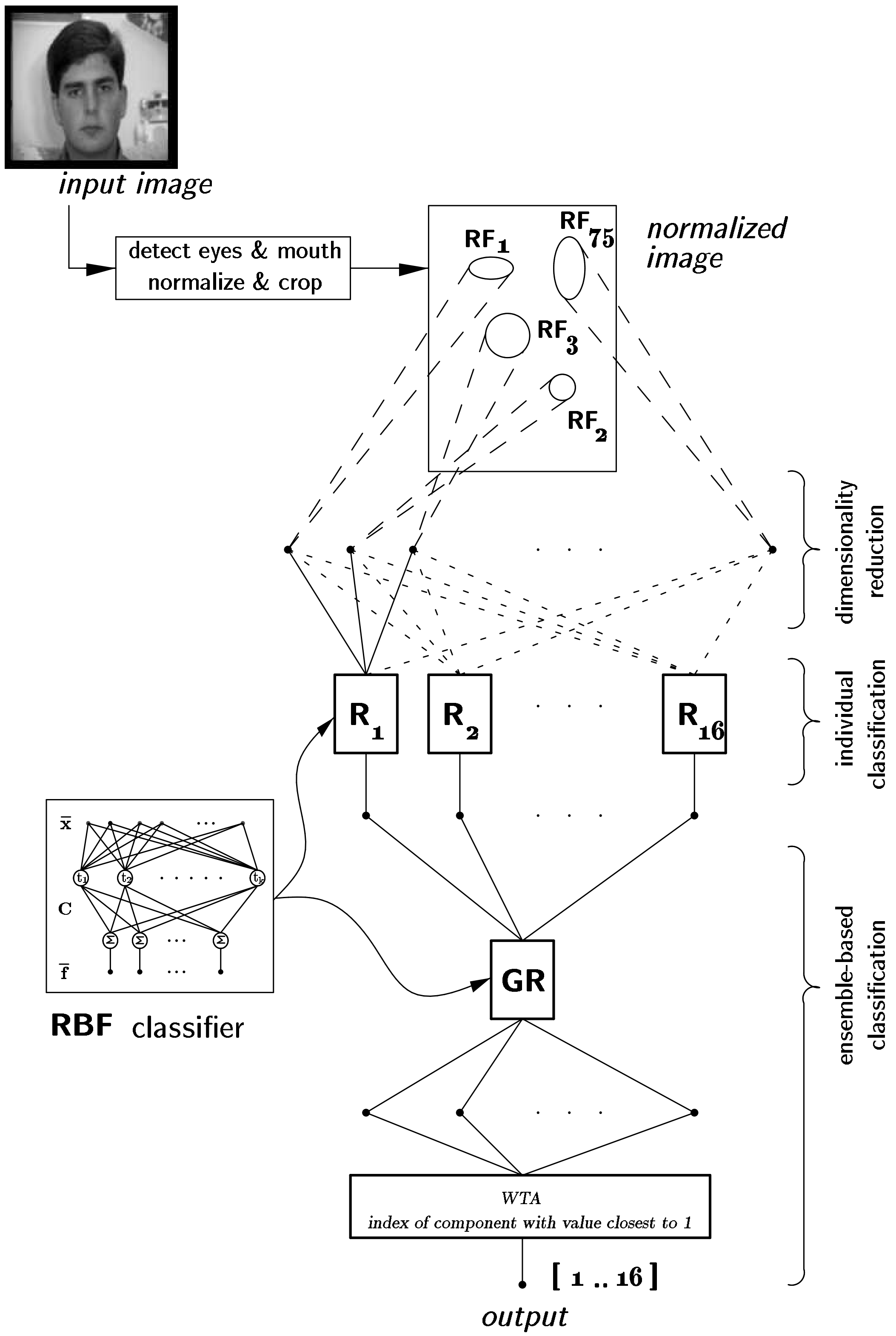

Figure 5: The two-stage scheme for face recognition (Edelman et al., 1992); see appendix A.

train\test	bil	bra	dav	foo	irf	joe	mik	min	pas	rob	sta	ste	tha	tre	vmb	wav
bil	**0.1**		0.2	0.1					0.1			0.2				
bra						0.4							0.4			
dav																
foo				**0.1**												
irf		0.3		0.1	**0.1**	0.4			0.1		0.5	0.3	0.2	0.5	0.2	0.1
joe		0.1											0.3			
mik		0.1					**0.1**				0.1					
min	0.3		0.5	0.9		0.2		**0.1**	0.3	0.8		0.8		0.6		0.6
pas	1.0		1.0	0.2	0.1				**0.1**	0.5		0.8		0.4		0.5
rob				0.2						**0.1**		0.3		0.6		
sta		0.4				0.5					**0.1**	0.1	0.4		0.1	
ste												**0.1**		0.4		
tha						0.1										
tre				0.2				0.1		0.1		0.2		**0.1**		
vmb		0.3				0.2					0.1		0.4		**0.1**	
wav																**0.1**

Figure 6: A confusion table representation of the performance of the first stage of the face recognition system, described in appendix A. Entries along the diagonal correspond to the "miss" error rates; off-diagonal entries signify the "false-alarm" error rates (zeros omitted for clarity).